The Cavity Method in Coding Theory

THÈSE N^O 6088 (2014)

PRÉSENTÉE LE 21 FÉVRIER 2014
À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE THÉORIE DES COMMUNICATIONS
PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Marc VUFFRAY

acceptée sur proposition du jury:

Prof. E. Telatar, président du jury
Prof. R. Urbanke, Dr N. Macris, directeurs de thèse
Dr M. Chertkov, rapporteur
Prof. H.-A. Loeliger, rapporteur
Dr P. Vontobel, rapporteur

EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2014

Acknowledgements

I first thank my master co-advisor Philippe Choquard, a truly fascinated and passionate man about physics, for introducing me to Nicolas Macris. It was by chance that I met Nicolas during a conference, where he offered me the possibility to work on this exciting interdisciplinary subject that mixes statistical physics and information theory.

I'm deeply indebted to Nicolas without whom this thesis would not have been possible. He gave me a lot of his time and I benefited considerably from his valuable advice, insight and comments that he provided generously and patiently.

It has been a pleasure and a privilege to work under the direction of Rüdiger Urbanke and Nicolas Macris. I would like to thank them both to have welcomed me in the great family which is the Communication Theory Laboratory.

I thank my thesis committee members Emre Telatar, Misha Chertkov, Hans-Andrea Loeliger and Pascal Vontobel for their careful reading of my manuscript and their suggestions.

Many thanks to Muriel Bardet and Françoise Behn who work behind the scenes to manage the administration and to organize the events which consolidate the social links of the lab. I also thank Damir Laurenzi and Giovanni Cangiani for baking us delicious cakes and for teaching computers how to remain in operating states.

Life at the lab would not have been so colorful without the presence of my colleagues. I thank in particular Mine Alsan and Saeid Haghighatsoar with whom I discussed during tea breaks; the foosball partners Adrian Tarniceriu, Karol Kruzelecki, Stefano Rosati, Youg Jun Ko and Vassilis Kalofolias; Marc Desgroseillers who is constantly happy, cheerful and looking for a victim to whom he can explain the hidden beauty of some obscure theorem; my conference partner Hamed Hassani who is the most polite, discrete, modest but hard-working person I know. A special thank you to Alla Merzakreeva for sharing her contagious enthusiasm about travelling and skiing, and to Andrei Giurgiu for his kindness, sense of humor and passion for talking about history, geography and every other subject that is contained in the depths of Wikipedia.

Acknowledgements

I was very lucky to have two exceptional officemates who are Eren Sasoglu and Vahid Aref.

Eren is extremely curious about everything and I had numerous interesting discussions with him on broad topics ranging from the physics of playground swings to synesthesia. I will deeply miss sharing an office with Vahid with whom I built up a great spirit of collaboration and complicity. His open mind, his jokes, his talent and our discussions around a cup of tea made the time spent at the office infinitely more enjoyable and enriching.

My close friends enabled me to live extraordinary moments during this period whether by going through marshes and jungles with a little canoe or playing board games around a table in the comfort of a warm home. I especially thank Blaise Petitpierre, François Sahy, Sylvie Stucki, Patrick Mingard and Diane Perez for those wonderful adventures.

Above all I would like to thank Rafah for her help, encouragement, generosity, comprehension, affection, kindness and constant happiness. You gave me the best thymes.

I dedicate my thesis to my family and in particular to my sisters Muriel and Sandrine and my parents Michel and Martine. It is difficult to express in these lines how much I esteem them and how much I am indebted to them for their unwavering support, their kindness and love that they have provided during all these years. I will be forever grateful.

Lausanne, 5th February 2014   M. V.

Abstract

Systems found in nature are composed of numerous components that interact with each other. Sometimes a small change in the microscopic interactions is accompanied by an abrupt change in the material properties. This phenomenon is called a phase transition. A simple example of a phase transition is the transformation of a liquid into a solid. A transition that remained to this day one of the most mysterious is the glass transition. The glass transition is present in amorphous materials and manifests itself in a slowdown of the dynamical properties of the material. A glass can be seen in many aspects as a frozen liquid.

The phase and glass transition are not only found in nature, but also in artificial systems used in computer science and communications. In this context, these transitions appear as sudden changes in the execution or in the performances of algorithms. In particular, we retrieve the glass transition in systems where the interactions between constituents are represented by a sparse graph. It appears that these systems are similar to spin glasses that were developed in statistical physics to study the glass transition. Within this framework, a non-rigorous mathematical tool, called the cavity method, has been developed to analyze these models.

In this thesis we focus on the application of the cavity method in coding theory. Our work focuses on two topics related to code constructions based on sparse graphical models.

Our first contribution concerns two error correcting codes: low-density parity-check (LDPC) codes and low-density generator-matrix (LDGM) codes. These codes are used to correct the errors that appear when information is transmitted over an imperfect channel. The conditional entropy is an important quantity for analyzing the performance of a code, but difficult to compute. The cavity method predicts that the conditional entropy is asymptotically equal to another quantity that is easy to compute: the Bethe free energy. We present a rigorous proof that the prediction of the cavity method is exact for LDGM and LDPC codes over a binary symmetric channel (BSC) in a high noise regime. For this we develop a new technique that combines loop calculus and the polymer expansion from statistical physics. The conditional entropy is computed as a series expansion containing the Bethe free energy as its zero-th order term plus some correction terms. We show that the series is convergent for graphical model at high temperatures. Moreover if the graphs have large girth, we prove that the correction terms are asymptotically vanishing. This

Abstract

applies immediately to LDGM codes. The application to LDPC codes requires more care as the parity constraints in such codes correspond to a low temperature regime. Nevertheless, by combining the properties of graph expansion and counting arguments, we can prove that the conditional entropy is asymptotically equal to Bethe free energy.

Our second contribution concerns lossy source coding. Our objective is to compress a binary symmetric source. It is well-known that once compressed, the source cannot be reconstructed without a certain number of errors that is called "distortion". For a fixed compression rate, the distortion cannot be smaller than the Shannon distortion. We investigate an encoding scheme based on spatially-coupled LDGM codes. The LDGM codes that we consider have check nodes with regular degree and variable nodes with random degree distributed according to a Poisson distribution. We show that these codes equipped with a low-complexity encoding algorithm, approach the Shannon distortion for any compression rate. The algorithm we consider is a belief-propagation guided-decimation (BPGD) process. The BPGD algorithm depends on a probability distribution on the space of compressed sequences. This distribution can be interpreted as the Gibbs measure of a spin glass at finite temperature. We compare the properties of this Gibbs measure and the performance of the BPDG algorithm. Therefore, we study the phase diagram of this measure using the cavity method. In particular, we compute two transition temperatures called dynamical temperature and condensation temperature. We observe that, in the case of spatially coupled LDGM codes, the dynamical temperature saturates to the condensation temperature. This saturation mechanism is similar to the so-called threshold saturation that appears in channel coding with spatially-coupled codes. In addition, we show that, as the degree of the check nodes increases, the typical sequences of the Gibbs measure at the condensation temperature give rise to a distortion that tends to the Shannon distortion.

Keywords: Spin glass, cavity method, dynamical and condensation temperatures, Bethe free energy, polymer expansion, loop calculus, lossy source coding, low-density generator-matrix codes, low-density parity-check codes, spatial coupling, threshold saturation, rate-distortion bound, belief propagation, decimation

Résumé

Les objets qui nous entourent sont composés de nombreux constituants en interactions. Il arrive qu'une petite modification dans les interactions microscopiques s'accompagne d'un changement abrupt dans les propriétés du matériau. On parle alors de transition de phase. C'est ce qui arrive par exemple lorsque qu'un liquide se transforme en solide. Une des transitions qui reste à ce jour parmi les plus énigmatiques est la transition vitreuse. Présente dans les matériaux amorphes, elle se manifeste sous la forme d'un fort ralentissement de la dynamique sous-jacente. Ainsi un verre ressemble sous bien des aspects à un liquide figé.

Les transitions de phase et la transition vitreuse ne sont pas uniquement présentes dans la nature mais aussi dans certains systèmes artificiels que l'on rencontre en informatique ou en communications. Dans ce contexte, ces transitions se manifestent par des changements soudains dans le fonctionnement d'un algorithme ou dans les performances de celui-ci. On retrouve la transition vitreuse en particulier sur des modèles où les interactions entre les constituants sont représentées par des graphes creux. Ces derniers possèdent une structure mathématique proche de ce qui est appelé des verres de spins avec interactions diluées. Les verres de spins sont des modèles théoriques de physique statistique qui furent développés pour étudier la transition vitreuse. Plusieurs outils mathématiques furent inventés dans ce cadre pour comprendre et étudier ces modèles.

Dans cette thèse nous nous concentrons sur l'application dans le codage de l'un de ces outils, qui est non rigoureux, appelé la méthode de la cavité. Notre travail porte sur deux sujets d'étude relatifs à des codes construits à partir de graphes creux.

Notre première contribution concerne deux codes correcteurs d'erreurs : les codes à contrôle de parité de faible densité (LDPC) et les codes à matrice génératrice creuse (LDGM). Ces codes sont utilisés pour se protéger contre une altération de données sur un canal de transmission imparfait. La qualité d'un code se mesure par le nombre d'erreurs qu'il est capable de corriger. Une quantité difficile à calculer mais primordiale dans le contrôle de la qualité d'un code est l'entropie conditionnelle. La méthode de la cavité prédit que cette entropie conditionnelle est asymptotiquement égale à une quantité aisément calculable : l'énergie libre de Bethe. Nous prouvons rigoureusement que dans le cas d'un canal de transmission binaire fortement bruité, la prédiction de la méthode de la cavité est exacte pour les codes LDGM et LDPC. Pour ce faire nous développons une nouvelle

Abstract

technique combinant la série des boucles et le développement en série de polymères de la physique statistique. L'entropie conditionnelle apparait comme étant égale à une somme dont le premier terme est l'énergie libre de Bethe suivi de termes correctifs. Nous montrons que pour les modèles graphiques à haute température cette somme est convergente. De plus si les graphes en question ont une grande maille, nous prouvons que les termes correctifs à l'énergie libre de Bethe sont asymptotiquement nuls. Ce résultat s'applique immédiatement aux codes LDGM. Le cas des codes LDPC est plus délicat car les contraintes de parité correspondent à un régime de basse température. Néanmoins en combinant les propriétés d'expansion des graphes et des arguments de comptage, nous prouvons que les termes correctifs à l'énergie libre de Bethe sont asymptotiquement nuls.

Notre deuxième contribution concerne le domaine du codage de source avec perte. L'objectif est ici de compresser une source d'information binaire et symétrique. Il est bien connu qu'une fois comprimée, cette source ne peut pas être reconstruite qu'avec un certain nombre d'erreurs appelé distorsion. Pour chaque taux de compression, il existe une limite inférieure à la distorsion atteignable qui s'appelle la distorsion de Shannon. Notre étude porte sur un procédé d'encodage basé sur des codes LDGM spatialement couplés. Les codes LDGM que nous considérons ont des nœuds de contrainte réguliers et des nœuds de variable ayant des degrés aléatoirement distribués selon une loi de Poisson. Nous montrons que ces codes associés à un algorithme de basse complexité peuvent approcher la distorsion de Shannon quel que soit le taux de compression. L'algorithme dont il est question est un processus de décimation de graphe guidé par une propagation de croyance (BPGD). L'algorithme BPGD est dépendant d'une distribution de probabilité sur les séquences compressée. Cette distribution de probabilité peut être interprétée comme une mesure de Gibbs d'un verre de spins à température finie. Nous comparons les propriétés de cette mesure de Gibbs aux performances de l'algorithme BPGD. Pour cela nous étudions le diagramme de phase de cette mesure en utilisant la méthode de la cavité. En particulier nous calculons deux températures de transition appelées température dynamique et température de condensation. Nous observons que dans le cas des codes LDGM spatialement couplé, la température dynamique sature vers la température de condensation. De plus nous montrons que lorsque le degré des nœuds de contrainte augmente, les séquences typiques de la mesure de Gibbs à la température de condensation donnent lieu à une distorsion approchant la distorsion de Shannon.

Mots clés : Verre de spin, méthode de la cavité, température dynamique et de condensation, énergie libre de Bethe, développement en série de polymères, série des boucles, codage de source avec perte, codes à contrôle de parité de faible densité, codes à matrice génératrice creuse, couplage spatial, saturation de seuil, courbe de distorsion, propagation de croyance, décimation

Contents

Contents

List of Figures

List of Tables

Introduction Part I

Preamble

Systems found in nature involve many constituents that interact locally with each other. Simple relationships between parts can give rise to complex collective behaviors of a whole system. We cite the examples of ants exchanging pheromones and building an anthill, neurons transmitting electrical pulses or avalanches in a pile of colliding sand grains. This mechanism is not only found in nature but also in artificial systems such as those that are encountered in information theory and coding, statistical inference, theoretical computer science, bio-informatics and signal processing. In these latter models, the dependencies between constituents are often represented by sparse random graphs. It is surprising that these systems are similar to spin glasses with dilute interactions, a subject studied in statistical mechanics.

Spin glass models are toy examples that were developed in the 1970s to understand the behavior of disordered systems and try to unravel the mystery of the glass transition in amorphous materials. In 1975, D. Sherrington and S. Kirkpatrick introduced one of the first exactly solvable models of spin glasses that now bears their name [1]. Four years later G. Parisi solved the SK model, perfecting the replica trick [2, 3], a highly non-rigorous method involving ill-defined mathematical objects. In their 1985 paper "SK Model: The Replica Solutions without Replicas" [4], M. Mezard, G. Parisi and M. A. Virasoro developed the cavity method in an attempt to replace the lack of rigor of the replica trick by a more probabilistic approach.

The cavity method combines the advantages of reproducing the replica trick results and is much more intuitive. The cavity method can be regarded as a series of successive approximation steps called "replica symmetry breaking (RSB) steps" and is a non-trivial generalization for loopy graphs of the Bethe-Peierls approach on trees. This makes its formalism closely related to algorithms such as belief propagation (BP) or min-sum algorithm (MSA). For this reason the cavity method has been applied fruitfully outside statistical physics in fields as diverse as biology [5, 6], compressed sensing [7, 8], machine learning [9, 10], neural networks [11, 12], coding theory [13, 14] and constraint satisfaction problems [15, 16]. The strength of the cavity method is its ability to provide theoretical insights, as well as algorithms to compute entropies, thresholds or ground-state energies. However, formulas predicted by the cavity method remain unproven in most of cases.

Despite the success of the rigorous approach by Guerra and Talagrand [17, 18] for the SK model, the whole field is very much at its beginning, and there remain many open problems.

This thesis focuses on the applications of the cavity method in coding theory. The core of the present document consists of two parts.

Part II concerns channel coding on the binary symmetric channel with codes based on sparse graphical models. In this context the cavity method is at its minimal level of complexity (no RSB steps) and reduces to the Bethe-Peierls mean field theory. It has already been used extensively to provide computationally efficient decoding algorithms and formulas that predict the performance of such codes. *Our contribution lies in a mathematically rigorous proof of the validity of the Bethe formula for computing the conditional entropy of code instances. The techniques we present are new and are based on a combinatorial tool from statistical physics, called the polymer expansion.*

Part III presents a different flavor in that it consists in an exploratory work in lossy source coding theory and in the use of the cavity method for interpreting algorithmic results. *We show that spatially-coupled codes approach the rate-distortion bound under a low-complexity message-passing and decimation algorithm.* Spatial coupling, a graphical construction recently discovered in channel coding, has quickly proved fruitful in other areas of study. Using the full machinery of the cavity method, we give an interpretation of the performance improvements in terms of glass transitions. We also provide tools for computing performance thresholds and show that the so-called threshold saturation mechanism takes place.

The present introductory part consists of three chapters. In Chapter 1, we first present the general formalism of graphical models and explain briefly the basic assumptions of the cavity method. Most of the concepts that are at the basis of the cavity method come from glass theory and statistical physics. Therefore we present a review of the glass phenomenology and the main results on the Ising model, along with the cavity method. In Chapter 2, we express two problems from coding theory in the language of spin glasses. The first is the computation of the conditional entropy for sparse graph codes in channel coding, and the second is the application of spatially-coupled codes in lossy source coding theory. We also provide a general review of the channel coding theory and the rate distortion theory. Chapter 3 contains the statement of the main contributions, along with a summary of Part II and Part III.

1 Spin Glasses and the Cavity Method

In this chapter, we provide an introduction to spin glasses, a particular type of random graphical models.

In Section 1.1, we present certain concepts from statistical physics. In particular, we introduce the free energy as a central object to study graphical models.

In Section 1.2, we illustrate the concepts of the preceding section with the square lattice Ising model and introduce the concept of extremal measures.

In Section 1.3, we give a review of the glass phenomenology and describe some properties of the glass transition.

In Section 1.4, we present the notion of quenched disorder and describe some of the most iconic spin glasses.

In Section 1.5, we provide a high level introduction of the main concepts of the cavity method.

1.1   Graphical Models and Concepts from Statistical Physics

In coding or spin glass theory, we encounter probability distributions that take values on binary alphabets that could be either bits $u \in \{0, 1\}$ or equivalently spins $\sigma \in \{-1, 1\}$. A factor graph $\Gamma = (V, C, E)$ is a bipartite graph that represents the factorization properties of such distributions. It is defined by a set of variable nodes V, a set of function nodes or check nodes C and a set of edges $E \subset V \times C$. The number of variable nodes is $n = |V|$ and the number of check nodes is $m = |C|$. The letters $i, j, k \in V$ refer to variable nodes and $a, b, c \in C$ to check nodes. We also use the notation $\partial i = \{a \in C \mid (i, a) \in E\}$ (resp. $\partial a = \{i \in V \mid (i, a) \in E\}$) to denote the variable neighborhood (resp. check neighborhood). Variable nodes $i \in V$ are associated with

binary variables $u_i \in \{0,1\}$ and check nodes are associated with non-negative weight functions ψ_a whose arguments are $\underline{u}_{\partial a} := \{u_i \mid i \in \partial a\}$. The weight functions have the general form

$$\psi_a \left(\underline{u}_{\partial a}\right) = \exp\left(-\beta H \left(\underline{u}_{\partial a} \mid x_a\right)\right), \tag{1.1}$$

where H is a real valued function, possibly infinite, that could depend on a local parameter x_a and on a global parameter $\beta > 0$. The collection of local parameters is denoted by $\underline{x}$. By analogy with the corresponding parameter entering the Gibbs distribution of physical models, β is called the inverse temperature. The Gibbs measure associated with the factor graph Γ is

$$\mu_\beta \left(\underline{u} \mid \underline{x}\right) = \frac{1}{Z \left(\beta \mid \underline{x}\right)} \prod_{a \in C} \psi_a \left(\underline{u}_{\partial a}\right). \tag{1.2}$$

The normalization factor Z in Equation (1.2) is called the partition function and is a sum over all configurations $\underline{u} \in \{-1,1\}^n$ of the product of the interaction functions

$$Z \left(\beta \mid \underline{x}\right) = \sum_{\underline{u} \in \{0,1\}^n} \prod_{a \in C} \psi_a \left(\underline{u}_{\partial a}\right). \tag{1.3}$$

1.1.1   The Free Energy

Most often, we are interested in computing the average of a quantity $A \left(\underline{u}\right)$ under the measure μ. This average is traditionally denoted with brackets as

$$\langle A \rangle_{\beta|\underline{x}} = \sum_{\underline{u}} \mu_\beta \left(\underline{u} \mid \underline{x}\right) A \left(\underline{u}\right). \tag{1.4}$$

It turns out that this can always be mapped to the computation of a central object called the free energy

$$f \left(\beta \mid \underline{x}\right) = -\frac{1}{n\beta} \ln Z \left(\beta \mid \underline{x}\right). \tag{1.5}$$

Finding $\langle A \rangle$ is equivalent to first computing a perturbed version of the free energy (here $\epsilon > 0$ is a small perturbation parameter)

$$f \left(\beta; \epsilon \mid \underline{x}\right) = -\frac{1}{n\beta} \ln \left(\sum_{\underline{u}} e^{\epsilon A(\underline{u}_I)} \prod_{a \in C} \psi_a \left(\underline{u}_{\partial a}\right)\right), \tag{1.6}$$

and taking a derivative with respect to the perturbation parameter

$$-\frac{1}{n\beta} \langle A \rangle_{\beta|\underline{x}} = \left. \frac{d}{d\epsilon} f \left(\beta; \epsilon \mid \underline{x}\right) \right|_{\epsilon=0}. \tag{1.7}$$

For this reason, it is often said that the free energy contains all the "useful information" about the Gibbs measure μ. However the perturbed free energy (1.6) could be much harder to compute than the free energy (1.5) of the original problem.

The free energy is a central object to understanding a graphical model. Unfortunately it is an object that is a priori hard to compute. The partition function (1.3) is a sum involving 2^n terms. When there exist symmetries in a problem, for instance if the graph is a lattice or if the interactions are uniform ($\psi_a = \psi$), the complexity can sometimes be reduced drastically. This is, however, never the case for glassy systems that are at the core of many problems in computer science. These are "disordered" systems with no explicit symmetries.

1.2 An Important Paradigm: the Square Lattice Ising Model

Graphical models involve a large number of degrees of freedom (say $u_i \in \{0, 1\}$); for instance 10^4 in coding or 10^{23} in statistical physics. In general, we are interested in the asymptotic properties of a system. The limit when the number of variables tends to infinity is called the thermodynamic limit. The thermodynamic limit of the free energy probes the sudden changes of the structure of Gibbs measures as a control parameter (e.g., the temperature) varies. Consider the free energy in the thermodynamic limit as a function of the control parameter. When this quantity has a point of non-analyticity, we say that the system undergoes a phase transition. Phase transitions are ubiquitous in nature; for instance, water turns to steam or snow, or a piece of iron acquires a magnetization. Phase transitions are also present in information science when an overload of noise, pushing a system off capacity, suddenly disrupts all communications.

One of the earliest attempts to model a phase transition was the square lattice Ising model. This model combines the benefits of being both simple enough to be analytically solvable and complex enough to have a phase transition.

The Ising model for a one-dimensional chain of spins was first introduced and solved by Ising [19] in his 1924 thesis. Although this model was developed for studying the ferromagnetic transition in matter, there is no phase transition in 1D. The two dimensional version of the problem is much harder and was solved analytically in 1944 by Onsager [20]. Unlike the one-dimensional system, the Ising model on the square lattice possesses a paramagnetic-ferromagnetic phase transition at finite temperature where a spontaneous magnetization develops.

Denote by $\Lambda_L = (V, E)$ the square lattice with $n = L^2$ spins with periodic boundary conditions (Λ_L is equivalent to the torus $(\mathbb{Z}/L\mathbb{Z})^2$). We will use the letters i, j for vertices in V and we denote by a couple $(i, j) \in E$ an edge of the lattice. In this example we can dispose of the bipartite structure of factor graphs as the function nodes are in one-to-one correspondence with the edges. A spin is a binary variable $\sigma_i \in \{-1, 1\}$ associated with a vertex $i \in V$ of the lattice. A spin configuration is a vector of n binary variables $\underline{\sigma} \in \{-1, 1\}^n$. The occurrence probability of a configuration at temperature β^{-1} is given

by the Gibbs measure

$$\mu_n\left(\underline{\sigma}\right) = \frac{1}{Z_n} e^{-\beta H_n(\underline{\sigma})}, \tag{1.8}$$

where the energy cost function, called the Hamiltonian, contains only nearest neighbor interactions

$$H_n\left(\underline{\sigma}\right) = \sum_{(i,j)\in\Lambda_L} -J\sigma_i\sigma_j. \tag{1.9}$$

The coupling constant J in the Hamiltonian (1.9) is non-negative. Such interactions between spins are called ferromagnetic interactions.

The free energy of this model first computed by Onsager is given in terms of an elliptic integral. We first introduce the dimensionless rescaled temperature

$$z := \tanh\left(\beta J\right). \tag{1.10}$$

The temperature is minimal when $z = 0$ and becomes maximal when $z = 1$. The free energy in terms of the rescaled temperature is

$$f_0\left(z\right) = \frac{\ln 2}{2} + \frac{1}{2\pi}\int_0^\pi d\theta \ln K\left(z,\theta\right), \tag{1.11}$$

where

$$K\left(z,\theta\right) = \frac{\left(1+z^2\right)^2 + \sqrt{1 - 4z^2 + 22z^4 - 4z^6 + z^8 - 8z\left(1-z^2\right)^2\cos\left(2\theta\right)}}{\left(1-z^2\right)^2}. \tag{1.12}$$

The free energy is a continuously differentiable function of z, but it is not analytic on the whole interval $z \in [0,1]$. The second derivative becomes singular at the point

$$z_c = \sqrt{2} - 1 \approx 0.414. \tag{1.13}$$

The temperature z_c is the critical temperature (recall $z = \tanh\left(\beta J\right)$) at which the square-lattice Ising model undergoes a phase transition. To understand the nature of the phase transition and its implication on the structure of the Gibbs measure (1.8), we introduce the concept of extremal measures, which is central in Gibbs measures theory. It is well known that in the thermodynamic limit, Gibbs measures of the form (1.8) are contained in a set closed under convex combinations. Different Gibbs measures can be obtained using different boundary conditions. The pure states or extremal measures are Gibbs measures that cannot be expressed as a convex combination. For periodic boundary conditions the thermodynamic limit of the Gibbs measure (1.8) yields

$$\mu = \begin{cases} \mu_0 & z < z_c, \\ \frac{1}{2}\mu_+ + \frac{1}{2}\mu_- & z \geq z_c, \end{cases} \tag{1.14}$$
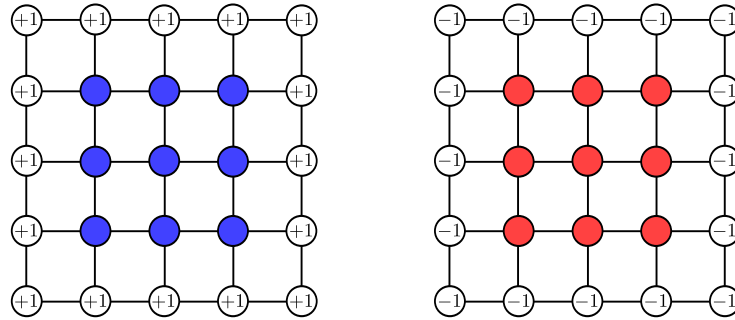
Figure 1.1: Squared lattice with $\pm$ boundary conditions. On the left $\Lambda_{L=5}^{+}$ and on the right $\Lambda_{L=5}^{-}$.

where μ_0 is the unique high-temperature measure and μ_+ and μ_- are the two extremal measures at low temperature [21].

The extremal measures at low temperature are induced by the square lattices Λ_L^+ and Λ_L^- with variables at the boundary fixed to $+1$ and -1 respectively. Figure 1.1 shows the square lattices Λ_L^+ and Λ_L^- for $L = 5$. We describe some property of the pure states μ_+ and μ_-. Denote by $\langle\cdot\rangle$ (periodic boundary condition), $\langle\cdot\rangle_+$ ($+$ boundary condition) and $\langle\cdot\rangle_-$ ($-$ boundary condition) the average with respect to μ, μ_+ and μ_-. The average value of a spin is called the magnetization. Take without loss of generality the spin $\sigma_{i=0}$ at the center of the lattice. The magnetization $m = \langle\sigma_0\rangle$ is always equal to zero. This is because the Gibbs measure (1.8) is symmetric under spin configuration flips (this is valid for periodic boundary conditions)

$$\mu_n\left(\sigma_1, ..., \sigma_n\right) = \mu_n\left(-\sigma_1, ..., -\sigma_n\right). \tag{1.15}$$

The magnetizations $m_+ = \langle\sigma_0\rangle_+$ and $m_- = \langle\sigma_0\rangle_-$ are computed exactly for the Ising model [22]

$$m_\pm = \pm\left(1 - \left(\frac{z^2 - 1}{2z}\right)^4\right)^{1/8}. \tag{1.16}$$

The magnetizations as a function of the rescaled temperature are depicted in Figure 1.2. The values $m_\pm$ are called the spontaneous magnetizations. If the system is perturbed by an arbitrarily small magnetic field, it will be described by the measure μ_+ or μ_-, breaking the symmetry (1.15). This mechanism is responsible for the spontaneous magnetization that occurs in ferromagnetic material. The spontaneous magnetizations informs us about the typical sequences or support of the pure states. The support of the extremal measure μ_+ (resp. μ_-) contains typical spin configurations with positive (resp. negative) average magnetization. At low temperatures, the convex combination (1.14) is thus supported on two "clusters" of spin configurations with mostly $+1$ or -1 values. Whereas at high temperatures, the support of the measure μ_0 is a giant cluster containing typical
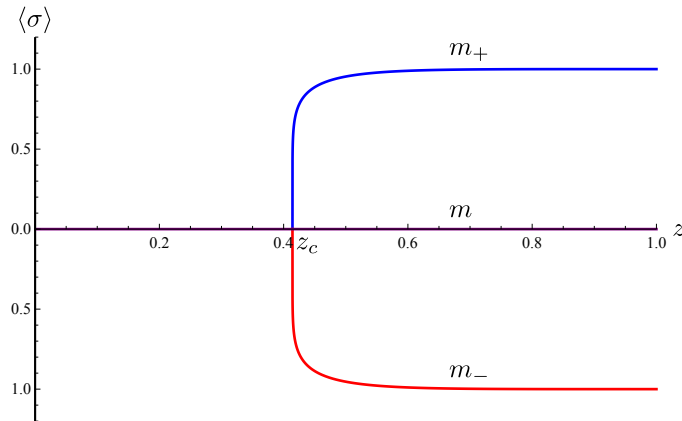
Figure 1.2: Magnetizations for the Ising model. The spontaneous magnetizations m_+ and m_- are shown in blue and red respectively. The magnetization $m = 0$ is shown in purple.
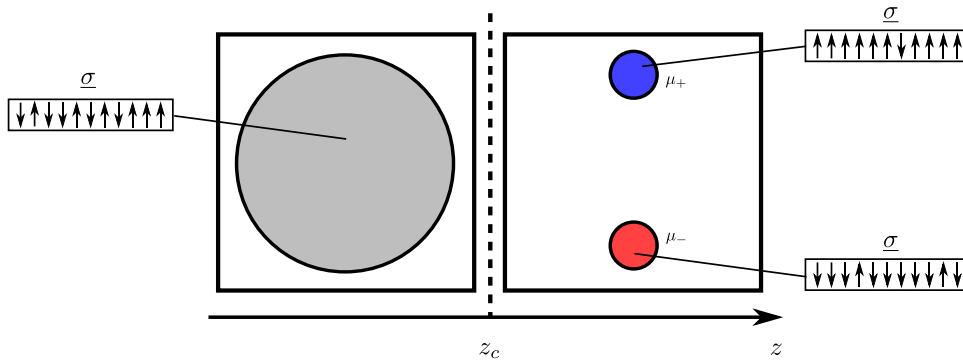
Figure 1.3: Pictorial representation of the phase transition in the Ising model. The boxes represent the state space $\{-1, 1\}^n$ and the circles represent the states on which the Gibbs measure is supported. A typical configuration of each support is depicted. An arrow up (resp. down) represents a $+1$ spin (resp. -1).

sequences with zero average magnetization. The clustering of the measure is depicted in the diagram in Figure 1.3

We make one last remark about the pure states and the clustering in the Ising model. The clustering can also be characterized by point-to-point correlations

$$\mathrm{corr}\,(\sigma_i, \sigma_j) := \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle. \tag{1.17}$$

The point-to-point correlations were computed exactly in 1976 by Wu et al. [23]; but their form is too complicated, hence we simply give the essential property. Away from the phase transition, the point-to-point correlation functions for the pure states μ_0, μ_+

and μ_- are vanishing if spins are far away. That is,

$$\lim_{|i-j|\to\infty} \mathrm{corr}_{0,\pm}\,(\sigma_i,\sigma_j) = 0, \tag{1.18}$$

where $|i-j|$ is the Manhattan distance between two spins. The decay of correlations is even exponential with the distance. However, because of the convex decomposition (1.14), point-to-point correlations for distant spins are non-zero in the low-temperature phase

$$\lim_{|i-j|\to\infty} \mathrm{corr}\,(\sigma_i,\sigma_j) = m_+^2 > 0. \tag{1.19}$$

1.3 Glass Phenomenology

Numerous materials do not possess a crystalline structure, and it is common to find amorphous solids without long-range ordering. A glass is a particular class of amorphous solids, which exhibits a glass transition. In this section, we provide a brief description of the glass phenomenology based on the topical reviews [24, 25, 26].

The material made of sand, lime and soda constituting our windows is one of the best known examples of glass. The method of its manufacture has been known since the beginning of antiquity [27]. We find glass in various aspects of our daily lives, some of which are unexpected. Glasses made of sugars and water is important in cereal-based food processing [28] and in the preservation of vaccines [29]. Glasses are also used for making optical fibers [30] and solar cells [31].

The structure of a glass is similar to that of a liquid, but its mechanical properties are similar to those of a solid. Various methods exist to create glass, but the most common way is by rapidly cooling a melt. This reduces the mobility of the molecules and avoids a possible rearrangement into a crystal. The rate of cooling necessary to reach a glassy state depends strongly on the material considered. For example, it attains the value of 50 K/s for salol, 10^7 K/s for water and more than 10^{10} K/s for silver [32]. During the cooling process the viscosity[1] of the material increases by several order of magnitude. Although this transition is abrupt, it does not involve discontinuous changes in any physical property.

One definition of the glass transition temperature T_g is the temperature at which the shear viscosity reaches 10^{13} Pa·s (in comparison, the viscosity of water is about 10^{-3} Pa·s). T_g depends upon the cooling rate; the slower the cooling rate is, the lower the glass transition temperature is [33]. However, the dependence on the cooling rate is rather weak (in practice, it is observed that T_g is proportional to the logarithm of the cooling rate [34]). The typical relation between the volume and the temperature of a liquid is depicted in Figure 1.4. The glass transition temperature is below the melting point

---

[1]The viscosity measures the resistance of a fluid to a shear stress.
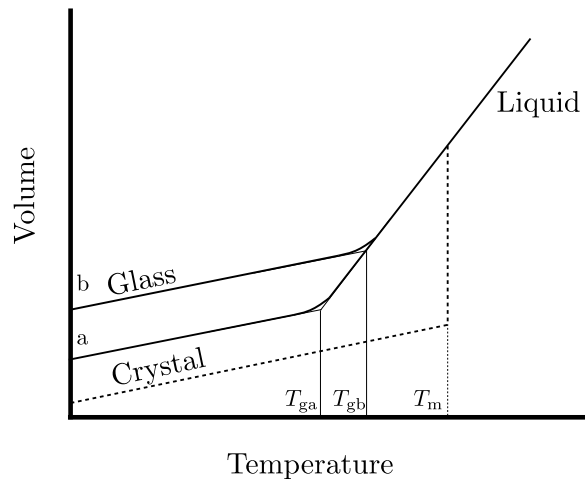
Figure 1.4: Typical relation between the volume and the temperature of glass in the liquid (solid line), glassy (solid lines a, b) and crystalline states (dashed line). T_{m} is the melting temperature, T_{ga} and T_{gb} are the glass transition temperatures corresponding to a slow (a) and a fast (b) cooling rate. Unlike crystallization, vitrification is a smooth transition. (From [35].)

T_{m} (if it exists) and two different cooling scenarios give rise to different glass transition temperatures.

There exists a theoretical lower-limit to the glass transition temperature. If it were possible to obtain a transition at an arbitrarily low temperature, the entropy of the glass would be lower than the entropy of the crystal. Since the entropy of the crystal vanishes when $T \to 0$, this would imply that the glass thus formed would have a negative entropy at low temperature. The entropy being intrinsically a non-negative quantity[2], this would be paradoxical. The lowest temperature before the paradox occurs is called the Kauzmann temperature T_{K} [35]. To avoid an entropy crisis at exactly T_{K}, the material should form an ideal glass with a unique (or at least sub-exponential) configuration. In practice the Kauzmann temperature is not reachable as it would require an infinitely slow cooling rate. The critical point T_{K} is only estimated from extrapolations of the liquid behavior below T_{g}. Thus the existence of an ideal-glass state remains debated [36].

To interpret the phenomenology just described, Goldstein [37] developed a convenient framework that is known today as the landscape paradigm. The potential energy[3] of a material is an n-dimensional function of the configurations of the particles or molecules present in the system. Goldstein assumes that the energy landscape generated by this function provides information about the dynamics of relaxation towards equilibrium. At high temperatures, the particles have sufficiently high kinetic energy to visit any

---

[2]In thermodynamics, the entropy is the logarithm of the states accessible by the system.

[3]The total energy is the sum of the potential energy, depending on the positions of the constituents and the kinetic energy associated with their motion.
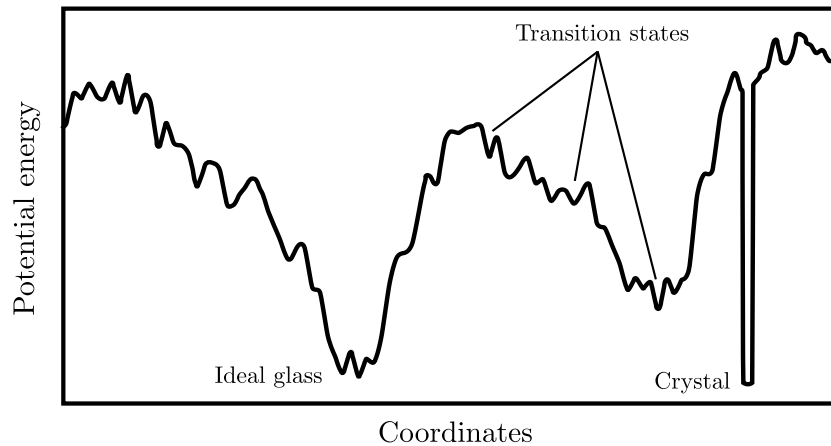
Figure 1.5: Schematic representation of a glass energy landscape. The number of transition states is exponential in the size of the system. A sub-exponential fraction of transition states are deep minima and correspond to ideal glass states. (Adapted from [36].)

configuration. When the temperature is cooled down below the melting point T_{m}, it is likely that the system remains trapped in a local minimum of the potential energy. Due to fluctuations, the system jumps from a local minimum to another in a time that is exponential in the energy barrier between those minima. In non-glassy materials, the system eventually relaxes from these transition states to the crystalline configuration that is a global minimum of the potential. However in glass, the energy landscape is complex and consists of a multitude of local minima. A schematic representation of a complex energy landscape is depicted in Figure 1.5. The transition states are exponentially numerous and they are clustered into basins separated by a high energy barrier. As a consequence of both effects, the relaxation time towards equilibrium becomes exponential in the system size and the system remains trapped in transition states. Glasses are peculiar in that they are off-equilibrium systems with properties constantly evolving very slowly with time.

1.4 Spin Glasses

Like glass, "spin glasses" have a complex energy landscape. This is due to the fact that, unlike the Ising model, spin glasses have little or no explicit symmetry. A system is said to present quenched disorder if some parameters controlling its properties are random and do not evolve with time [38]. The word "quenched" is borrowed from the terminology of the manufacture of alloys. There is a presence of impurities in the metal when it is in the molten state. To obtain a metallic glass, the alloy is subject to a rapid cooling. This procedure, called quenching, freezes the positions of the impurities and prevents the material from entering into an ordered state. Although two quenched workpieces do not have the same global structure, their properties are similar.

We can distinguish between two types of quenched disorder entering in the Gibbs measure (1.2) of a spin glass.

Structural disorder: The graph Γ which shapes the dependencies between the variables is sampled randomly from a prescribed ensemble.

Interaction disorder: The parameters controlling the interactions (the x_a in (1.2)) are random.

The systems that we are studying in this thesis have both structural and interaction disorder. Therefore the Gibbs measures that we consider are random objects.

The average over the disorder is called a quenched average and is denoted by $\mathbb{E}_{\Gamma, \underline{X}}[\cdot]$. In all spin glass models, much attention is given to the calculation of the quenched average of the free energy

$$\overline{f}(\beta) = \lim_{n \to \infty} \mathbb{E}_{\Gamma, \underline{X}}[f(\beta \mid \underline{x})]. \tag{1.20}$$

In general we expect that the free energy is a self-averaging quantity that converges to a non-random object as n grows. Thereby the quenched average (1.20) should capture the typical behavior of the disorder realizations.

We mention below some of the most iconic spin glasses.

Edwards-Anderson (EA) model [39]: The graph is a D dimensional lattice. Like the Ising model, the interactions are between spin pairs $J_{ij}^{\mathrm{EA}} \sigma_i \sigma_j$. However the coupling constants are independent and identically distributed (i.i.d.) random variables that have a Gaussian distribution of mean 0 and variance J^2. The EA model is the most difficult spin glass to analyze and little is known beyond mean-field approximations. As the interaction graph is a regular lattice, we do not expect that mean-field methods (such as the cavity method) are exact for finite D. However, in the asymptotic regime $D \to \infty$, we recover the SK model.

Sherrington-Kirkpatrick (SK) model [1]: The graph is the complete graph $\mathcal{K}_n$ with n vertices, where n is the number of spins. The interactions are also between pairs of spins $J_{ij}^{\mathrm{SK}} \sigma_i \sigma_j$. The coupling constants are i.i.d. distributed according to a Gaussian with mean 0 and variance J^2/n. The $1/n$ scaling factor ensures that the free energy is not trivial in the asymptotic limit as the number of interacting pairs is quadratic in n. So far, it is proven rigorously that the cavity method gives the correct quenched free energy for the SK model [17, 18]. Using similar techniques it is also proven that the quenched free energy of the SK model is self-averaging [40].

Dilute p-spin models: The graph is a random bipartite graph and function nodes are regular with degree p. The interactions involve the product of p spins $J_a \sigma_{a1} \cdots \sigma_{ap}$. The coupling constants are i.i.d. random variables with Gaussian distribution

$\mathcal{N}\left(0, J^2\right)$ or with Bernoulli symmetric distribution, depending on the models considered. For variable nodes with Poisson-distributed random degree and at zero temperature, it is proven that the cavity method predicts the correct free energy. The proof is algorithm-based and exploits the property that at zero temperature the system is identical to XORSAT [41].

The models encountered in coding theory on which this thesis focuses are similar to p-spin models.

1.5 Introduction to the Cavity Method

Computing the free energy is generally a hard task and is a fortiori even harder for glassy systems. Rather than calculating the exact quantity, we can look at an approximation. Mean-field approximations enable us to replace the exhaustive computation of the free energy by solving a system of several equations. The price to pay is that in general those approximations are not accurate when the correlations between distant variables are strong.

The approximation method at the basis of the cavity method is the Bethe-Peierls mean-field theory. First introduced in 1935 by Bethe [42] and extended the following year by Peierls [43], it can be considered as a sophisticated mean-field approximation that goes beyond the slightly older Curie-Weiss theory. Particularly well suited to graphical models that are locally tree-like, it is even conjectured for these graphs that it is asymptotically exact in the presence of short-range correlations. The cavity method is a non-trivial generalization of this approximation to systems that exhibit long-range correlations between variables.

1.5.1 Graphical Models on Trees: the Bethe Free Energy

Consider the Gibbs measure

$$\mu_\beta\left(\underline{u} \mid \underline{x}\right) = \frac{1}{Z\left(\beta \mid \underline{x}\right)} \prod_{a \in C} \psi_a\left(\underline{u}_{\partial a}\right). \tag{1.21}$$

If the graphical model is a tree, there exists an iterative procedure for computing exactly and efficiently the free energy.

We call "message" any probability distribution over a single binary variable. Assign to edge $(ia) \in E$ a pair of messages $\nu_{ia}(u_i)$ and $\widehat{\nu}_{ai}(u_i)$ that satisfies the belief propagation (BP) equations

$$\begin{cases} \nu_{ia}(u_i) = \dfrac{\prod_{b \in \partial i \setminus a} \widehat{\nu}_{bi}(u_i)}{\sum_{s_i} \prod_{b \in \partial i \setminus a} \widehat{\nu}_{bi}(s_i)}. \\[4mm] \widehat{\nu}_{ai}(u_i) = \dfrac{\sum_{u_{\partial a \setminus i}} \psi_a(\underline{u}_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(u_j)}{\sum_{s_{\partial a}} \psi_a(\underline{s}_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}. \end{cases} \qquad (1.22)$$

By convention when $\partial i \setminus a$ (resp. $\partial a \setminus i$) is an empty set, the product over $b \in \partial i \setminus a$ (resp. $j \in \partial a \setminus i$) is equal to one. The BP equations, also called sum-product equations, admit a unique solution on a tree. The iterative version of equations (1.22) is the sum-product algorithm. On trees, we find a solution after t^* iterations, where t^* is the diameter of the tree.

Denote the collection of all messages by $(\underline{\nu}, \widehat{\underline{\nu}})$. The Bethe free energy is a functional of the messages $(\underline{\nu}, \widehat{\underline{\nu}})$ associated with a fixed point of the BP equations (1.22)

$$f^{\text{Bethe}}(\underline{\nu}, \widehat{\underline{\nu}}) = \frac{1}{n} \left(\sum_{a \in C} F_a + \sum_{i \in V} F_i - \sum_{(i,a) \in E} F_{ia} \right), \qquad (1.23)$$

where

$$\begin{cases} F_a = \frac{-1}{\beta} \ln\{\sum_{u_{\partial a}} \psi_a(\underline{u}_{\partial a}) \prod_{j \in \partial a} \nu_{ja}(u_j)\}, \\[4mm] F_i = \frac{-1}{\beta} \ln\{\sum_{u_i} \prod_{b \in \partial i} \widehat{\nu}_{bi}(u_i)\}, \\[4mm] F_{ia} = \frac{-1}{\beta} \ln\{\sum_{u_i} \nu_{ia}(u_i)\widehat{\nu}_{ai}(u_i)\}. \end{cases} \qquad (1.24)$$

It is well known that on trees, the Bethe free energy is equal to the free energy.

Lemma 1.1 (Exactness of the Bethe free energy on trees)**.** *If the graphical model is a tree, then*

$$f = f^{\text{Bethe}}(\underline{\nu}, \widehat{\underline{\nu}}), \qquad (1.25)$$

where $(\underline{\nu}, \widehat{\underline{\nu}})$ satisfies the BP equations (1.22).

A proof of this lemma can be found in [44] or in Section 4.2 as a consequence of the loop sum identity.

1.5.2 Loopy Graphs: Predictions From the Cavity Method

It is tempting to conjecture that Lemma 1.1 is true for locally tree-like graphs, at least in the asymptotic regime $n \to \infty$. This is however *not the case* in general and many systems are known to behave differently such as the XORSAT or K-SAT models [41, 45]. The general picture is much more flavored and interesting.

The cavity method assumes that the random Gibbs distribution (1.2) can, in the limit of $n \to +\infty$, be decomposed into a convex superposition of $\mathcal{N}$ "extremal measures"

$$
\mu_\beta(\underline{u} \mid \underline{x}) = \sum_{p=1}^{\mathcal{N}} w_p \, \mu_{\beta,p}(\underline{u} \mid \underline{x}), \tag{1.26}
$$

each of which occurs with a weight $w_p = e^{-\beta n(f_p - f)}$, where f_p is a free energy associated with the extremal measure $\mu_{\beta,p}$. Notice the similarity between the hypothesis contained in Equation (1.26) and the energy landscape paradigm in Section 1.3. Here pure states play the role of the transition states.

The convex decompositions of the Gibbs distribution into bona fide extremal measures are under mathematical control for "simple" models such as the square lattice Ising model [46]. But for spin glass models, is it not known how to construct or even precisely define the extremal measures. One important conceptual difference with respect to the Ising model, which has a small number of extremal states, is that for spin glasses we envision the possibility of having an exponentially large (in n) number of terms in the decomposition (1.26).

In the context of sparse graphical models it is further assumed that there are "extremal" Bethe measures that are a good proxy for the "extremal measures" discussed above. The Bethe measures are those measures that have marginals given by BP marginals. When the BP equations have many fixed point solutions there are many possible Bethe measures. Heuristically, the extremal ones correspond to the minima of the Bethe free energy[4]. Similarly, it is assumed that the Bethe free energies corresponding to solutions of the BP equations are good proxies for the free energies f_p. Moreover we expect that the free energies f_p are self-averaging quantities.

Since the weights w_p have to sum to 1, we have

$$
e^{-\beta n f} \approx \sum_{p=1}^{\mathcal{N}} e^{-\beta n f_p} \approx e^{-\beta n \min_\varphi (\varphi - \beta^{-1} \Sigma(\varphi; \beta))}, \tag{1.27}
$$

where $e^{n\Sigma(\varphi;\beta)}$ counts the number of extremal states $\mu_{\beta,p}$ with free energy $f_p \approx \varphi$. Once we choose to replace f_p by the Bethe free energies, then the counting function $\Sigma(\varphi; \beta)$ and

---

[4]Remarkably, it is not very important to be able to precisely select the "extremal" ones because at low temperatures we expect that they outnumber the others.

the free energy f can be computed through a fairly technical procedure, and a number of remarkable predictions about the decomposition (1.26) emerge.

The cavity method predicts the existence of two sharply defined thresholds β_{d} and β_{c} at which the nature of the convex decomposition (1.26) changes drastically. Figure 1.6 gives a pictorial view of the transitions associated with the decomposition (1.26).

Replica-symmetric (RS) phase: For $\beta < \beta_{\mathrm{d}}$ the measure $\mu_{\beta}(\underline{u} \mid \underline{x})$ is extremal, in the sense that $\mathcal{N} = 1$ in (1.26). In this phase, the Bethe free energy should be asymptotically exact

$$\lim_{n\to\infty} \mathbb{E}_{\Gamma,\underline{X}}\left[|f - f_{\mathrm{Bethe}}|\right] = 0. \tag{1.28}$$

Dynamical replica-symmetry breaking (d-RSB) phase: For $\beta_{\mathrm{d}} \leq \beta < \beta_{\mathrm{c}}$ the measure is a convex superposition of an exponentially large number of extremal states. The exponent $\varphi - \beta^{-1}\Sigma(\varphi;\beta)$ in (1.27) is minimized at a value $\varphi_{\mathrm{int}}(\beta)$ such that $\Sigma(\varphi_{\mathrm{int}}(\beta);\beta) > 0$. Then the growth rate (as $n \to +\infty$) of the number of extremal states that dominate the convex superposition of pure states (1.26) is defined as

$$\Sigma(\beta) \equiv \Sigma(\varphi_{\mathrm{int}}(\beta);\beta) = \beta(\varphi_{\mathrm{int}}(\beta) - f(\beta)), \tag{1.29}$$

and is strictly positive. This quantity is called the complexity. It appears that the complexity is a decreasing function of β, which becomes equal to zero at β_{c}. In summary, for $\beta \in [\beta_{\mathrm{d}}, \beta_{\mathrm{c}}[$, an exponentially large number of extremal states with the same free energy φ_{int} contribute significantly to the Gibbs distribution. Surprisingly, the convex combination is itself a Bethe measure. Therefore there exists a BP fixed point for which the corresponding Bethe free energy should be exact. In practice this enables us to compute the quenched free energy but not the free energy of a particular instance, as there is no known method to discriminate this particular fixed point from the (exponentially numerous) others.

Static replica-symmetry breaking (s-RSB) phase: For $\beta \geq \beta_{\mathrm{c}}$ the number of extremal states that dominate the measure is finite. It is said that the measure is *condensed* over a small number of extremal states. In fact, there could still be an exponential number of extremal states but they do not contribute significantly to the measure, because their weight is exponentially smaller than the dominant ones. In this phase, the Bethe free energy is not exact and we should use the full machinery of the cavity method (see Chapter 8).

There exists a mathematically more precise definition of β_{d} and β_{c} in terms of correlation functions. When these correlation functions are computed within the framework of the cavity method, the resulting values of β_{d} and β_{c} agree with those given by the complexity curve $\Sigma(\beta)$. In the RS phase and the d-RSB phase, point-to-point correlation functions of type (1.17) decay with respect to the distance between spins. Like the low-temperature
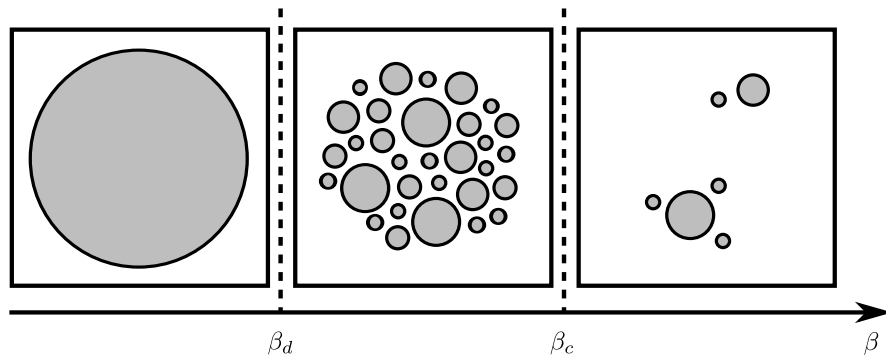
Figure 1.6: Pictorial representation of the decomposition of the Gibbs distribution into a convex superposition of extremal states. Balls represent extremal states (their size represents their internal entropy). For $\beta < \beta_d$ there is one extremal state. For $\beta_d < \beta < \beta_c$ there are exponentially many extremal states (with the same internal free energy φ_{int}). For $\beta > \beta_c$ there is a finite number of extremal states.

case of the Ising model, this is not the case in the s-RSB phase. There exists another type of correlations relevant to the characterization of the d-RSB: the point-to-set correlations. Point-to-set correlations measure how correlated a spin is with respect to any change in the set of spins located at a distance R from it. Although point-to-set correlations decay exponentially fast with R in the RS phase, this is not the case for the d-RSB and s-RSB phases [47, 48].

What is the significance of the transitions at β_d and β_c? The condensation threshold is a thermodynamic phase transition point: the free energy $f(\beta)$ is not analytic at β_c. At β_d the free energy has no singularity: in particular, its *analytical expression* does not change in the whole range $0 < \beta < \beta_c$. At β_d the transition is dynamical: Markov chain Monte Carlo algorithms have an equilibration time that diverges when $\beta \uparrow \beta_d$, and are unable to sample the Gibbs distribution for $\beta > \beta_d$. For more details we refer to [44]. The d-RSB and s-RSB are thus very similar to the phenomenological glass transition discussed in Section 1.3, where β_d and β_c have to be linked with T_g (the glassy transition temperature) and T_K (the Kauzmann transition temperature), respectively.

2 Coding Theory and Sparse Graph Constructions

In this chapter, we express two problems from coding theory in the language of spin glasses.

In Section 2.1, we start by an introduction to the channel coding problem. We present two code constructions based on sparse graphical models, the low-density parity-check codes and the low-density generator-matrix codes. We show that the measure of their error correction performances is linked to the computation of the free energy of a spin glass.

In Section 2.2, we give a brief review of the lossy source coding problem and rate distortion theory. We present an encoding scheme based on low-density generator-matrix codes for binary sources. We express the encoding process as a problem of sampling according to the Gibbs measure of a spin glass. Finally, we present a code construction technique called spatial coupling and describe the corresponding spatially-coupled low-density generator-matrix codes.

2.1 Channel Coding

The transfer of information through a noisy medium is one of the most fundamental topics in information theory. Everyone has experienced talking to a person across the street during a windy day. To be heard, we have to sometimes repeat the same word several times. We can then ask if this way of proceeding is optimal. Is there a minimum amount of information that must be retransmitted in order to be understood? And just as importantly, if not more, how can we efficiently achieve this optimal number of retransmissions? The first question was answered by Shannon [49] in his 1948 paper, thus creating the field of information theory. The second question still stimulates the active research field of channel coding. Since then, many different code constructions were invented. We mention the Hamming codes [50] discovered in 1950 among the early algebraic codes, or more recently the Polar codes that are provable capacity-achieving

codes developed in 2009 by E. Arikan [51].

Our study concerns codes constructed from sparse graphical models, and more specifically low-density parity-check codes (LDPC) and their dual construction, the low-density generator-matrix codes (LDGM). LDPC codes were first developed in Gallager's thesis [52] in 1963 but overlooked because they were too complicated for the technology of the 1960s. The importance of these codes was recognized three decades later, when in 1996 they were rediscovered independently by Neil and McKay [53] and by Sipser and Spielman [54]. LDPC codes perform very well under distributed low-complexity decoding algorithms and have been proven to achieve capacity over binary erasure channels (BEC) [55]. Since 1996, they have been integrated into standards for wireless data transmissions, computer chips and commercial cell phones.

In this section, we start by reviewing the problem of communicating through a noisy memoryless channel. Then we follow with an overview of the LDPC and LDGM code construction and how to measure the quality of their error correction through the conditional entropy of the source given the received message. We finish by showing that the computation of the conditional entropy of an LDPC or an LDGM code is tantamount to computing the free energy of a spin glass system.

According to the prediction of the cavity method, the Bethe free energy of an LDGM or an LDPC code should be equal to its the free energy. In this thesis we prove that the prediction of the cavity method is exact over binary-symmetric channels (BSC) at high noise. For this we use a tool from statistical physics called the polymer expansion.

Let us briefly comment on the connections of this work with other recent approaches. For the class of graphical models that describe communication with LDPC and LDGM codes over BSC we have ample evidence that the replica-symmetric solution[1] is exact. Bounds between the replica-symmetric and true free energy were derived in [56, 57, 58], and for the special case of the binary erasure channel equality was proved in [59, 60]. These results are based on specific methods such as combinatorial calculations for the binary erasure channel, and the interpolation method for the bounds on general channels. In [61] a more generic approach is taken based on cluster expansions combined with duality. The cluster expansions used in [61] are sophisticated forms of polymer expansions. It is proven that correlations between pairs of distant (with respect to graph distance) bits decay exponentially fast for LDGM codes in the regime of large noise, and LDPC codes in the regime of small noise. This has also allowed us to conclude that the replica-symmetric formulas are exact in these regimes for general binary-symmetric memoryless channels. A case where the cluster expansions of [61] does not work is that of LDPC codes over BSC in the regime of large noise, which is the case considered in this thesis.

---

[1]Replica-symmetric formulas are quenched averaged forms of the Bethe formulas.
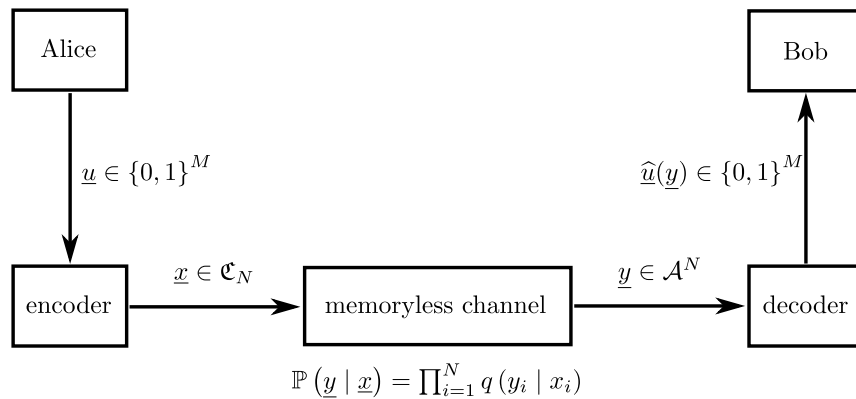
Figure 2.1: Communication through a noisy channel

2.1.1 Communication Through Noisy Memoryless Channels

Alice wants to send messages to Bob. The source of the messages is modeled as a random variable $\underline{U}$ which generates strings of M bits $\underline{u} \in \{0,1\}^M$ uniformly at random. To protect the messages against interferences, Alice adds redundancy. This procedure is called the encoding process and can be viewed as a one-to-one mapping from $\underline{u}$ to a longer string of bits $\underline{x} \in \{0,1\}^N$, where $N \geq M$. The message with redundancy is called a codeword, and the set of codewords is called the codebook $\mathfrak{C}_N$. The measure of the redundancy is the rate. It is defined as the ratio between the number of bits M that carry the information and the number of bits N that compose a codeword.

$$R = \frac{M}{N}. \tag{2.1}$$

Alice sends a codeword $\underline{x} \in \mathfrak{C}_N$ to Bob through a noisy memoryless channel. The string of bits that Bob receives, denoted by $\underline{y} \in \mathcal{A}^N$, is a corrupted version of the codeword. The corrupted codeword does not necessarily belong to a binary alphabet. For instance, if we are in the presence of erasures, the alphabet is ternary $\mathcal{A} = \{0, 1, \text{erasure}\}$. Once Bob receives the corrupted codeword of Alice, he will decode it by constructing an estimate of the codeword $\widehat{\underline{x}}\left(\underline{y}\right)$ or of the original message $\widehat{\underline{u}}\left(\underline{y}\right)$. The communication process is summarized in Figure 2.1.

When we say that the channel is memoryless, we mean that the probability to receive the particular bit y_i in $\underline{y}$ depends only on x_i in $\underline{x}$. The term "memoryless" refers to the fact that if bits are sent one after the other, the state of the channel (or the noise realization) is independent of its previous state. Thus the probability that Bob receives $\underline{y}$ takes the simple form of a product

$$\mathbb{P}_{\underline{Y}|\underline{X}}\left(\underline{y} \mid \underline{x}\right) = \prod_{i=1}^{N} q\left(y_i \mid x_i\right), \tag{2.2}$$
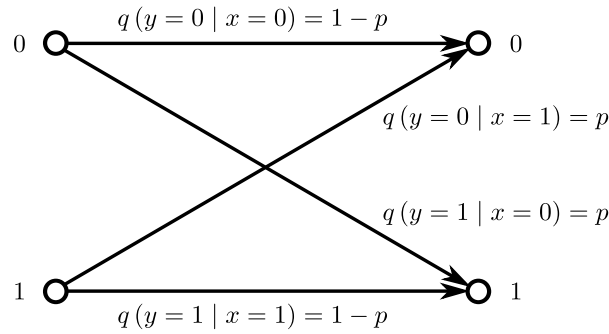
Figure 2.2: Transition probability of the BSC (p) channel.

where $q\,(y \mid x)$ is a transition probability that characterizes the channel completely. The transition probabilities of a channel are traditionally represented by a diagram. For example, the transition probabilities of the BSC are depicted in Figure 2.2. The BSC is a model for channels that corrupt bits by transforming a zero into a one or vice versa with an equal flipping probability p.

For a given pair $(\underline{u}, \underline{\widehat{u}})$, we measure the fraction of incorrectly decoded bits by the relative Hamming distance

$$d_M\,(\underline{u}, \underline{\widehat{u}}) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}\,(u_i \neq \widehat{u}_i)\,, \qquad (2.3)$$

where $\mathbb{I}$ denotes the indicator function. The quality of the reconstruction is measured by the probability of block code error

$$P_M = \mathbb{P}_{\underline{U},Y}[\underline{u} \neq \underline{\widehat{u}}], \qquad (2.4)$$

and by the average bit-error probability

$$D_M = \mathbb{E}_{\underline{U},Y}[d_M\left(\underline{u}, \underline{\widehat{u}}\left(\underline{y}\right)\right)]. \qquad (2.5)$$

In the present case the average bit-error probability is equal to the average distortion.

Alice decides to send messages with a rate R and she tolerates a fraction of bit-error D. She asks if a sequence of encoding schemes exists with the prescribed requirement that $\lim_{M\to\infty} D_M = D$. This question is answered by the famous Shannon theorems. The version presented below is transcribed from MacKay's book [62, p. 162].

Theorem 2.1 (Noisy-Channel theorem). *Define the capacity of a binary input memoryless channel as the maximum of the mutual information over all probability distributions for $X \in \{0,1\}$*

$$C = \max_{\mathbb{P}_X} I\,(X;Y)\,. \qquad (2.6)$$
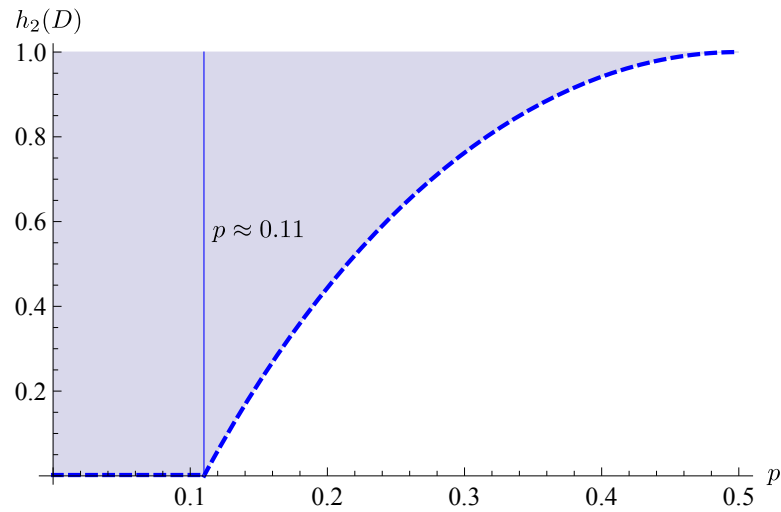
Figure 2.3: The achievable coding region for the BSC (p) with a rate $R = 1/2$ is shown in blue. For $p \leq 0.11$, it is possible to achieve a vanishing block code error.

1. If $R < C$, there exist codes and decoders achieving $\lim_{M \to \infty} P_M = 0$.

2. If $R \geq C$ and $h_2(D) \geq 1 - \frac{C}{R}$, there exist codes and decoders achieving $\lim_{M \to \infty} D_M = D$.

3. For any D, rates higher than $\frac{C}{1 - h_2(D)}$ are not achievable.

In Theorem 2.1 we introduced the binary entropy function

$$h_2(p) = -p \log_2 p - (1 - p) \log_2 (1 - p). \tag{2.7}$$

To illustrate the theorem, we take an example with $R = \frac{1}{2}$ and $D = 0$ on the binary symmetric channel. The capacity of the BSC(p) is

$$C_{\text{BSC}}(p) = 1 - h_2(p). \tag{2.8}$$

For flipping probability less than $p \approx 0.11$, the rate is smaller than the capacity and a vanishing block code error probability is achievable. For higher flipping probabilities $p \gtrsim 0.11$, the average bit-error probability can be at best $D_{\text{Sh}} = h_2^{-1} \left(1 - \frac{1 - h_2(p)}{R} \right)$ as depicted in Figure 2.3.

2.1.2 Codes Based on Sparse Graphical Models

LDGM and LDPC codes are linear codes constructed from sparse matrices. A common advantage of both types of codes is that they can be described in terms of sparse factor graphs. The decoding process for these codes tantamount to an optimization problem on

a graphical model.

LDGM Codes

A low-density generator-matrix code is defined by its $M \times N$ generator-matrix G, where G is created randomly. The fraction of rows with row weight[2] s, $1 \leq s \leq l_{\max}$, is $\Lambda_s \geq 0$, and the fraction of columns with column weights t, $1 \leq t \leq r_{\max}$, is $P_t \geq 0$. It is easy to see that $\sum_{s=1}^{l_{\max}} \Lambda_s = \sum_{t=1}^{r_{\max}} P_t = 1$ and the average row and column weight fractions $l = \sum_{s=1}^{l_{\max}} \Lambda_s s$, $r = \sum_{t=1}^{r_{\max}} P_t t$ satisfy

$$Ml = Nr. \tag{2.9}$$

It is convenient to use the generating functions $\Lambda(z) = \sum_{s=1}^{l_{\max}} \Lambda_s z^s$, $P(z) = \sum_{t=1}^{r_{\max}} P_t z^t$ to refer to the fractions of rows and columns weights. The ensemble of generator-matrices with the same fractions (Λ, P) is denoted by $\mathrm{LDGM}(\Lambda, P, M)$. A generator-matrix is selected uniformly at random[3] from the ensemble $\mathrm{LDGM}(\Lambda, P, M)$. For example, the following matrix belongs to the ensemble $\mathrm{LDGM}(z^4, z^2, 3)$

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{2.10}$$

Note that in regime of interest $M \to \infty$, the matrix G becomes sparse. The design rate of the code is by definition $R_{\mathrm{LDGM}} = \frac{r}{l}$.

Given a message $\underline{u}$, the corresponding codeword is simply

$$\underline{x} = \underline{u}G, \tag{2.11}$$

where the matrix product is performed in $\mathbb{F}_2$. The optimal block decoder is the maximum a posteriori (MAP) decoder

$$\widehat{\underline{u}}^{\mathrm{MAP}} = \arg \max_{\underline{u}} \mathbb{P}_{\underline{U}|\underline{Y}} \left(\underline{u} \mid \underline{y} \right), \tag{2.12}$$

and the optimal bit decoder is the bit-MAP decoder

$$\widehat{u}_i^{\mathrm{MAP}} = \arg \max_{u_i} \mathbb{P}_{U_i|\underline{Y}} \left(u_i \mid \underline{y} \right). \tag{2.13}$$

The block MAP decoder (2.12) minimizes the probability of block error (2.4) and the bit MAP decoder (2.13) minimizes the average bit-error probability (2.5).

---

[2]The weight of a row or column is the number of non-zero elements that it contains.

[3]This ensemble is not practical to generate in practice. One uses slightly different ensembles that have the same asymptotic properties. For details we refer to [59].
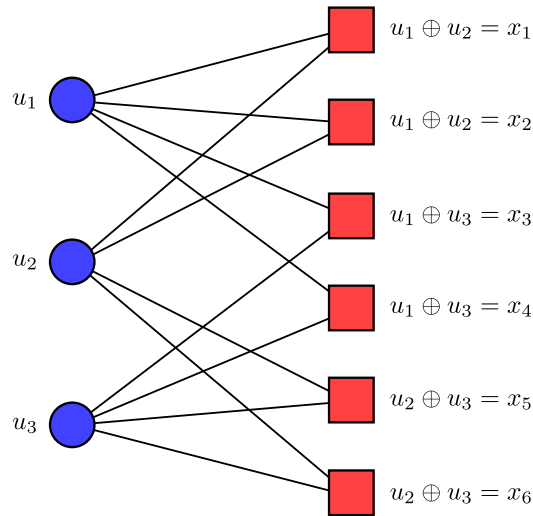
Figure 2.4: Graph of LDGM code corresponding to the matrix (2.10).

The conditional probability distribution and its marginal which appear in (2.12) and (2.13) are Gibbs distributions. Using Bayes rules and the memoryless property of the channel (2.2), we obtain

$$\mathbb{P}_{\underline{U}|\underline{Y}}\left(\underline{u} \mid \underline{y}\right) \propto \prod_{a=1}^{N} \left(\frac{q\left(y_a \mid 0\right)}{q\left(y_a \mid 1\right)}\right)^{\frac{(-1)^{x_a(\underline{u}_{\partial a})}}{2}}. \tag{2.14}$$

The underlying graphical model is constructed by taking the number of variables $n = M$, the number of check nodes $m = N$ and drawing an edge (i, a) iff $G_{ia} = 1$. Figure 2.4 shows the graph corresponding to the generator-matrix G in (2.10). The probability (2.14) is the following Gibbs measure

$$\mathbb{P}_{\underline{U}|\underline{Y}}\left(\underline{u} \mid \underline{h}\right) = \frac{1}{Z_{\text{LDGM}}} \prod_{a \in C} e^{h_a \prod_{i \in \partial a}(-1)^{u_i}}, \tag{2.15}$$

where the normalization factor

$$Z_{\text{LDGM}} = \sum_{\underline{u} \in \{0,1\}^n} \prod_{a \in C} e^{h_a \prod_{i \in \partial a}(-1)^{u_i}}, \tag{2.16}$$

is the partition function of the LDGM measure. The quantities h_a which appear in the measure (2.15) are the half log-likelihood ratios of the channel observations y_a

$$h_a = \frac{1}{2} \ln \frac{q\left(y_a \mid 0\right)}{q\left(y_a \mid 1\right)}. \tag{2.17}$$

The LDGM Gibbs measure (2.15) is that of a disordered system. The factor graph Γ is selected uniformly at random among the ensemble of bipartite graphs with fractions

of variable nodes with degree s equal to Λ_s and with fractions of function nodes with degree t equal to P_t. The channel observations h_a are also randomly distributed. For the BSC (p), as the code is linear and the optimal decoder is symmetric, we can assume without loss of generality that the all-zero codeword is sent (see Appendix A), which implies that the probability distribution of h_a is

$$q\left(h_a\right) = (1-p)\,\delta\left(h_a - h\right) + p\,\delta\left(h_a + h\right),\tag{2.18}$$

where

$$h = \frac{1}{2}\ln\frac{1-p}{p}.\tag{2.19}$$

The type of interaction in (2.15) is similar to the magnetic interactions discussed in the Ising model section. The difference is that the coupling constant could be negative or positive, and the number of variables (or spins) involved are in general more than two. The intensity of the half log-likelihood ratios (2.19) is viewed as the temperature of the system.

LDPC Codes

A low-density parity-check code (LDPC) is a dual construction to the LDGM code. An LDPC code from the ensemble LDPC(Λ, P, N) is generated from an $(N - M) \times N$ parity-check matrix H, where the fraction of row weights (respectively, columns weights) is prescribed by $\Lambda\left(z\right)$ (respectively $P\left(z\right)$). For instance, the following matrix belongs to the ensemble LDPC$(z^2, z^4, 3)$

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.\tag{2.20}$$

If and only if a sequence $\underline{x}$ is a codeword, it satisfies

$$H\underline{x}^T = 0,$$

and the codebook is then the subspace $\mathfrak{C}_N = \left\{\underline{x} \in \{0,1\}^N \mid H\underline{x}^T = 0\right\}$. The design rate of the code is $R_{\text{LDPC}} = 1 - \frac{l}{r}$. The optimal block and bit decoder are also given by their respective maximum a posteriori (MAP) decoder

$$\widehat{\underline{x}} = \arg\max_{\underline{x}} \mathbb{P}_{\underline{X}|\underline{Y}}\left(\underline{x} \mid \underline{y}\right),\tag{2.21}$$

and

$$\widehat{x}_i = \arg\max_{x_i} \mathbb{P}_{X_i|\underline{Y}}\left(x_i \mid \underline{y}\right).\tag{2.22}$$

Using the memoryless property of the channel, the conditional probability distribution
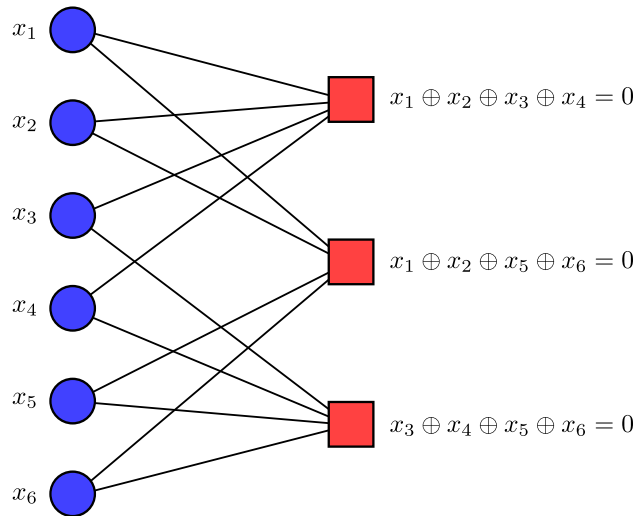
Figure 2.5: The graph associated with the parity-check matrix (2.20).

in (2.21) and (2.22) can be factorized in a Gibbs measure

$$\mathbb{P}_{\underline{X}|\underline{Y}}\left(\underline{x} \mid \underline{y}\right) \propto \prod_{a=1}^{M} \mathbb{I}\left(\left(H\underline{x}^T\right)_a = 0\right) \prod_{i=1}^{N} \left(\frac{q\left(y_i \mid 0\right)}{q\left(y_i \mid 1\right)}\right)^{\frac{(-1)^{x_i}}{2}}. \qquad (2.23)$$

To construct the underlying graphical model, we take the number of variables to be $n = N$, the number of check nodes to be $m = N - M$, and we draw an edge (i, a) iff $H_{ia}^T = 1$. Figure 2.5 shows the graph corresponding to the parity-check matrix H in (2.20).

The conditional probability (2.23) becomes

$$\mathbb{P}_{\underline{X}|\underline{Y}}\left(\underline{x} \mid \underline{h}\right) = \frac{1}{Z_{\text{LDPC}}} \prod_{a \in C} \left(1 + \prod_{i \in \partial a} (-1)^{x_i}\right) \prod_{i \in V} e^{h_i(-1)^{x_i}}, \qquad (2.24)$$

where the normalization factor

$$Z_{\text{LDPC}} = \sum_{\underline{x} \in \{0,1\}^n} \prod_{a \in C} \left(1 + \prod_{i \in \partial a} (-1)^{x_i}\right) \prod_{i \in V} e^{h_i(-1)^{x_i}}, \qquad (2.25)$$

is the partition function of the LDPC code. Similarly to LDGM codes, the quantities h_i are the half log-likelihood ratios of the channel observations y_i

$$h_i\left(y_i\right) = \frac{1}{2} \ln \frac{q\left(y_i \mid 0\right)}{q\left(y_i \mid 1\right)}. \qquad (2.26)$$

The Gibbs measure (2.24) is a measure of a disordered system. Similarly to the LDGM

case, the factor graph Γ is selected uniformly at random among the ensemble of bipartite graphs with fractions of variable nodes with degree s equal to Λ_s and with fractions of function nodes with degree t equal to P_t. Under the all-zero codeword assumption for the BSC (p), we recover the distribution of half log-likelihood ratios

$$q\left(h_i\right) = \left(1 - p\right)\delta\left(h_i - h\right) + p\delta\left(h_i + h\right),\tag{2.27}$$

where h is the amplitude of the half log-likelihood ratios given by (2.19).

The LDPC Gibbs measure is, however, very different from the LDGM measure (2.15) in that it contains parity-check constraints. These interaction functions are "hard" in the sense that they are equal to zero for certain variable configurations (when $\underline{x} \notin \mathfrak{C}$). The amplitude $h = \frac{1}{2}\ln\frac{1-p}{p}$ cannot be interpreted as a temperature for LDPC codes. The LDPC Gibbs measure can in fact be viewed as a spin glass measure at zero temperature (infinite energy cost if $\underline{x} \notin \mathfrak{C}$) with a ground state degeneracy (there is an exponential number of codewords).

2.1.3 The Conditional Entropy as a Measure of Optimal Performance

We want to measure the behavior of the average bit-error probability (2.5) of a code as the channel parameter changes. The performance of the code is lower-bounded by the optimal bit-MAP decoder (2.13) for LDGM codes and (2.22) for LDPC codes. The average bit-error probability D_M^{MAP} of the MAP decoder is a quantity difficult to analyze and control. Fano's inequality [63] provides a lower bound on the error in terms of the conditional entropy

$$h_2\left(D_M^{\text{MAP}}\right) \geq \frac{1}{M}H\left(\underline{U} \mid \underline{Y}\right) = \frac{1}{RN}H\left(\underline{X} \mid \underline{Y}\right).\tag{2.28}$$

The averaged conditional entropy over the LDPC $\left(z^3, z^6\right)$ ensemble is depicted in Figure 2.6. The Shannon lower bound on the entropy of the average error probability is shown for comparison. The smallest value of p, for which the conditional entropy is non-zero, is called the MAP threshold p_{MAP} and according to Fano's inequality $D_M^{\text{MAP}} > 0$ for $p > p_{\text{MAP}}$. The conditional entropy is trivial at low noise and is non-analytic at $p_{\text{MAP}} \approx 0.101$. The picture for LDGM codes is slightly different at low noise, because these codes exhibit an error floor $h_2\left(D_M^{\text{MAP}}\right) > \frac{1}{M}H\left(\underline{U} \mid \underline{Y}\right) > 0$. This issue can be fixed, however, by adding a precoder, thus making LDGM codes still useful in practice.

If we define the free energy for the LDGM and LDPC codes, respectively, as

$$f_{\text{LDGM}} = \frac{1}{n}\ln Z_{\text{LDGM}} \text{ and } f_{\text{LDPC}} = \frac{1}{n}\ln Z_{\text{LDPC}},\tag{2.29}$$

then the conditional entropy on the BSC (p) channel is directly related to the free energy
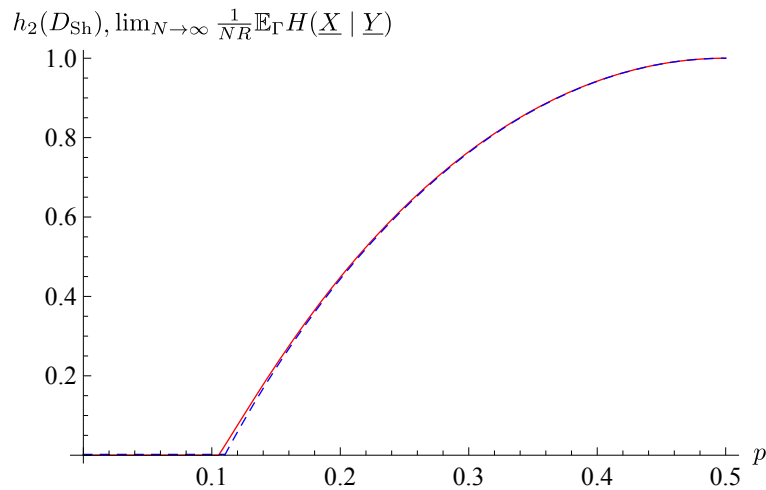
Figure 2.6: Conditional entropy and Shannon bound on the BSC (p). The average conditional entropy for the LDPC(z^3, z^6) ensemble is shown in red. Its threshold is $p_{\text{MAP}} \approx 0.101$. The optimal Shannon bound is shown in dashed blue and its threshold is $p_{\text{Sh}} \approx 0.110$. (Adapted from [59]).

of the codes through the formulas (see [64] for a derivation)

$$\frac{1}{M} H_{\text{LDGM}} \left(\underline{U}|\underline{Y}\right) = \mathbb{E}_{\underline{h}} \left[f_{\text{LDGM}}\right] - \frac{l}{r} \frac{1 - 2p}{2} \ln \frac{1 - p}{p}, \tag{2.30}$$

and

$$\frac{1}{N} H_{\text{LDPC}} \left(\underline{X}|\underline{Y}\right) = \mathbb{E}_{\underline{h}} [f_{\text{LDPC}}] - \frac{1 - 2p}{2} \ln \frac{1 - p}{p}. \tag{2.31}$$

Therefore the analysis of the performance of LDGM or LDPC codes under MAP decoding is tantamount to the computation of the free energy of spin glasses.

2.2 Lossy Source Coding

Lossy source coding is one the oldest and most fundamental problems in communications. It was originally created to answer the following question: Up to which fidelity can we reconstruct a source of information if we only have a limited amount of memory to describe it? The measure of fidelity depends on the application considered. If the source is a picture or a song, the fidelity will depend on the human perception and its aesthetic sense. The concept of lossy compression gave rise to numerous efficient encoding algorithms, notably for movies, music or photos. An example: a Beethoven symphony takes about three times less memory space with an MP3 encoder than with a lossless sound compression algorithm, such as FLAC, and this is without a perceptible difference of quality for the human ear.

The rate-distortion theory was created by Shannon in 1948. Far from the technical complexity of current lossy-encoding techniques, it nevertheless poses a solid theoretical foundation. Being a dual problem of channel coding, the rate-distortion theory can be explained within a common framework. The fidelity of the reconstruction is measured by the distortion that is identical to the average error probability encountered previously in channel coding. Thus lossy compression can be viewed as communication through a perfect channel but with codewords shorter than the original message. Therefore Shannon's theorems for channel coding apply, characterizing the achievable region of distortion versus a compression-rate diagram. The question of an explicit construction of an efficient encoding/decoding scheme remains unanswered by the theorem and forms an active research area.

It has been known since 1963 [65] that linear codes can achieve the Shannon rate-distortion bound. However, it is of fundamental importance to find encoding/decoding schemes with a low computational complexity. One of the first attempts is the work by Viterbi and Omura [66] who designed in 1974 trellis codes approaching the rate-distortion bound for memoryless sources and bounded distortion measures. Although linear in the number of encoded bits, the complexity increases exponentially as we approach the distortion bound. In the past few years, different techniques emerged for designing efficient algorithms. We mention the concatenation of small codes under optimal encoding [67], and also the polar codes that achieve the rate-distortion bound with almost linear complexity [68].

It has been shown that LDGM codes equipped with low-complexity message-passing algorithms approach the rate-distortion curve for optimized degree distributions (Λ, P) [69, 70, 71]. However, these results are empirical and there is no real principle for the choice of the degree distribution.

The construction investigated in this thesis is based on simple spatially-coupled LDGM codes. Spatially-coupled codes were first introduced in the context of channel coding in the form of convolutional LDPC codes [72] and it is now well established that the performance of such ensembles under BP decoding is consistently better than the performance of the underlying ensembles [73, 74, 75]. This is also true for coupled LDGM ensembles in the context of rateless codes [76]. The key observation is that the BP threshold of a coupled ensemble saturates towards the maximum a posteriori MAP threshold of the underlying ensemble as the width of the coupling window grows. A proof of this *threshold saturation phenomenon* is in [77, 78]. An important consequence is that spatially-coupled regular LDPC codes with large degrees universally achieve capacity. Recently, more intuitive proofs based on replica-symmetric energy functionals have been given in [79, 80]. Spatial coupling has also been investigated beyond coding theory in other models such as the Curie-Weiss chain, random constraint satisfaction problems [81, 82, 83], and compressed sensing [84, 7, 85].

Our motivation to study spatially-coupled LDGM codes in lossy source coding is two-fold.

First we want to know if the spatially-coupled version of an LDGM code can approach the rate-distortion bound under low-complexity message-passing decoding when the underlying code cannot. Second, we want to assess if, as in channel coding, a threshold saturation phenomenon occurs and, if so, what is the specific nature of such thresholds.

Our focus is on LDGM codes with Poisson-distributed degrees for function nodes and regular degrees for variable nodes. This degree distribution is known to achieve the rate-distortion bound under optimal decoding [69] but shows poor performance under message-passing decoding. A similar conclusion is reached (by non-rigorous means) by the cavity [86] and replica [14] methods from statistical physics.

In this section, we start by reviewing the basics of the rate-distortion theory. Subsequently, we present the way in which an LDGM construction can be used as a lossy source encoding /decoding scheme. Finally, we introduce the concept of spatial coupling and the related code construction.

2.2.1 The Rate-Distortion Theory

Consider a source of information $\underline{X}$ generating sequences of N bits $\underline{x} \in \{0,1\}^N$. Without loss of generality, we consider that this source has maximal entropy $H(\underline{X}) = N$. Thus X_i are i.i.d. Bernoulli 1/2 random variables. We want to encode this information into a smaller sequence that possesses only $M < N$ bits. We have then to compress an information sequence to a smaller string of bits $\underline{u} \in \{0,1\}^M$. The reproduction of the source based on the stored sequence is denoted by $\widehat{\underline{X}}$. The measure of fidelity is simply the Hamming distance between the original signal and its reproduction

$$d_N\left(\underline{x}, \widehat{\underline{x}}\right) = \frac{1}{N} \sum_{i=1}^{N} |x_i - \widehat{x}_i| . \tag{2.32}$$

The distortion is the average of the Hamming distance (2.32) over the source realizations and over the encoding/decoding realization[4]

$$D_N = \mathbb{E}_{\underline{X}, \widehat{\underline{X}}} \left[d_N\left(\underline{x}, \widehat{\underline{x}}\right) \right] . \tag{2.33}$$

The inverse ratio between the length of the source and its representation in the memory measures the compression rate

$$R = \frac{M}{N} . \tag{2.34}$$

The problem of determining, for a given rate, what is the achievable distortion is then isomorphic to a channel coding problem. Lossy source coding can be viewed as communication through a noiseless channel at a rate R^{-1}. The distortion (2.33) can be, in this context, interpreted as the average bit-error probability (2.5). The following is

---

[4]The encoding/decoding scheme can be a probabilistic algorithm.
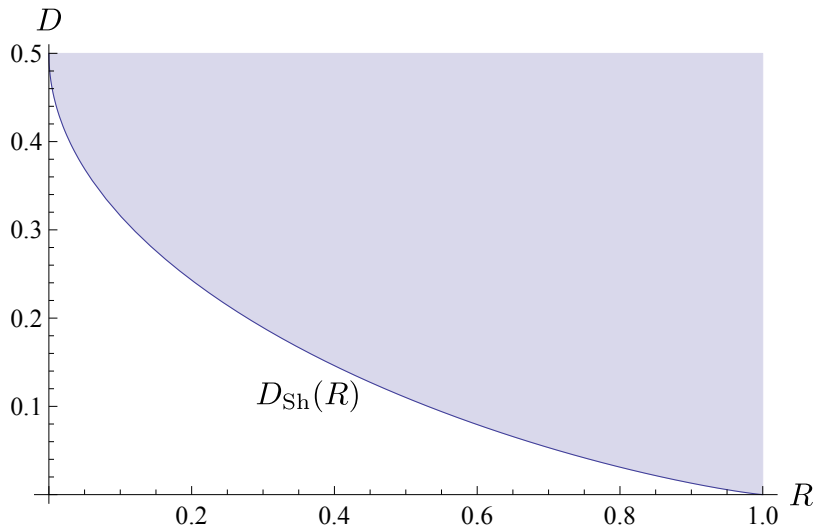
Figure 2.7: Shannon rate-distortion curve and the achievable region.

thus a consequence of Theorem 2.1 (see [62, p. 167]).

Theorem 2.2 (Lossy source coding theorem)**.** *Define the Shannon rate-distortion curve as*

$$D_{\mathrm{Sh}}(R) = h_2^{-1}(1 - R). \tag{2.35}$$

1. *If $D \geq D_{\mathrm{Sh}}(R)$, there exists codes achieving $\lim_{N \to \infty} D_N = D$.*

2. *For any R, distortions lower than $D_{\mathrm{Sh}}(R)$ are not achievable.*

The rate-distortion curve is shown in Figure 2.7. The blue area represents the achievable region.

2.2.2 LDGM as a Lossy Source Compression Scheme

Let G be an $M \times N$ generator-matrix from an LDGM (Λ, P, M) ensemble. We start by describing the decoding process. The reconstructed message is defined as the product of the compressed sequence $\underline{u}$ and the generator-matrix

$$\widehat{\underline{x}}(\underline{u}) = \underline{u}G. \tag{2.36}$$

As the matrix G contains only Nl non-zero entries, decoding is an operation of linear complexity. The difficulty comes from the encoding process. The optimal encoder finds a compressed sequence $\underline{u}$ such that the resulting distortion is minimized

$$\underline{u}^{\mathrm{opt}} = \arg \min_{\underline{u}} d(\underline{x}, \widehat{\underline{x}}(\underline{u})). \tag{2.37}$$
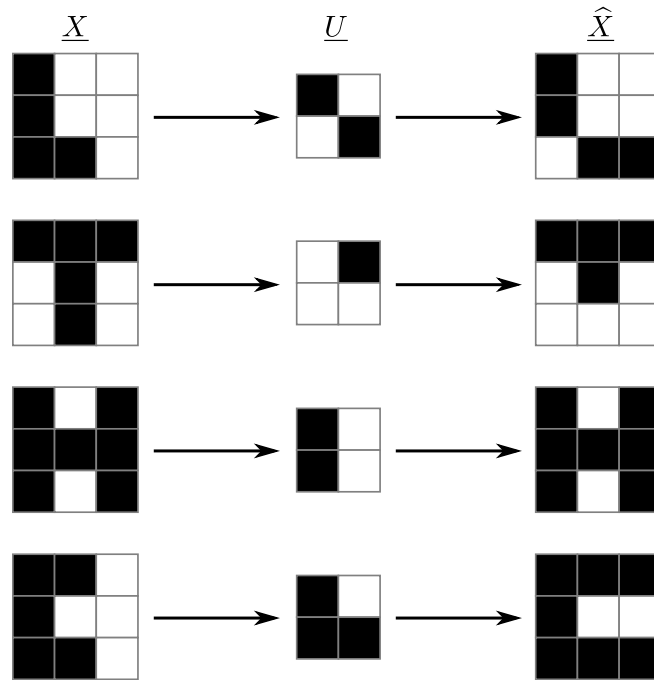
Figure 2.8: Coding of 9-bit images representing the word "LTHC" into 4-bit images. This code is based on the matrix (2.38).

So far, no polynomial complexity algorithms that can solve the optimization problem (2.37) are known. In fact for the regular distribution $P = z^l$, this problem turns out to be MAX-l-XORSAT, which is known to be NP-hard [87, p. 366], thus leaving little chance that such algorithms exist.

An example of an LDGM encoding/decoding process is depicted in Figure 2.8. The 3×3-bit images representing the word "LTHC" are compressed to 2×2-bit images, then recovered with some distortion. The images are transformed into a string of bits starting with the pixel in the upper left corner and going from left to right and from top to bottom. The matrix used to realize this example is the following:

$$
G_{\text{example}} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}. \tag{2.38}
$$

We use the optimal encoding scheme (2.37), which requires looking through 2^4 configurations. For compressed images of size 16×16, the number of configurations would exceed the number of atoms in the universe.

2.2.3   Decoding by Sampling: BP Guided Decimation

The optimization problem (2.37) can be formulated in terms of a graphical model. We consider $n = M$ variable nodes and $m = N$ check nodes, and we draw an edge (i, a) iff $G_{ia} = 1$. On this graph we define the Gibbs measure

$$
\begin{aligned}
\mu_\beta\left(\underline{u} \mid \underline{x}\right) &= \frac{1}{Z_\beta(\underline{x})} e^{-2\beta N d_N\left(\underline{x}, \widehat{\underline{x}}(\underline{u})\right)} \\
&= \frac{1}{Z_\beta(\underline{x})} \prod_{a \in C} e^{-2\beta|x_a - \bigoplus_{i \in \partial a} u_i|},
\end{aligned}
\tag{2.39}
$$

where $\beta > 0$, the inverse temperature, is an open parameter. The normalization factor in Equation (2.39) is the partition function

$$
Z_\beta\left(\underline{x}\right) = \sum_{\underline{u} \in \{0,1\}^n} \prod_{a \in C} e^{-2\beta|x_a - \bigoplus_{i \in \partial a} u_i|}.
\tag{2.40}
$$

Suppose for the moment that we are able to sample a configuration $\underline{u}^*$ according to the distribution (2.39). As β increases, the resulting distortion becomes lower on average and in the limit we have

$$
\lim_{\beta \to \infty} d\left(\underline{x}, \widehat{\underline{x}}(\underline{u}^*)\right) = \min_{\underline{u}} d\left(\underline{x}, \widehat{\underline{x}}(\underline{u})\right).
\tag{2.41}
$$

Sampling a configuration for a value of β large enough is equivalent to solving the optimization problem (2.37).

We can design a sampling algorithm if we can compute the free energy of a system. We recall that the perturbation of the free energy gives us access to the expectation of bits and thus to their marginal distributions (see Eqn. (1.6)). By the Bayes rule:

$$
\mu_\beta(\underline{u} \mid \underline{x}) = \prod_{i=1}^{n} \mu_\beta(u_i|\underline{x}, u_1, \ldots, u_{i-1}).
\tag{2.42}
$$

Thus we can sample $\underline{u}$ by first sampling u_1 from $\mu_\beta(u_1|\underline{x})$, then u_2 from $\mu_\beta(u_2|\underline{x}, u_1)$ and so on.

The sampling procedure is summarized by the Algorithm 2.1. This algorithm fixes a variable at every step and then removes it from the variable set. Such a process is called decimation. The graphical model thus changes after each iteration, but it is not hard to see that it remains in the same form (2.39) albeit with a different graph.

In order to have a low-complexity algorithm, the estimation of the marginal distributions (2.43) can be made through the Bethe free energy. We then refer to a belief propagation guided decimation (BPGD) algorithm. If, as in Algorithm 2.1, the value of u_i^* is sampled

---

Algorithm 2.1 Sampling by decimation

---

for $i = 1$ *to* $i = n$ **do**

compute the conditional marginal distribution

$$\mu_\beta \left(u_i \mid \underline{x}, u_1^*, \ldots, u_{i-1}^* \right) ; \qquad (2.43)$$

sample u_i^* according to (2.43);

return $\underline{u}^*$;

---

from the marginal distribution (2.43), we say that we have a BPGD algorithm with randomized rounding (BPDG-r).

We also mention the existence of another type of decimation widely used for constraint satisfaction problems. The bit u_i^* is fixed to a value that maximizes the marginal distribution (2.43). Such a decimation process is call hard and the associated algorithm is denoted by BPDG-h. We stress that these algorithms are not samplers. The reasons for using hard decimation procedures is primarily empirical. BPGD-h algorithms are deterministic and they achieve distortions lower than BPGD-r algorithms. Unfortunately there is no clear interpretation of the output sequence of BPGD-h algorithms in terms of the Gibbs measure (2.15).

2.2.4 Spatially-Coupled LDGM Codes

Spatially-coupled LDGM codes are codes based on a special class of $M \times N$ generator-matrices. These matrices are block lower-band matrices. There are L column partitions, $L + w - 1$ row partitions and the blocks are formed of $\frac{M}{L+w-1} \times \frac{N}{L}$ matrices $G_{u,v}$. The integer L is called the spatial extent or length and the integer $w \le L$ is called the coupling window. The reason for having a terminology related to space will appear once the graphical model is built. A generator-matrix G of a spatially-coupled LDGM code has the following form

$$G = \begin{pmatrix} G_{1,1} & 0 & 0 \\ \vdots & \ddots & 0 \\ G_{1+w-1,1} & & G_{L,L} \\ 0 & \ddots & \vdots \\ 0 & 0 & G_{L+w-1,L} \end{pmatrix} . \qquad (2.44)$$

If the columns of G have a weight distributed according to P and the rows in the partitions $w - 1 < u < L + 1$ have a weight distributed according to Λ, we say that the matrix belongs to the ensemble $\text{SCLDGM}(\Lambda, P, L, w, M)$. Notice that the rows in the partition $u \le w - 1$ or $u \ge L + 1$ have a different weight distribution. Those partitions play an important role in coupled codes.
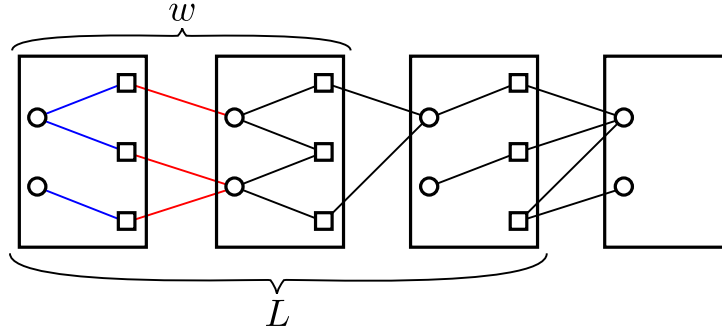
Figure 2.9: Factor graph associated with the spatially-coupled matrix (2.45).

The principal feature of graphical models associated with spatially-coupled codes is that they possess an ordering along a "spatial" direction. The spatial position is denoted by the non-negative integer $z \leq L + w - 1$. The variable and function node sets are also composed of subsets labeled by their respective spatial position $V = \cup_{z=1}^{L+w-1} V_z$ and $C = \cup_{z=1}^{L} C_z$. The subset of function nodes C_z at position z contains m/L elements and the subset of variable nodes V_z at position z contains $n/(L + w - 1)$ elements. We draw an edge $(i, a) \in E$ iff there exists a non-zero element in ith row and the ath column of the matrix G. Note that an edge that links a variable node $i \in V_v$ to a function node $a \in C_u$ corresponds to an element in the block $G_{u,v}$. Therefore due to the band structure of the matrix (2.45), a variable node $i \in V_v$ can be connected to a function node $a \in C_u$ only if $0 \leq v - u \leq w$.

The following matrix belongs to the SCLDGM $\left(\frac{1}{4}z^1 + \frac{1}{2}z^3 + \frac{1}{4}z^4, z^2, 4, 2, 8 \right)$ ensemble

$$
G_{\text{SC}} = \begin{pmatrix}
\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & 0 & 0 \\
\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} & 0 \\
0 & \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\
0 & 0 & \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}
\end{pmatrix}. \tag{2.45}
$$

The corresponding graphical model is depicted in Figure 2.9. Nodes at a common position z are enclosed in a black box that is called a "copy" of the underlying LDGM code. Blocks $G_{u,u}$ in a spatially-coupled matrix are associated with edges drawn inside a copy, as the other blocks are associated with edges between distinct copies. The blocks $G_{1,1}$ (resp. $G_{1,2}$) in (2.45) and its associated edges in Figure 2.9 are shown in blue (resp. red).

2.2.5 The Poisson-Regular Degree Distributions

The LDGM ensemble studied in this thesis has p-regular check nodes and no constraints on the degree distribution of the variable nodes.

The $M \times N$ LDGM matrix is randomly generated as follows: Initialize the matrix to the all-zero matrix. For each column, choose p entries at random and assign their values to one. The degree of the variable nodes is then a random variable with Binomial distribution $\mathrm{Bi}(N, p/M)$. In the asymptotic regime of large N, M with $M/N = R$ the code-bit node degrees are i.i.d. Poisson distributed with an average degree p/R.

The coupled version of the LDGM matrix is generated analogously. For each column in the partition $z \in \{1, \ldots, L\}$, we choose p entries at random among the rows in the partition $u \in \{z, \ldots, z + w - 1\}$ and assign their values to 1.

To simplify notation, we will denote these ensembles by $\mathrm{LDGM}(p, R, N)$ and their coupled version by $\mathrm{SCLDGM}(p, R, L, w, \rho)$, where $\rho = N/L$ is the density of function nodes per copy.

One reason to look at the variable Poisson-degree distribution is that it possesses a universality property that the other distributions do not. Suppose that the amount of memory for compressing a sequence is fixed to M but the length of the sequence is unknown, except that it is less than $N < M/R^*$, for some rate R^*. One way of treating this issue could be to have several different matrices with rates ranging from R^* to 1. This would require the storage of several matrices. The Poisson distribution, however, enables us to cover all rates $R > R^*$ using only one matrix. To do so, we first generate an $M \times M/R^*$ matrix G from the $\mathrm{LDGM}(p, R^*, M/R^*)$ ensemble (resp. from the $\mathrm{SCLDGM}(p, R^*, L, w, M/LR^*)$ ensemble). If $N = M/R < M/R^*$, we delete from G the columns $k(1 - R^*/R)^{-1}$ with $k \in \{1, \ldots, M(1/R^* - 1/R)\}$. The resulting matrix $\widetilde{G}$ belongs to the $\mathrm{LDGM}(p, R, N)$ ensemble (resp. belongs to the $\mathrm{SCLDGM}(p, R, L, w, N/L)$ ensemble).

3 Contribution and Organization

In Part II of the thesis, we prove the exactness of the Bethe free energy for LDGM and LDPC codes over a BSC at high noise. These results are based on the paper by Macris and *Vuffray* [88].

In Part III of the thesis, we show that spatially-coupled LDGM codes with BPGD algorithms approach the rate-distortion bound. This work and its presentation are based on the paper by Aref, Macris and *Vuffray* [89].

In this chapter, we summarize the main contributions of the thesis.

Exactness of the Bethe Free Energy for Sparse Graph Codes (Part II)

We prove that the Bethe free energy enables us to compute the conditional free entropy of LDGM and LDPC code instances on a BSC at high noise. By this we mean that the Bethe free energy associated with an LDPC code or an LDGM code used over a BSC in a large noise regime is, with high probability, asymptotically exact as the block length grows. The proof of this fact is obtained with the help of a tool from statistical mechanics, called the polymer expansion, which is new in the context of coding theory. The polymer expansion has an easy application to the case of low-density generator-matrix (LDGM) codes for high noise, and more generally, to graphical models in a high-temperature regime. It is interesting that the polymer expansion has to be combined with special features of the graphical model associated with LDPC codes (features that are not needed in the usual applications of the polymer expansion).

The Polymer Expansion (Chapter 4)

In 2006 Chertkov and Chernyak [90] developed a loop-sum representation for the partition function of graphical models. The virtue of this representation is that the partition function factorizes as the product of the Bethe contribution and a finite sum of terms over subgraphs (not necessarily connected) with no dangling edges. Each term of the sum involves only the belief-propagation messages on the edges contained in the subgraphs and in its edge boundary . In [90] these subgraphs are called loops. It is tempting to use the loop-sum representation not only as a mere formal tool, but to compare the true and Bethe free energies.

One of our contributions in the thesis is the development of this idea from a systematic point of view. We recognize that the loop-sum is itself the partition function of a system of polymers (Equation (4.22) on page 54). A loop can be seen as an union of connected subgraphs with no dangling edges, called polymers. Each polymer has an associated weight that depends only on the belief-propagation messages on the edges contained in the polymer and in its edge boundary. The representation of the loop-sum as the partition function of a polymer system with infinitely repulsive interactions opens the way to the computation of the logarithm of this sum via a combinatorial expansion known in statistical mechanics as the polymer expansion [91] (Lemma 4.1 on page 56).

If this expansion converges, then we can in principle compute corrections to the Bethe free energy (which corresponds to the zero-th order term of the expansion) to an arbitrarily high order. A sufficient condition for the convergence of the expansion is that the activity of a polymer decreases exponentially fast enough in the polymer size (Corollary 4.1 on page 59).

At the end of this chapter, we provide a pedagogical application of the polymer expansion for the square lattice Ising model. We derive the first correction terms to the Bethe free energy for this model and compare the results to the exact solutions. We show that the first correction terms given by the polymer expansion significantly improve the Bethe estimation of the critical temperature.

High-Temperature Models (Chapter 5)

If the girth of the graph is large, all contributions beyond the zero-th order Bethe free energy only come from large loops and, if these contributions become small as the size of a loop increases, we could expect that, provided the expansion converges uniformly in system size, the Bethe free energy is asymptotically exact. More generally, this mechanism could occur for typical instances of graphs from a random Erdős-Rényi type ensemble of sparse graphs, because the neighborhood of a given vertex is tree-like.

We show that the polymer expansion is absolutely convergent for general models pro-

vided that the temperature is high enough. We give an immediate application of this result to models whose factor graph has a large girth in the sense that the girth grows logarithmically with the size of the graph. For such models, we prove that the Bethe free energy is asymptotically exact in the thermodynamic limit (Theorem 5.1 on page 75).

We directly apply this result for irregular LDGM codes in the high-noise regime. We prove that the free energy of an instance drawn at random from an irregular LDGM ensemble is, with high probability, given by the Bethe formula in the large block length limit (Theorem 5.2 on page 79).

LDPC Codes (Chapter 6)

In this chapter we consider regular LDPC codes used over a BSC in the high-noise regime. For this model the solutions of the BP equations are not unique. In particular, the low-noise solution is always a solution. We impose a restriction on the class of BP solutions in which we evaluate the Bethe free energy, and we call high-noise solutions the solutions obtained from this restriction (Hypothesis 6.1 on page 83). In the case of LDPC codes, we cannot prove that the polymer expansion is absolutely convergent. The reason is that the check-node constraints are indicator functions. Such constraints cannot be viewed as high-temperature constraints but rather of low (even zero) temperature. Therefore the polymer expansion cannot be used directly by itself.

It is interesting that we can show that a truncated form of this expansion does converge absolutely (uniformly in the system size), using expander properties of typical instances from the LDPC ensemble. Moreover, we prove that the convergent truncated expansion accounts for the biggest part of the corrections to the Bethe free energy, up to a remainder of order $O(e^{-n\epsilon})$, $\epsilon > 0$ (Theorem 6.1 on page 94). This remainder is not controlled with the polymer expansion but estimated by a combinatorial counting method. The final result is again that the Bethe free energy is asymptotically exact with high probability in the large size limit (Theorem 6.2 on page 96).

Approaching the Rate-Distortion Bound by Spatial Coupling, a Perspective from the Cavity Method (Part III)

We investigate an encoding scheme for lossy source compression of a binary symmetric source which is based on spatially-coupled LDGM codes. The degree of the check nodes is regular, and the degree of the variable nodes is Poisson distributed with an average depending on the compression rate. We show that under a low-complexity BPGD-h and BPGD-r algorithm, we approach the rate distortion curve for any compression rate.

The BPGD encoders are based on a random Gibbs measure with non-zero "temperature."

We investigate the links between the algorithmic performance of the BPGD encoders and the phase diagram of this Gibbs measure. Using the cavity method, we investigate the phase diagram that predicts a dynamical and condensation thresholds for the Gibbs measure. We observe that the dynamical threshold of the spatially-coupled code saturates towards the condensation temperature. Moreover, for large degrees the condensation temperature approaches the temperature related to the information theoretic Shannon test-channel noise parameter of rate-distortion theory. This provides heuristic insight into the excellent performance of the BPGD algorithms.

Approaching the Rate-Distortion Bound by Spatial Coupling (Chapter 7)

Recall that p denotes the check-node degree, R denotes the rate, L denotes the coupling length, w denotes the coupling window and ρ denotes the number of check nodes per copy.

The Gibbs measure (2.39) is the basis for setting up the BPGD algorithms. At each step of the decimation procedure, we compute with BP the marginal distribution of individual bits. The main idea is that, when the marginal distributions are close to be uniform, a random bit from the boundary of the chain is fixed to a random value; and as long as there exist bits with biased marginal distribution, they are eliminated from the chain by fixing them and decimating the graph. We consider two forms of BPGD algorithms (Algorithm 7.1 on page 113). The BPGD-h algorithm which turns out to perform slightly better, is based on hard decisions. The BPGD-r algorithm uses a randomized decimation rule for fixing the bits.

For both algorithms, we observe that the rate-distortion curve of the coupled ensemble approaches the Shannon limit when $\rho \gg L \gg w \gg 1$ and the node degree grows large (Figure 7.3 on page 114 and Figure 7.4 on page 115). We cannot assess if the Shannon limit is achieved based on our numerical results. However we observe that, in order to avoid finite size effects, the degree has to grow large only after the other parameters grow large in the specified order. In practice though $\rho = 2000$, $L = 64$, $w = 3$ and check degrees equal to $p = 3$ yield good results for a compression rate $1/2$.

The performance of the BPGD algorithms depends on the inverse temperature parameter in the Gibbs measure and is optimal for a specific $\beta^*(R, p)$ (Figure 7.5 on page 116). Using the prediction of the cavity method for the average distortion, we compare the quality of sampling of BPGD-r algorithms. We observe that the BPGD-r algorithm correctly samples the Gibbs measure (2.39) down to inverse temperatures roughly higher than β^* (Figure 7.6 on page 117).

A Perspective from the Cavity Method (Chapter 8)

We provide a summary of the main technical ideas of the cavity method at the beginning of this chapter. We then proceed to derive the equations necessary for computing the dynamical and condensation thresholds β_d and β_c of the Gibbs measure (2.39). The cavity equations take the form of six fixed-point integral equations (Equations (8.51) on page 131 and (8.52) on page 131 and Equations (8.54) on page 131 and (8.55) on page 132) .

Using population dynamics we solve the fixed point integral and compute the values of these thresholds for the coupled and underlying ensembles (Table 8.1 on page 139 and Table 8.2 on page 139). We observe that for both ensembles we always have $\beta^* < \beta_d$. In particular, the BPGD-r algorithm does not sample correctly after β_d. We also observe that when $\rho \gg L \gg w \gg 1$ and for every p, there is threshold saturations $\beta^* \to \beta_d \to \beta_c$ for the coupled ensembles. This explains the good performance of the BPGD-r algorithm. On the coupled ensembles, the BPGD-r correctly samples at low-temperatures configurations that have a small distortion.

Moreover, we observe by population dynamics that two of the six cavity equations are satisfied by a trivial fixed point. This is justified by a theoretical analysis for large degrees p. When this trivial fixed point is used, the remaining four equations reduce to two fixed-point integral equations that have the form of usual density evolution equations for a BSC (Equations (8.63) on page 136 and (8.55) on page 132). This simplification is slightly surprising because although the original Gibbs measure does not possess channel symmetry[1], this symmetry emerges as a solution of the cavity equations. Within this framework the saturation of the dynamical temperature towards the condensation temperature appears to be very similar to threshold saturation, in the context of channel coding with LDPC codes.

The simplifications of the cavity equations enables us to perform an asymptotic analysis in the limit $p \to \infty$. We show that in this limit the dynamical threshold β_d goes to zero for the underlying ensembles. Moreover, for both ensembles $\beta_c \to \beta_{\mathrm{Sh}}$, where β_{Sh} corresponds to the amplitude of the half log-likelihood ratios of a BSC test channel at capacity $C_{\mathrm{BSC}} = R$ (Equation (8.73) on page 138). This shows that the rate-distortion curve can be interpreted as a line of condensation thresholds for each R.

---

[1]In the context of spin glass theory this is the Nishimori gauge symmetry.

Exactness of the Bethe Free Energy for Sparse Graph Codes

Part II

4 The Polymer Expansion

In this chapter, we present the polymer expansion and show that it enables us to compute the correction terms of the Bethe free energy in a systematic way.

In Section 4.1, we set up the notation and recall the definition of the Bethe free energy.

In Section 4.2, for the convenience of the reader we provide a derivation of the loop sum identity of Chertkov and Chernyak.

In Section 4.3, we show that the loop sum can be represented as the partition function of a system of polymers (Equation (4.22) on page 54). We also present the polymer expansion that enables us to compute the corrections to the Bethe free energy (Lemma 4.1 on page 56).

In Section 4.4, we give a sufficient condition which ensures the uniform convergence of the polymer expansion. The condition is that the activity of a polymer decreases exponentially fast enough in the polymer size (Corollary 4.1 on page 59).

In Section 4.5, we provide a pedagogical application of the polymer expansion for the Ising model. To our knowledge this formalism has never been applied on this model.

4.1   Settings

The LDGM and LDPC codes are special cases of general factor graph models. We find it convenient to develop the formalism of the loop sum and polymer expansions in a unified manner which applies to general models.

We begin with a few definitions and notations. Consider two vertex sets: V a set of n variable nodes and C a set of m check nodes. We think of n and m large. We consider bipartite graphs – call them Γ – connecting V and C. The set of edges is E. Letters i, j will always denote nodes in V and letters a, b nodes in C. We reserve the notations ∂i

(resp. ∂a) for the sets of nodes that are neighbors of i (resp. a) in Γ.

Consider a bipartite graph Γ. We construct a general factor graph model or spin system as follows. We attach spin degrees of freedom $s_i \in \{-1, +1\}$ to nodes $i \in V$. A spin configuration is an assignment $\underline{s} = (s_i)_{i \in V}$. To each check node a we associate a weight depending on spins $i \in \partial a$. The collection of spins s_i with $i \in \partial a$ and the associated weights are denoted $s_{\partial a}$ and $\psi_a(s_{\partial a})$, respectively. The partition function of the factor graph model (or spin system) is

$$Z = \sum_{\underline{s} \in \{-1, +1\}^n} \prod_{a \in C} \psi_a(s_{\partial a}). \tag{4.1}$$

The free energy is defined by

$$f = \frac{1}{n} \ln Z \tag{4.2}$$

and its thermodynamic limit is taken in the limit $n \to +\infty$.

If we restrict ourselves to the class of strictly positive weights their most general form is

$$\psi_a(s_{\partial a}) = \exp\{\beta \sum_{I \subset \partial a} J_I \prod_{i \in I} s_i\}, \tag{4.3}$$

where $\beta > 0$ has the interpretation of an inverse temperature and $J_I \in \mathbb{R}$ have the interpretation of coupling constants.

The cavity method enables us the computation of candidate approximations, called Bethe free energies, for $f = \frac{1}{n} \ln Z$. In the case of LDPC and LDGM codes it enables us the computation of candidate approximations for the free energies f_{LDPC} and f_{LDGM}. As explained in the introduction, controlling in a rather systematic way the quality of these approximations is the object of Part II.

Let us first recall the Bethe formulas for general factor graph models. This involves a set of messages $\nu_{ia}(s_i)$ and $\widehat{\nu}_{ai}(s_i)$ attached to the edges of $(ia) \in E$. Messages are probability distributions over the single spin space $\{-1, 1\}$. The collection of all messages is denoted $(\underline{\nu}, \widehat{\underline{\nu}})$; they satisfy the belief propagation fixed-point equations

$$\begin{cases} \nu_{ia}(s_i) = \dfrac{\prod_{b \in \partial i \backslash a} \widehat{\nu}_{bi}(s_i)}{\sum_{s_i} \prod_{b \in \partial i \backslash a} \widehat{\nu}_{bi}(s_i)}. \\[4mm] \widehat{\nu}_{ai}(s_i) = \dfrac{\sum_{s_{\partial a \backslash i}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \backslash i} \nu_{ja}(s_j)}{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \backslash i} \nu_{ja}(s_j)} \end{cases} \tag{4.4}$$

The Bethe free energy functional associated with a particular solution of these equations

is

$$f^{\text{Bethe}}(\underline{\nu}, \widehat{\underline{\nu}}) = \frac{1}{n} \left(\sum_{a \in C} F_a + \sum_{i \in V} F_i - \sum_{(i,a) \in E} F_{ia} \right), \tag{4.5}$$

where

$$\begin{cases} F_a = \ln\{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{j \in \partial a} \nu_{ja}(s_j)\}, \\[2ex] F_i = \ln\{\sum_{s_i} \prod_{b \in \partial i} \widehat{\nu}_{bi}(s_i)\}, \\[2ex] F_{ia} = \ln\{\sum_{s_i} \nu_{ia}(s_i)\widehat{\nu}_{ai}(s_i)\}. \end{cases} \tag{4.6}$$

Notice that the functional $f^{\text{Bethe}}(\underline{\nu}, \widehat{\underline{\nu}})$ is well defined for any sets of messages $\underline{\nu}$ and $\widehat{\underline{\nu}}$. The stationary points of $f^{\text{Bethe}}(\underline{\nu}, \widehat{\underline{\nu}})$ considered as a function of the messages over $\mathbb{R}^E \times \mathbb{R}^E$, satisfy the belief propagation equations (see [92]).

4.2 Loop Sum Identity

The "loop sum identity" is a representation of the error term between the free energy and the Bethe free energy. It takes the form of the logarithm of a sum over sub-graphs that are non-necessarily connected. This identity was first derived for graphical models with binary variables by Chertkov and Chernyak in [90] and later generalized for variables on a q-ary alphabet by the same authors in [93]. The extension of the loop sum identity to continuous alphabet has been carried out by Xiao and Zhou in [94]. The present section contains a short derivation of the loop sum identity based on the original paper [90]. There exists other representations of the loop sum identity and its generalization notably as the holographic transformation of a normal factor graph [95, 96].

Consider the problem of computing the partition function of a factor graph model (Equation (4.1)). The loop expansion takes a natural form on graphical models called vertex models, where variables are attached to edges. We introduce the auxiliary set of spins $\sigma_{ia}, \widehat{\sigma}_{ai} \in \{-1, 1\}$ attached to directed edges $(i \to a)$ and $(a \to i)$ respectively. We denote by $\sigma_{\partial a} = \{\sigma_{ja} | j \in \partial a\}$ the collection of spins that are on edges pointing toward a and we denote by $\widehat{\sigma}_{\partial i} = \{\widehat{\sigma}_{bi} | b \in \partial i\}$ the collection of spins that are on edges pointing toward i. We can rewrite (4.1) as a partition function of a vertex model

$$Z = \sum_{\underline{\sigma}, \widehat{\underline{\sigma}} \in \{-1,1\}^{|E|}} \prod_{a \in C} \psi_a(\sigma_{\partial a}) \prod_{i \in V} \phi_i(\widehat{\sigma}_{\partial i}) \prod_{(a,i) \in E} \frac{1 + \sigma_{ia}\widehat{\sigma}_{ai}}{2}, \tag{4.7}$$

where

$$\phi_i(\widehat{\sigma}_{\partial i}) = \prod_{b,c \in \partial i} \frac{1 + \widehat{\sigma}_{bi}\widehat{\sigma}_{ci}}{2}. \tag{4.8}$$

Let us comment on the expression (4.7). The new factors (4.8) ensure that all spins on edges outgoing from a variable node i take the same value s_i. The last product in (4.7) forces spins on the same edge to be equal. The key idea in the loop expansion is to "soften" the constraints on the edges before performing the expansion.

Using the following identity, valid for any binary distributions ν_{ia} and $\widehat{\nu}_{ai}$,

$$\frac{1 + \sigma_{ia}\widehat{\sigma}_{ai}}{2} = \frac{\nu_{ia}(\sigma_{ia})\widehat{\nu}_{ai}(\widehat{\sigma}_{ai}) + \sigma_{ia}\widehat{\nu}_{ai}(-\sigma_{ia})\widehat{\sigma}_{ai}\nu_{ia}(-\widehat{\sigma}_{ai})}{\sum_{s\in\{-1,1\}}\nu_{ia}(s)\widehat{\nu}_{ai}(s)},\tag{4.9}$$

we can rewrite the partition function (4.7) as

$$Z = \sum_{\underline{\sigma},\widehat{\underline{\sigma}}\in\{-1,1\}^{|E|}}\prod_{a\in C}\psi_a(\sigma_{\partial a})\prod_{j\in\partial a}\nu_{ja}(\sigma_{ja})$$
$$\times \prod_{i\in V}\phi_i(\widehat{\sigma}_{\partial i})\prod_{b\in\partial i}\widehat{\nu}_{bi}(\widehat{\sigma}_{bi})$$
$$\times \prod_{(a,i)\in E}\left(\sum_{s\in\{-1,1\}}\nu_{ia}(s)\widehat{\nu}_{ai}(s)\right)^{-1}$$
$$\times \prod_{(a,i)\in E}\left(1 + \sigma_{ia}\frac{\widehat{\nu}_{ai}(-\sigma_{ia})}{\nu_{ia}(\sigma_{ia})}\widehat{\sigma}_{ai}\frac{\nu_{ia}(-\widehat{\sigma}_{ai})}{\widehat{\nu}_{ai}(\widehat{\sigma}_{ai})}\right).\tag{4.10}$$

We use the "generalized binomial formula" on graphs. For any function χ defined on the edges $e \in E$ of a graph Γ, the following relation holds

$$\prod_{e\in E}(1 + \chi(e)) = 1 + \sum_{g\subset\Gamma}\prod_{e\in E\cap g}\chi(e),\tag{4.11}$$

where the sum runs on every non-empty subset of edges represented by subgraphs g whose vertices are incident to the edges in the subset. Expanding the last product in (4.10) with the generalized binomial formula yields

$$Z = \exp(nf^{\text{Bethe}}(\underline{\nu},\widehat{\underline{\nu}})) \times \left(1 + \sum_{g\subset\Gamma}K(g)\right).\tag{4.12}$$

The quantity that is factorized in the expansion appears to be the Bethe free energy functional (4.5) evaluated at the distributions $\underline{\nu}$ and $\widehat{\underline{\nu}}$. The weights $K(g)$ associated with each subgraph can be distributed in contributions coming from vertices in g

$$K(g) = \prod_{i\in g\cap V}K_i\prod_{a\in g\cap C}K_a,\tag{4.13}$$

where

$$K_i(g) = \frac{\sum_{s_i} \prod_{a \in \partial i \setminus g} \widehat{\nu}_{ai}(s_i) \prod_{a \in \partial i \cap g} s_i \nu_{ia}(-s_i)}{\sum_{s_i} \prod_{a \in \partial i} \widehat{\nu}_{ai}(s_i)}, \tag{4.14}$$

and

$$K_a(g) = \frac{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{i \in \partial a \setminus g} \nu_{ia}(s_i) \prod_{i \in \partial a \cap g} s_i \widehat{\nu}_{ai}(-s_i)}{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{i \in \partial a} \nu_{ia}(s_i)}. \tag{4.15}$$

The sum over non-empty subgraphs in (4.12) is the loop sum identity. Note that for the moment the binary distributions entering in (4.9) are completely arbitrary. The transformation (4.9) is crucial in that it allows the preservation of the correlations between neighboring spins. Messages $\widehat{\nu}_{a \to i}$ directed toward a spin σ_i can be interpreted as an interaction from the neighboring variables $\underline{\sigma}_{\partial a \setminus i}$ that bias the average value of the spin σ_i. Expanding the Kronecker delta in (4.7) directly is equivalent as taking $\underline{\nu}$ and $\underline{\widehat{\nu}}$ as being uniform distributions. Such an expansion would be accurate only in a regime were the spins are almost independent from each other and almost uniformly distributed between $+1$ and -1. This is the case for instance in the high-temperature regime. Thanks to the transformation (4.9), the effect of correlations between neighboring spins can be captured by the distributions $\underline{\nu}$ and $\underline{\widehat{\nu}}$. Thus, for appropriate choices of messages $(\underline{\nu}, \underline{\widehat{\nu}})$, the expansion can also be accurate in the low-temperature regime.

In order for the loop sum identity to be useful, one has to choose the "correct" binary distributions. We call $d_i(g) = |\partial i \cap g|$ (resp. $d_a(g) = |\partial a \cap g|$) the induced degree of node i (resp. a) in a subgraph g. If $d_i(g) \geq 2$ and $d_a(g) \geq 2$ for all $i, a \in g$, we say that g is a loop. In other words a loop has no dangling edge. The natural requirement for sparse locally tree-like graphs is that every subgraph g that is not a loop must have a zero weight. In particular if Γ is a trees, it implies that the loop sum is equal to zero. This condition requires that the distributions ν and $\widehat{\nu}$ must be chosen such that $d_a(g) = 1$ and $d_i(g) = 1$ implies $K_a(g) = 0$ and $K_i(g) = 0$ respectively. The requirement $K_a(g) = 0$ is fulfilled by distributions $\widehat{\nu}_{ai}$ that satisfy the following equation

$$\sum_{s_i} s_i \widehat{\nu}_{ai}(-s_i) \sum_{s_{\partial a \setminus i}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j) = 0. \tag{4.16}$$

This is satisfied if $\widehat{\nu}_{ai}$ is a solution of the first belief propagation equation (4.4)

$$\widehat{\nu}_{ai}(s_i) = \frac{\sum_{s_{\partial a \setminus i}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}. \tag{4.17}$$

Similarly one can check that the requirement $K_i(g) = 0$ is fulfilled by the choice

$$\nu_{ia}(s_i) = \frac{\prod_{b \in \partial i \setminus a} \widehat{\nu}_{bi}(s_i)}{\sum_{s_i} \prod_{b \in \partial i \setminus a} \widehat{\nu}_{bi}(s_i)}. \tag{4.18}$$

$$g = \gamma_1 \cup \gamma_2 \qquad \emptyset = \gamma_1 \cap \gamma_2$$
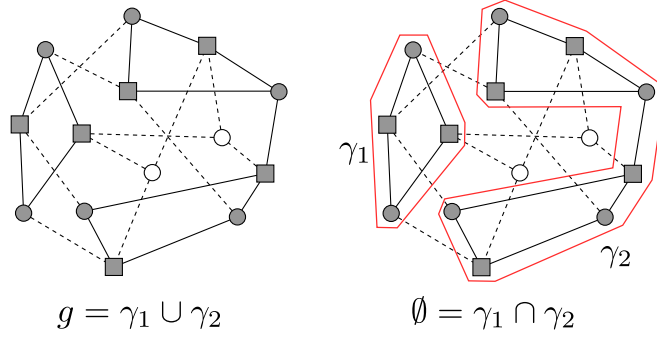
Figure 4.1: Example of $\Gamma \in \mathcal{B}(3,4,8)$. The loop g has two disjoint connected parts γ_1 and γ_2.

This is nothing else but the second belief propagation equation (4.4).

4.3 Polymer Expansion of the Loop Sum Identity

4.3.1 Polymer Representation

In the loop sum identity (4.12), only non-empty subgraphs $g \subset \Gamma$ with no isolated vertices and no dangling edges are present. These subgraphs are called "loops". Note that these loops are not necessarily cycles, and are not necessarily connected. Figure 4.1 shows an example.

Each generalized loop can be decomposed in a unique way as an union of its connected components

$$g = \cup_k \gamma_k, \tag{4.19}$$

where γ_k are the connected components of g. It is easy to see that the γ_k entering in (4.19) are non-empty connected loops and are mutually disjoint (see figure 4.1). The connected loops γ_k are called polymers. Remarkably, each weight $K(g)$ can be factorized (see Section 4.2, Eqn. (4.13)) in a product of contributions associated with the connected parts of g. We have

$$K(g) = \prod_k K(\gamma_k). \tag{4.20}$$

The factorization implies

$$1 + \sum_{g \subset \Gamma} K(g) \equiv Z^{\mathrm{polymer}}, \tag{4.21}$$

with

$$Z^{\mathrm{polymer}} = \sum_{M \geq 0} \frac{1}{M!} \sum_{\gamma_1,...,\gamma_M \subset \Gamma} \prod_{k=1}^{M} K(\gamma_k) \prod_{k<k'} \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset). \tag{4.22}$$

$$\mathcal{G}_1 \qquad U_1(\gamma_1) \qquad\qquad \mathcal{G}_3 \qquad U_3(\gamma_1, \gamma_2, \gamma_3)$$

①   1

$$(\mathbb{I}(\gamma_1 \cap \gamma_2 = \emptyset) - 1)$$
$$\times (\mathbb{I}(\gamma_1 \cap \gamma_3 = \emptyset) - 1)$$

$$+$$

$$(\mathbb{I}(\gamma_1 \cap \gamma_2 = \emptyset) - 1)$$
$$\times (\mathbb{I}(\gamma_2 \cap \gamma_3 = \emptyset) - 1)$$

$$\mathcal{G}_2 \qquad U_2(\gamma_1, \gamma_2)$$

$$+$$

① —— ② $(\mathbb{I}(\gamma_1 \cap \gamma_2 = \emptyset) - 1)$

$$(\mathbb{I}(\gamma_1 \cap \gamma_3 = \emptyset) - 1)$$
$$\times (\mathbb{I}(\gamma_2 \cap \gamma_3 = \emptyset) - 1)$$

$$+$$

$$(\mathbb{I}(\gamma_1 \cap \gamma_2 = \emptyset) - 1)$$
$$\times (\mathbb{I}(\gamma_1 \cap \gamma_3 = \emptyset) - 1)$$
$$\times (\mathbb{I}(\gamma_2 \cap \gamma_3 = \emptyset) - 1)$$

Figure 4.2: All the Mayer graphs for $M = 1, 2, 3$ and their associated Ursell functions.

In the second sum on the right-hand side, each γ_k runs over all polymers contained in Γ. The factor $\frac{1}{M!}$ accounts for the fact that a polymer configuration has to be counted only once. The indicator function ensures that the polymers do not intersect. By convention the term corresponding to $M = 0$ is equal to 1. The convention is also taken that if $M = 1$ the product over the indicator functions equals 1. Note that because of the non-intersection constraint of the polymers, the number of terms in the sums on the right-hand side is finite.

From a physical point of view (4.22) interprets the loop sum in Eqn. (4.12) as the *partition function of a gas of polymers* that can acquire any shape allowed by Γ. The polymers in this gas have an *activity*[1] $K(\gamma)$ and interact via a two body hard-core repulsion which precludes their overlap. This analogy enables us to use methods from statistical mechanics to analyze the corrections to the Bethe free energy.

4.3.2 A Combinatorial Tool: the Polymer Expansion

The polymer expansion is a powerful tool from statistical physics to expand the logarithm of a polymer partition function in powers of the activity. We give in this section a quick derivation of the polymer expansion based on [91].

We denote by $\mathcal{G}_M$ the set of all connected graphs G with M labeled vertices $1, \ldots, M$

---

[1] "Activity" is the name used by chemists for the prior probability weight $K(\gamma)$ of an isolated polymer. Note that here $K(\gamma)$ can be negative and this analogy is at best formal. We use the name "activity" rather than "prior weight" for $K(\gamma)$ precisely because they can be negative in the present context.

(see figure 4.2). These are called Mayer graphs.We associate to an ensemble of Mayer graphs $\mathcal{G}_M$ an Ursell function whose arguments are polymers

$$U_M(\gamma_1, \ldots, \gamma_M) = \begin{cases} \sum_{G \in \mathcal{G}_M} \prod_{(k,k') \in G} \left(\mathbb{I}\left(\gamma_k \cap \gamma_{k'} = \emptyset\right) - 1 \right) & \text{if } M \geq 2 \\ 1 & \text{if } M = 1 \end{cases}, \qquad (4.23)$$

were the product runs over the edges (k, k') of G.

Lemma 4.1 (Polymer expansion)**.** If Z^{polymer} is the partition function of a polymer system (Eqn. (4.22)), the following equality hold

$$\ln Z^{\text{polymer}} = \sum_{M=1}^{\infty} \frac{1}{M!} \sum_{\gamma_1, \ldots, \gamma_M \subset \Gamma} U_M(\gamma_1, \ldots, \gamma_M) \prod_{k=1}^{M} K(\gamma_k). \qquad (4.24)$$

Proof. Let us introduce some notations. We denote the complete graph with M labeled vertices by $\mathcal{K}_M$. A partition of the set $\{1, \ldots, M\}$ into q "blocks" is an unordered list $\{\mathcal{I}_1, \ldots, \mathcal{I}_q\}$ of disjoint nonempty subsets $\mathcal{I}_t \subset \{1, \ldots, M\}$. The partitions of M elements into q "blocks" form an ensemble denoted by $\mathcal{P}_M^q$.

We recall that polymers γ are connected subgraphs of Γ that cannot intersect due to the presence of the hard core constraints $\mathbb{I}\left(\gamma_k \cap \gamma_{k'} = \emptyset\right)$. The polymer expansion identity is based on the expansion of these hard core constraints using the binomial theorem on graphs (Eqn. (4.11))

$$\prod_{k<k'} \mathbb{I}\left(\gamma_k \cap \gamma_{k'} = \emptyset\right) = \prod_{(k,k') \in \mathcal{K}_M} \left(\mathbb{I}\left(\gamma_k \cap \gamma_{k'} = \emptyset\right) - 1 + 1 \right)$$

$$= 1 + \sum_{G \subset \mathcal{K}_M} \prod_{(k,k') \in G} \left(\mathbb{I}\left(\gamma_k \cap \gamma_{k'} = \emptyset\right) - 1 \right). \qquad (4.25)$$

The sum in (4.25) runs over non-empty subset of edges of $\mathcal{K}_M$ represented by subgraphs G whose vertices are incident to the edges in the subset. Notice that each general subgraph in $\mathcal{K}_M$ can be written as an union of disjoint connected subgraphs $G_1, \ldots, G_q$. This with the fact that $U_1(\gamma_k) = 1$ enable us to re-sum (4.25) as

$$1 + \sum_{G \subset \mathcal{K}_M} \prod_{(k,k') \in G} \left(\mathbb{I}\left(\gamma_k \cap \gamma_{k'} = \emptyset\right) - 1 \right) = \sum_{q=1}^{M} \sum_{\{\mathcal{I}_1, \ldots, \mathcal{I}_q\} \in \mathcal{P}_M^q} \prod_{t=1}^{q} U_{|\mathcal{I}_t|}\left((\gamma_k)_{k \in \mathcal{I}_t} \right). \qquad (4.26)$$

Figure 4.3 gives a pictorial representation of the formula (4.26) in the case $M = 3$.

Together with (4.25) and (4.26), the polymer partition function (4.22) can be rewritten
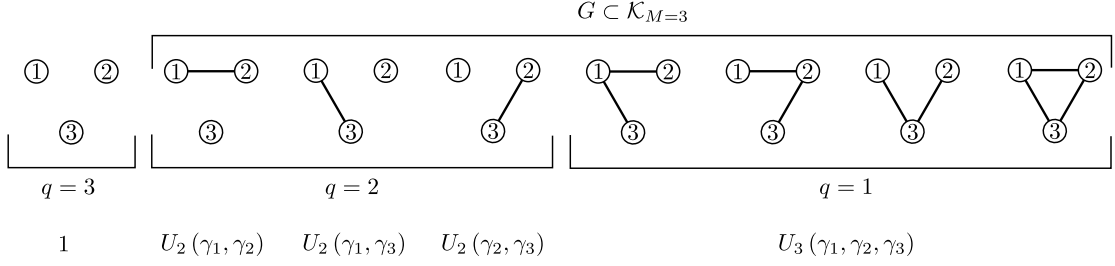
Figure 4.3: Example in the case $M = 3$ of the terms entering in the re-summation (4.26). The number of "blocks" in a partition of $\{1, 2, 3\}$ is denoted by q. Recall that $U_1(\gamma_k) = 1$.

as

$$Z^{\text{polymer}} = 1 + \sum_{M \geq 1} \frac{1}{M!} \sum_{q=1}^{M} \sum_{\{\mathcal{I}_1, \ldots, \mathcal{I}_q\} \in \mathcal{P}_M^q} \prod_{t=1}^{q} \phi(\mathcal{I}_t), \tag{4.27}$$

where

$$\phi(\mathcal{I}_t) := \sum_{\gamma_k \in \mathcal{I}_t} U_{|\mathcal{I}_t|}\left((\gamma_k)_{k \in \mathcal{I}_t}\right) \prod_{k \in \mathcal{I}_t} K(\gamma_k). \tag{4.28}$$

The function introduced in (4.28) depends only on the size of the ensemble

$$\phi(\mathcal{I}_t) = \phi(|\mathcal{I}_t|), \tag{4.29}$$

as $k \in \mathcal{I}_t$ in (4.28) are just dummy indices. The number of partitions of $\{1, \ldots, M\}$ with prescribed size $|\mathcal{I}_1| = m_1, \ldots, |\mathcal{I}_q| = m_q$ is

$$\sum_{\{|\mathcal{I}_1| = m_1, \ldots, |\mathcal{I}_q| = m_q\} \in \mathcal{P}_M^q} 1 = \frac{M!}{q!} \prod_{t=1}^{q} \frac{1}{m_t!}, \tag{4.30}$$

where $m_1, \ldots, m_q$ are non-zero integers satisfying $m_1 + \cdots + m_q = M$. These considerations allow us to rewrite (4.27) as

$$\begin{aligned} Z^{\text{polymer}} &= 1 + \sum_{M \geq 1} \sum_{q=1}^{M} \frac{1}{q!} \sum_{m_1 + \cdots + m_q = M} \prod_{t=1}^{q} \frac{\phi(m_t)}{m_t!} \cdot \\ &= 1 + \sum_{q=1}^{M} \frac{1}{q!} \sum_{M=q}^{\infty} \sum_{m_1 + \cdots + m_q = M} \prod_{t=1}^{q} \frac{\phi(m_t)}{m_t!} \\ &= 1 + \sum_{q=1}^{M} \frac{1}{q!} \left(\sum_{M=1}^{\infty} \frac{\phi(M)}{M!} \right)^q \\ &= \exp\left(\sum_{M=1}^{\infty} \frac{\phi(M)}{M!} \right). \end{aligned} \tag{4.31}$$

The argument in the exponential of (4.31) is equal to the right-hand side of (4.24). $\quad\square$

Lemma 4.1 is rather general and is a combinatorial feature of polymer systems. It requires no special property from the polymers γ or from their containing graph Γ.

4.4 Convergence Criterion and Control of the Polymer Expansion

The polymer expansion presented in Section 4.3 gives us a way to compute the logarithm of the loop sum. Unlike the polymer partition function (4.22), the polymer expansion (4.24) contains an infinite number of terms as polymers can overlap. It is then necessary to have tools to ensure that the series is absolutely convergent. Besides in order to be able to take the limit $n \to \infty$ terms by terms in the series, we have to ensure that the convergence is uniform in the system size. This is ensured by the following lemma

Lemma 4.2 (Convergence criterion). *Let*

$$Q = \sum_{t=0}^{\infty} \frac{1}{t!} \sup_{x \in V \cup C} \sum_{\gamma \ni x} |\gamma|^t |K(\gamma)|, \tag{4.32}$$

where the last sum in (4.32) runs over polymers γ containing x. If $Q < 1$ then the polymer expansion (4.24) is absolutely convergent. Moreover we have the following bound

$$\left| \sum_{M=1}^{\infty} \frac{1}{M!} \sum_{\gamma_1,\dots,\gamma_M \subset \Gamma} U_M(\gamma_1,\dots,\gamma_M) \prod_{k=1}^{M} K(\gamma_k) \right| \leq \left(\frac{1}{1-Q} \right) \sum_{x \in V \cup C} \sum_{\gamma \ni x} |K(\gamma)| e^{|\gamma|}.$$

Proof. We use the remarkable inequality [91]

$$|U_M(\gamma_1,\dots,\gamma_M)| \leq \sum_{T \in \mathcal{T}_M} \prod_{(k,k') \in T} |\mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset) - 1|$$

$$= \sum_{T \in \mathcal{T}_M} \prod_{(k,k') \in T} \mathbb{I}(\gamma_k \cap \gamma_{k'} \neq \emptyset), \tag{4.33}$$

where $\mathcal{T}_M$ is the set of trees on M vertices labeled $1,\cdots,M$. Using (4.33) we find that the term of order M in (4.24) is smaller than

$$\left| \sum_{\gamma_1,\dots,\gamma_M \subset \Gamma} U_M(\gamma_1,\dots,\gamma_M) \prod_{k=1}^{M} K(\gamma_k) \right| \leq \sum_{T \in \mathcal{T}_M} \sum_{\gamma_1,\dots,\gamma_M} \prod_{(k,k') \in T} \mathbb{I}(\gamma_k \cap \gamma_{k'} \neq \emptyset) \prod_{k=1}^{M} |K(\gamma_k)|. \tag{4.34}$$

We will now estimate the sum over $\gamma_1,\dots,\gamma_M$ for each tree T. Let $t_1,\dots,t_M$ be the degrees of the nodes $1,\dots,M$. One can decide that γ_1 is the root of T and that the leafs are among $2,\dots,M$. We first perform recursively the sum over $\gamma_2,\dots,\gamma_M$ by starting

from the leaf nodes in this set. One finds the estimate

$$\sum_{\gamma_2,\ldots,\gamma_M} \prod_{k=2}^{M} |K\left(\gamma_k\right)| \prod_{(k,k')\in T\setminus\{1\}} \mathbb{I}\left(\gamma_k \cap \gamma_{k'} \neq \emptyset\right) \leq |\gamma_1|^{t_1} \prod_{k=2}^{M} \left(\sup_{x\in V\cup C} \sum_{\gamma\ni x} |\gamma|^{t_k-1} |K\left(\gamma\right)|\right). \tag{4.35}$$

This implies

$$\sum_{\gamma_1,\ldots,\gamma_M} \prod_{k=1}^{M} |K\left(\gamma_k\right)| \prod_{(k,k')\in T} \mathbb{I}\left(\gamma_k \cap \gamma_{k'} \neq \emptyset\right) \leq \sum_{y\in V\cup C} \sum_{\gamma_1\ni y} |K\left(\gamma_1\right)| |\gamma_1|^{t_1}$$
$$\times \prod_{k=2}^{M} \left(\sup_{x\in V\cup C} \sum_{\gamma\ni x} |\gamma|^{t_k-1} |K\left(\gamma\right)|\right). \tag{4.36}$$

Now it is easy to estimate the sum over T in (4.34). According to Cayley's formula, the number of trees with M labeled vertices of degrees $t_1,\ldots,t_M$ is equal to

$$\frac{(M-2)!}{(t_1-1)!\cdots(t_M-1)!}, \tag{4.37}$$

so we find that (4.34) is upper bounded by

$$\sum_{y\in V\cup C} \sum_{\gamma_1\ni y} |K\left(\gamma_1\right)| e^{|\gamma_1|} \frac{1}{M} \left(\sum_{t=0}^{+\infty} \frac{1}{t!} \sup_{x\in V\cup C} \sum_{\gamma\ni x} |\gamma|^t |K\left(\gamma\right)|\right)^{M-1}. \tag{4.38}$$

We see that in this expression the quantity in parenthesis is

$$Q = \sum_{t=0}^{+\infty} \frac{1}{t!} \sup_{x\in V\cup C} \sum_{\gamma\ni x} |\gamma|^t |K\left(\gamma\right)|, \tag{4.39}$$

which is smaller than 1 by hypothesis. $\qquad\square$

An immediate application of this lemma is the following corollary which gives a convergence condition independent of n.

Corollary 4.1 (Uniform convergence criterion). *Recall that the degree of the variable nodes is upper-bounded by l_{max} and the degree of the check nodes is upper-bounded by r_{max}. Define by R_x the minimal size of polymers containing $x \in V \cup C$. Let $\epsilon < \frac{1}{2}$. If*

$$|K\left(\gamma\right)| \leq \left(\epsilon e^{-1-\max(\mathrm{l_{max}},\mathrm{r_{max}})}\right)^{|\gamma|} \tag{4.40}$$

then:

1. $\frac{1}{n} \ln Z^{\mathrm{polymer}}$ *has an absolutely uniformly (in n) convergent expansion.*

2. $\frac{1}{n} \left| \ln Z^{\mathrm{polymer}} \right| \leq \frac{4}{n} \sum_{x \in V \cup C} \epsilon^{R_x}$.

Proof. The number of polymers with size $|\gamma| = \tau$ containing a node $x \in V \cup C$ is upper-bounded by $e^{\max(l_{\max}, r_{\max})\tau}$ (See Appendix B). Therefore as a polymer contains at least two nodes

$$Q \leq \frac{\epsilon^2}{1-\epsilon} \leq \frac{1}{2}. \tag{4.41}$$

Application of Lemma 4.2 ends the proof. □

An important consequence of Corollary 4.1 is that it implies in particular that the polymer free energy is analytic as a function of $\{K(\gamma), \gamma \subset \Gamma\}$.

A mathematically precise and simple way to express the analyticity of the series is to replace $K(\gamma)$ by $zK(\gamma)$, $z \in \mathbb{C}$, $|z| < z_0$, where $z_0 > 1$ is fixed. Then the polymer free energy becomes a function of the complex variable z,

$$\frac{1}{n} \ln Z^{\mathrm{polymer}}(z) \tag{4.42}$$

and (4.1) becomes a series expansion in z^M, $M \geq 1$. If the convergence criterion (4.40) holds with $K(\gamma)$ replaced by $z_0 K(\gamma)$ we can conclude by Montel's theorem [97] that the series is holomorphic for $|z| < z_0$. Moreover the limit $n \to +\infty$, as long as it exists, is also holomorphic for $|z| < z_0$. In practice, existence of the limit requires some regularity structure on the sequence of graphical models (which is not the case in the present formulation), and it can be checked term by term on the series expansion. We take $z_0 > 1$ in order to then apply the results to the case of interest $z = 1$.

As will be seen in Chapter 5 it is fairly easy to check that the inequality (4.40) is satisfied for high-temperature general-models and also for typical instances of LDGM codes in the large noise regime. These cases also serve as pedagogical examples to better understand the difficulties that arise in the case of LDPC codes. In fact for LDPC codes we are not able to satisfy this criterion as such. However the criterion holds if Γ is an expander and the sum in Z^{polymer} is restricted to small polymers of size $|\gamma| < \lambda n$, $0 < \lambda < \lambda_0 \ll 1$. The contribution of "large" polymers $|\gamma| > \lambda n$ is treated differently.

4.5 Illustration with the 2D Ising Model

The section's goal is to illustrate the concepts presented above with the square lattice Ising model encountered in Chapter 1. We compare quantities computed within the Bethe approximation and the exact solution and we show how the polymer expansion is effective to computes the corrections to the Bethe approximation.

4.5.1 Preamble

Denote by $\Lambda_n = (V, E)$ the square lattice with n spins with periodic boundary conditions. We will use the letters i, j for vertices in V and denote by a couple $(i, j) \in E$ an edge of the lattice. A spin is a binary variable $\sigma_i \in \{-1, 1\}$ associated with a vertex $i \in V$ of the lattice. A spin configuration is a vector of n binary variables $\underline{\sigma} \in \{-1, 1\}^n$. The Hamiltonian or energy-cost of the Ising model contains only ferromagnetic nearest neighbor interactions

$$H_n(\underline{\sigma}) = \sum_{(i,j) \in \Lambda_n} -J\sigma_i\sigma_j. \tag{4.43}$$

The occurrence probability of a configuration at temperature β^{-1} is given by the Gibbs measure

$$\mu_N(\underline{\sigma}) = \frac{1}{Z_n} e^{-\beta H_n(\underline{\sigma})}, \tag{4.44}$$

where Z_n is the partition function

$$Z_n = \sum_{\underline{\sigma}} e^{-\beta H_n(\underline{\sigma})}. \tag{4.45}$$

We denote by brackets $\langle \cdot \rangle_\mu$ the average with respect to the measure μ. Note that the measure (4.44) can be interpreted as a factor graph model on a bipartite graph if for each edge $(i, j) \in E$ we create a check node with factor $\exp(\beta J\sigma_i\sigma_j)$.

We are interested in the properties of the system in the thermodynamic limit $n \to \infty$. The quantities of interest are the free energy

$$f_{\text{Ising}} = \lim_{n \to \infty} \frac{1}{n} \ln Z_n, \tag{4.46}$$

the dimensionless internal energy

$$u = \lim_{n \to \infty} \frac{1}{Jn} \langle H_n \rangle_{\mu_n}, \tag{4.47}$$

the entropy

$$s = \lim_{n \to \infty} \frac{-1}{n} \langle \ln \mu_n \rangle_{\mu_n}, \tag{4.48}$$

and the heat capacity

$$C = \frac{d}{d\beta^{-1}} (Ju). \tag{4.49}$$

We introduce at this point the dimensionless rescaled temperature

$$z := \tanh(\beta J). \tag{4.50}$$

The temperature is minimal when $z = 0$ and becomes maximal when $z = 1$. Quantities (4.47), (4.48), (4.49) can be obtained by taking derivatives[2] of the free energy (4.46). It is straightforward to derive the following identities

$$u = -\left(1 - z^2\right) \frac{d}{dz} f_{\text{Ising}}, \tag{4.51}$$

$$s = f_{\text{Ising}} - \frac{1 - z^2}{2} \ln\left(\frac{1 + z}{1 - z}\right) \frac{d}{dz} f_{\text{Ising}}, \tag{4.52}$$

$$C = \frac{1 - z^2}{4} \left(\ln\left(\frac{1 + z}{1 - z}\right)\right)^2 \left(\left(1 - z^2\right) \frac{d^2}{dz^2} f_{\text{Ising}} - 2z \frac{d}{dz} f_{\text{Ising}}\right). \tag{4.53}$$

4.5.2 The Bethe Free Energy of the Ising Model

Because the measure (4.44) is translation invariant, the belief propagation messages are uniform over the graph. The BP messages $\nu(\sigma)$ are parametrized by a single number

$$\xi = \sum_\sigma \sigma \nu(\sigma).$$

The Bethe free energy for the square-lattice Ising model takes the following form in the limit $n \to \infty$

$$f_\xi^{\text{Bethe}}(z) := \ln\left((1 + \xi z)^4 + (1 - \xi z)^4\right) - 2\ln\left(1 + \xi^2 z\right) - \ln\left(1 - z^2\right), \tag{4.54}$$

where the value of the parameter $\xi \in [-1, 1]$ satisfies the following belief propagation equation

$$\xi = \frac{(1 + z\xi)^3 - (1 - z\xi)^3}{(1 + z\xi)^3 + (1 - z\xi)^3}. \tag{4.55}$$

Equation (4.55) has several solutions and we should define a way to select one of them in which to evaluate $f_\xi^{\text{Bethe}}(z)$. For the Ising model, it appears that the polymers activities $K(\gamma)$ are non-negative independently of the BP fixed point. It implies that the Bethe free energy is a lower bound to the free energy

$$f_\xi^{\text{Bethe}}(z) \leq f_{\text{Ising}}(z), \tag{4.56}$$

reproducing the results of Sudderth et al. [99] and Ruozzi [100] for more general graphical models. Inequality (4.56) implies that following fixed-point selection principle minimizes

---

[2]The free energy is a convex function of z. This enables us to exchange the derivative with the thermodynamic limit [98, p. 203].

the error

$$f^{\text{Bethe}}(z) := \max_{\xi} f_{\xi}^{\text{Bethe}}(z).$$ (4.57)

A phase transition occurs in the number of solutions of the belief propagation equation (4.55) at

$$z_c^{\text{Bethe}} = \frac{1}{3}.$$ (4.58)

The high-temperature regime, $z < z_c^{\text{Bethe}}$, is characterized by a unique, trivial solution

$$\xi_{\text{high}} = 0.$$ (4.59)

As for low temperatures, $z \geq z_c^{\text{Bethe}}$, two solutions which differs only by a sign emerge

$$\xi_{\text{low}} = \pm\sqrt{\frac{1 - 3z}{(z - 3)z^2}}.$$ (4.60)

The associated high and low-temperature free energies are respectively

$$f_{\text{high}}^{\text{Bethe}}(z) = \ln 2 - \ln\left(1 - z^2\right),$$ (4.61)

and

$$f_{\text{low}}^{\text{Bethe}}(z) = \ln\frac{16z^2}{(1 + z)(1 - z)(1 + 6z - z^2)}.$$ (4.62)

Note that for $z \geq z_c^{\text{Bethe}}$, the low-temperature free energy is always bigger than the high-temperature free energy. Thus according to our selection principle (4.57) z_c^{Bethe} also corresponds to a phase transition in the Bethe approximation.

4.5.3 The Polymer Expansion for the Ising Model

According to Section 4.2, the difference between the Bethe free energy and the free energy is given by

$$f^{\text{Ising}}(z) - f_{\xi}^{\text{Bethe}}(z) = f_{\xi}^{\text{Polymer}}(z),$$ (4.63)

where

$$f_\xi^{\text{Polymer}} = \lim_{n \to \infty} \frac{1}{n} \sum_{M \geq 1} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M} \prod_{k=1}^{M} K_\xi(\gamma_k) \sum_{G \in \mathcal{G}_M} \prod_{(k,k') \in G} \left(\mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset) - 1 \right). \tag{4.64}$$

In (4.64), weights are given by the following formula

$$K_\xi(\gamma) = z^{|\gamma \cap E|} \prod_{i \in \gamma \cap V} \frac{(1 - \xi)^{q_i} (1 + z\xi)^{4-q_i} + (-1)^{q_i} (1 + z\xi)^{q_i} (1 - z\xi)^{4-q_i}}{(1 + z\xi)^4 + (1 - z\xi)^4}, \tag{4.65}$$

where $q_i = |\{ j \in V \mid (i,j) \in \gamma \cap E \}|$ is the induced degree in γ of the vertex i.

In the next subsection we compute the first terms of the expansion (4.64) which are the leading correction terms of the Bethe free energy. Up to the third order correction terms we find that the phase transition occurs at a lower temperature than predicted by the Bethe free energy[3]

$$z_c^{\text{Bethe+corr}} \approx 0.405. \tag{4.66}$$

This transition temperature is found by applying the selection principle of the fixed point (4.57) to the Bethe free energy to which we added the first correction terms from the polymer expansion. We define accordingly the corrected Bethe free energy as

$$f^{\text{Bethe+corr}} = \max_\xi \left(f_\xi + \text{corr}_\xi \right). \tag{4.67}$$

High-Temperature Expansion

At the high-temperature fixed-point (4.59), weights (4.65) become

$$K_{\text{high}}(\gamma) = z^{|\gamma \cap E|} \prod_{i \in \gamma \cap V} \frac{1 + (-1)^{q_i}}{2}. \tag{4.68}$$

According to (4.68) a polymer is any subgraph with vertices of even induced degree. The activity of a polymer is exponentially decreasing in the number of edges contained in the polymer.

Figure 4.4 shows the first polymers entering in the expansion. The multiplicity of a polymer described in Figure 4.4 is the ratio between the number of polymers of a particular shape (up to rotations and translations) and the number of vertices. Let us define more precisely the notion of multiplicity. We say that two polymers γ_1 and γ_2 have the same shape if there exists a transformation T composed of a translation and a rotation such that $\gamma_1 = T[\gamma_2]$. We clearly see that "having the same shape" defines

---

[3]In general mean-field methods neglect fluctuations and overestimate the critical temperature.

γ	$K_{\text{high}}(\gamma)$	multiplicity $\chi([\gamma])$	# of edges
	z^4	1	4
	z^6	2	6
	z^8	4	8
	z^8	2	8
	z^8	1	8
	z^8	2	8

Figure 4.4: Polymers entering in the first orders of the expansion at high temperature.

an equivalence relation between polymers. The interest of this equivalence relation is that polymers with the same shape give the same contribution in the polymer expansion. Therefore we can simplify the polymer expansion by considering only contributions given by a certain shape of polymer rather than contributions given by the polymers themselves. The multiplicity $\chi([\gamma])$ is a number associated to an equivalence class of polymers having the same shape. It represents in the thermodynamic limit the number of occurrence per variables of a certain shape in the lattice

$$\chi([\gamma]) = \lim_{n \to \infty} \frac{|[\gamma]|}{n}. \tag{4.69}$$

Using the polymer expansion (4.64) we compute the difference between the free energy and the Bethe free energy up to the first few small-order correction-terms

$$f^{\text{Ising}} - f_{\text{high}}^{\text{Bethe}} = z^4 + 2z^6 + \frac{9}{2}z^8 + O\left(z^{10}\right). \tag{4.70}$$

Low-Temperature Expansion

Similarly to the high-temperature case we compute the activity at the low-temperature fixed-point (4.60)

$$K_{\text{low}}(\gamma) = \prod_{i \in \gamma \cap V} \kappa_{q_i}, \tag{4.71}$$

where

$$\kappa_2 = \frac{(1-z)^2}{4z} = \frac{(1-z)^2}{4} + O\left((1-z)^3\right), \tag{4.72}$$

65

γ	$K_{\mathrm{low}}(\gamma)$	multiplicity $\chi\left([\gamma]\right)$	perimeter
	$\approx (1-z)^8$	1	8
	$\approx (1-z)^{10}$	2	10
	$\approx (1-z)^{12}$	2	12
	$\approx (1-z)^{12}$	4	12
	$\approx (1-z)^{12}$	2	12
	$\approx (1-z)^{12}$	1	12
	$\approx (1-z)^{12}$	2	12

Figure 4.5: Polymers entering in the first orders of the expansion at low temperature.

$$\kappa_3 = \frac{1-z^2}{8z^{3/2}}\sqrt{(3z-1)(3-z)} = \frac{1-z}{2} + O\left((1-z)^2\right), \qquad (4.73)$$

$$\kappa_4 = -\frac{1-10z^2+z^4}{8z^2} = 1 - \frac{(1-z)^2}{2} + O\left((1-z)^3\right). \qquad (4.74)$$

Polymers are subgraphs with no induced degree-one vertex. According to (4.71), the activity of a polymer is in the first order in $(1-z)$

$$K_{\mathrm{low}}\left(\gamma\right) = \left(\frac{1-z}{2}\right)^{4|V\cap\gamma|-2|E\cap\gamma|} + \text{high order terms}. \qquad (4.75)$$

But the exponent on the right-hand side of (4.75) is the number of edges attached to γ but not in γ. The activity of the polymer is exponentially decreasing in the perimeter size, where the perimeter are edges which connect the polymer to the rest of the lattice.

Figure 4.5 shows the first polymers entering in the low-temperature expansion. The difference between the free energy and the Bethe free energy in the first few small-order

Table 4.1: Phase transition temperature for the different approximations

z_c	z_c^{Bethe}	$z_c^{\mathrm{Bethe+corr}}$
0.414	0.333	0.405

correction-terms

$$
\begin{aligned}
f - f_{\mathrm{High}}^{\mathrm{Bethe}} =& \kappa_2^4 + 2\kappa_2^4\kappa_3^2 + \kappa_2^4\kappa_3^4\kappa_4 + 2\kappa_2^6 + 2\kappa_2^4\kappa_3^4 + 4\kappa_2^5\kappa_3^2\kappa_4 + 2\kappa_2^6\kappa_4 \\
\approx& \frac{1}{256}\left(1-z\right)^8 + \frac{1}{64}\left(1-z\right)^9 + \frac{21}{512}\left(1-z\right)^{10} + \frac{45}{512}\left(1-z\right)^{11} \\
&+ \frac{687}{4096}\left(1-z\right)^{12} + \frac{609}{2048}\left(1-z\right)^{13} + O\left(1-z\right)^{14}.
\end{aligned}
\tag{4.76}
$$

4.5.4 Comparison with the Exact Solution

We compare the value of the free energy (4.46), the internal energy (4.47), the entropy (4.48) and the heat capacity (4.49) obtained by the exact computation, by the Bethe free energy approximation and by the polymer expansion. Table 4.1 shows a recapitulation of the various phase transition temperatures predicted by the three approaches. In Figures 4.6, 4.7, 4.8, 4.9 and 4.10 the exact quantity is represented in black, the Bethe quantity is in blue and the Polymer corrections are in red.

Figure 4.6 depicts the various free energies. Figure 4.7 shows the relative difference between the exact solution and the approximations

$$
\frac{\Delta f^{\mathrm{approx}}}{f} = \frac{f - f^{\mathrm{approx}}}{f}.
\tag{4.77}
$$

In Figure 4.8 the free energy is reported. Figure 4.9 shows the entropy and Figure 4.10 shows the heat capacity.

The first few small-order correction-terms given by the polymer expansion already improves significantly the quantitative results predicted by the Bethe approximation. Notably, the relative error on the phase transition temperature predicted by the Bethe approximation is reduced from 24% to 2%.

4.5.5 On the Convergence of the Polymer Expansion

The mechanism which ensures that the polymer expansion is a convergent series is different between the high and the low-temperature regime.

In the high-temperature regime we can directly apply apply Corollary 4.1 by noticing
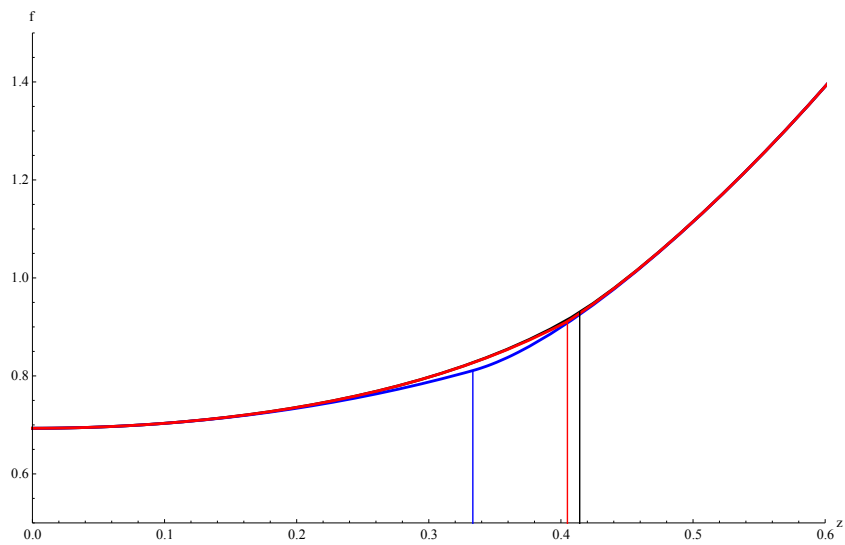
Figure 4.6: Free energy. The curves are: f in black, f^{Bethe} in blue and the corrections $f^{\text{Bethe+corr}}$ in red.
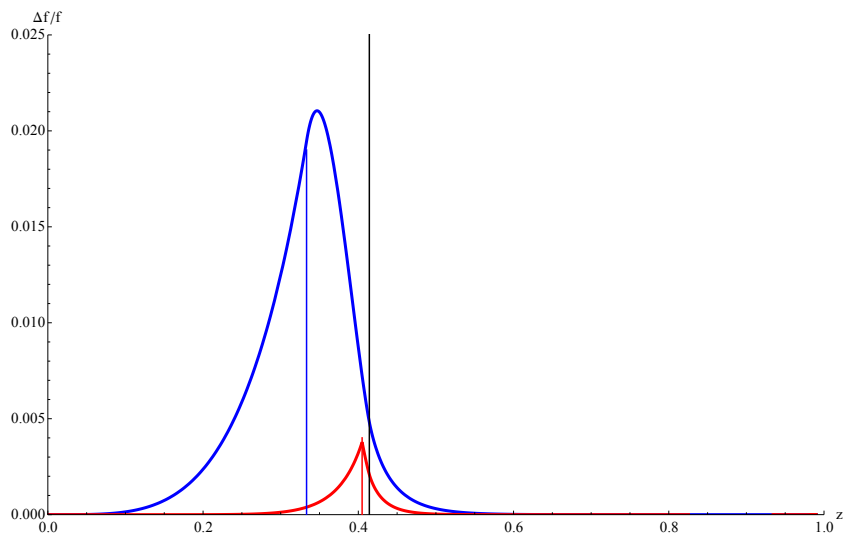
Figure 4.7: Free energy relative difference. The curves are $\frac{\Delta f^{\text{Bethe}}}{f}$ in blue and the corrections $\frac{\Delta f^{\text{Bethe+corr}}}{f}$ in red .
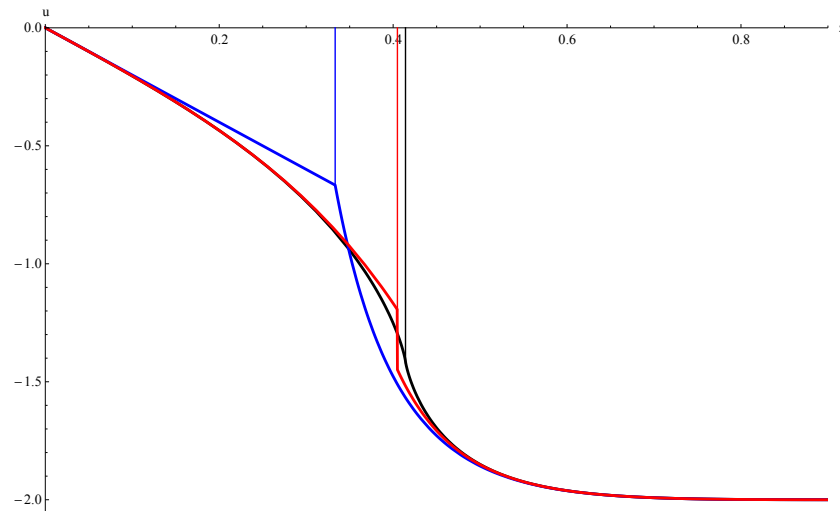
Figure 4.8: Internal energy. The curves are u in black, u^{Bethe} in blue and the corrections $u^{\text{Bethe+corr}}$ in red .
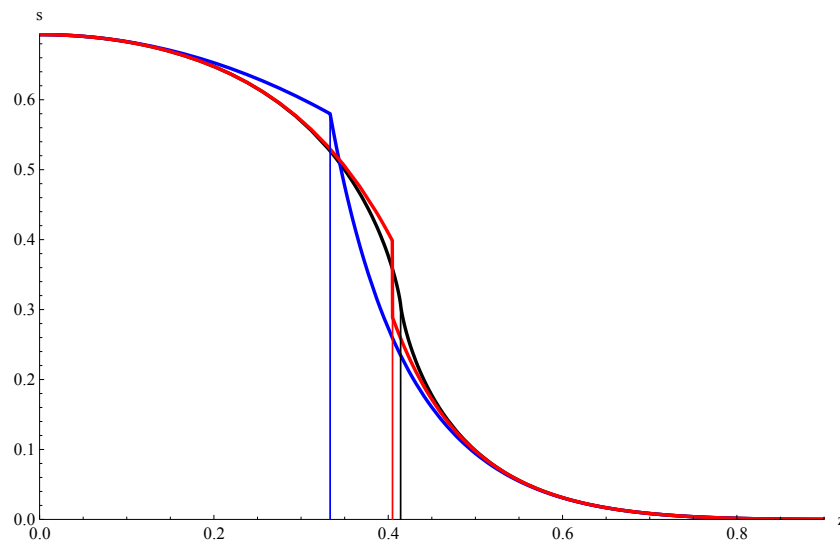
Figure 4.9: Entropy. The curves are s in black, s^{Bethe} in blue and the corrections $s^{\text{Bethe+corr}}$ in red.
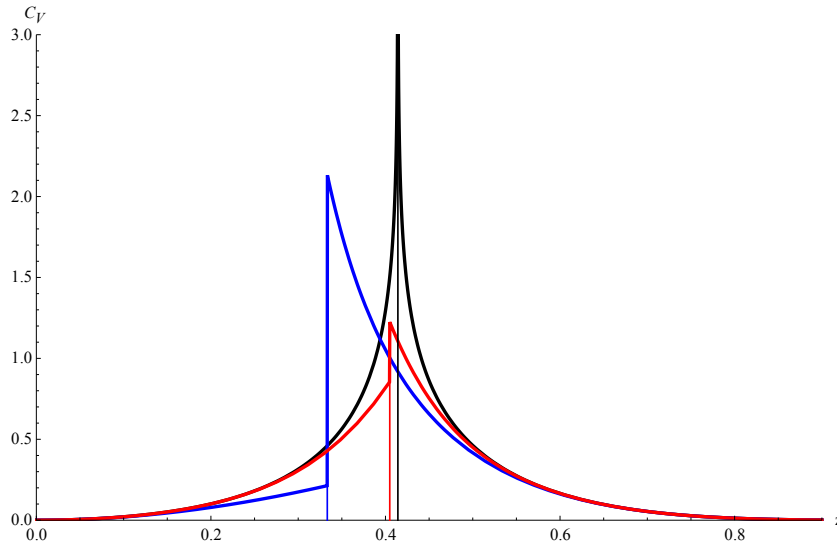
Figure 4.10: Heat capacity. The curves are C in black, C^{Bethe} in blue and the corrections $C^{\text{Bethe+corr}}$ in red.

with (4.68) that $|K_{\text{high}}(\gamma)| < z^{|\gamma \cap E|} < z^{|\gamma|}$. We thus find that for a temperature $z < \frac{1}{2}e^{-5} \approx 0.003$ the polymer expansion is an absolutely convergent series. This extremely small convergence radius contrasts with the result found in the previous subsection. Apparently from the comparison with the exact solution, it seems reasonable to think that the polymer series is convergent up to the phase transition $z < \sqrt{2} - 1$. However because of estimates made in Lemma 4.2 and more importantly the loose bound on the number of rooted polymers in Corollary 4.1, we cannot expect to ensure absolute convergence on a large interval.

The low-temperature regime is more challenging. The activity of polymers does not decrease fast enough with the system size uniformly in n (take for instance the activity of the whole lattice $K_{\text{low}}(\Lambda_n) \approx \left(1 - \frac{1}{2}(1-z)\right)^n$). We cannot directly apply Corollary 4.1 in this situation. However on the lattice, a polymer is almost completely determined by the position of the boundary nodes. Once the induced degree two nodes and three nodes are placed on the lattice, there are almost no other choice for placing the degree four nodes (They must "fill" the space between the degree two nodes and three nodes). It implies that the number of polymers containing a vertex $x \in V$ with a prescribed *boundary size* $|\partial \gamma| = \tau$ is exponentially increasing in τ. This combines with the fact that the activity is decreasing in the boundary size enables us to fulfill the convergence criterion of Lemma 4.2.

5 High-Temperature Models

In this chapter[1], we prove that the Bethe free energy is exact for high-temperature models.

In Section 5.1, we sketch the general strategy of the proof.

In Section 5.2, we show that the polymer expansion is absolutely convergent for general models provided that the temperature is high enough. A direct application of this result is the proof that the Bethe free energy is exact at high temperatures for factor graphs with large girth (Theorem 5.1 on page 75).

In Section 5.3, we apply the result of Section 5.2 to LDGM codes. We prove that the free energy of an instance drawn at random from an irregular LDGM code ensemble is, with high probability, given by the Bethe formula in the large block length limit (Theorem 5.2 on page 79).

5.1 Preamble

We recall that the partition function of a general graphical models for binary variables is

$$Z = \sum_{\underline{s} \in \{-1,+1\}^n} \prod_{a \in C} \psi_a(s_{\partial a}), \tag{5.1}$$

with weights

$$\psi_a(s_{\partial a}) = \exp\{\beta \sum_{I \subset \partial a} J_I \prod_{i \in I} s_i\}.$$

All the correction terms to the Bethe free energy are contained in the free energy of the

---

[1]The material of this chapter is based on [88].

polymer gas, namely

$$\frac{1}{n}\ln Z = f^{\text{Bethe}} + \frac{1}{n}\ln Z^{\text{polymer}}. \tag{5.2}$$

If a suitable fixed point of the belief propagation equations is chosen such that the Bethe free energy is a good approximation, then we expect that the polymer free energy is small (or negligible in an appropriate limit). One way that this could happen is if the activities of the polymers become small as the size of the polymers increase. Let us explain this point in more detail. We expect the activities to be exponentially small in $|\gamma|$ (as will become clear later for LDGM codes this is true for small $h = \frac{1}{2}\ln\frac{1-p}{p}$). For h small enough the smallness of the activities enables us to expand thelog in a power series in $K(\gamma)$. Since the polymers have no dangling edges, on a locally tree-like graph they have a typical size $|\gamma| \approx c\ln n$ for some small constant c. This means $K(\gamma) \approx O(e^{-c\ln n})$ and since the series expansion starts linearly with $K(\gamma)$, the polymer free energy is itself $O(e^{-c\ln n})$. Note that the polymer free energy could still be negligible even if the activities are not small because in general they have signs and cancellations could occur. However such cancellations would be difficult to control. The regimes investigated in the present chapter are those where the activities are small enough so that their weight counterbalances the entropy of the polymers and we do not need to track sign cancellations.

5.2 Exactness of the Free Energy at High Temperature

When we speak of the high-temperature regime it is meant that the coupling constants J_I are finite and $\beta > 0$ is small enough so that

$$\mu := 2\beta \sup_{a \in C} \sum_{I \subset \partial a} |J_I| < 1. \tag{5.3}$$

We remark for later use that in the high-temperature regime

$$\left|\psi_a(\{s_{\partial a}\}) - 1\right| \leq \mu. \tag{5.4}$$

It will become clear that for LDGM codes the high-temperature regime is equivalent to large noise. However for LDPC codes this is not true because these codes essentially correspond to have $J_I = +\infty$.

We begin with the high-temperature regime of general factor graph models. It has been proven [101] that when the high-temperature condition (5.4) is satisfied the belief propagation equations have a unique fixed-point solution. We assume in the present chapter that the Bethe free energy is computed at this fixed point.

Lemma 5.1 (High-temperature bound on the activities). For $\mu < \left(2l_{\max}^2 r_{\max}\right)^{-1}$, the

polymers activity is bounded by

$$|K(\gamma)| \leq (6e\mu)^{2|g|/(2+r_{\max})}.$$

Proof. In order to find bounds on the activities (4.14) and (4.15), we should control the behavior of the belief propagation messages. This is realized through the BP equations (4.17) and (4.18). We first choose to parametrize the BP distributions $\nu, \widehat{\nu}$ with real numbers $\zeta, \widehat{\zeta}$

$$\widehat{\nu}_{ai}(s_i) = \frac{1 + s_i \tanh \widehat{\zeta}_{a \to i}}{2} \text{ and } \nu_{ia}(s_i) = \frac{1 + s_i \tanh \zeta_{i \to a}}{2}. \tag{5.5}$$

The BP equation (4.17) now reads

$$\tanh \widehat{\zeta}_{a \to i} = \frac{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) s_i \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}. \tag{5.6}$$

Using the high-temperature condition (5.4) leads to the following bound

$$
\begin{aligned}
|\tanh \widehat{\zeta}_{a \to i}| &= \frac{\left| \sum_{s_{\partial a}} (\psi_a(s_{\partial a}) - 1 + 1) s_i \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j) \right|}{1 + \sum_{s_{\partial a}} (\psi_a(s_{\partial a}) - 1) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)} \\
&\leq \frac{\left| \sum_{s_{\partial a}} (\psi_a(s_{\partial a}) - 1) s_i \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j) \right|}{1 - \left| \sum_{s_{\partial a}} (\psi_a(s_{\partial a}) - 1) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j) \right|} \\
&\leq \frac{\sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}{1 - \sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)} \\
&\leq \frac{\mu}{1 - \mu} \\
&\leq 2\mu, \tag{5.7}
\end{aligned}
$$

where in the last line, we use the fact that $\mu < 1/2$.

The other BP equation (4.18) takes the form

$$\tanh \zeta_{i \to a} = \tanh \left(\sum_{b \in \partial i \setminus a} \widehat{\zeta}_{b \to i} \right). \tag{5.8}$$

Using the bound (5.7) on messages $\widehat{\zeta}_{a \to i}$ gives

$$
\begin{aligned}
|\tanh \zeta_{i \to a}| &\leq \tanh \left((l_{\max} - 1) \tanh^{-1}(2\mu) \right) \\
&\leq 2(l_{\max} - 1)\mu. \tag{5.9}
\end{aligned}
$$

The inequalities (5.7) and (5.9) can be restated in terms of distributions $\widehat{\nu}, \nu$ and take

the form

$$\begin{cases} \frac{1-2\mu}{2} \le \widehat{\nu}_{ai}(s_i) \le \frac{1+2\mu}{2} \\ \frac{1-2(l_{\max}-1)\mu}{2} \le \nu_{ia}(s_i) \le \frac{1+2(l_{\max}-1)\mu}{2}. \end{cases} \tag{5.10}$$

By noticing that $\sum_s s\widehat{\nu}_{ai}(s) = \tanh \widehat{\zeta}_{a\to i}$ and using the bound (5.7), we are in position to control the activity (4.15)

$$\begin{aligned} |K_a(\gamma)| &\le \frac{\sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{i\in\partial a\backslash\gamma} \nu_{ia}(s_i) \prod_{i\in\partial a\cap\gamma} \widehat{\nu}_{ai}(-s_i)}{1 - \sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{i\in\partial a} \nu_{ia}(s_i)} \\ &\quad + \frac{\prod_{i\in\partial a\cap\gamma} |\sum_{s_i} s_i\widehat{\nu}_{ai}(s_i)|}{1 - \sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{i\in\partial a} \nu_{ia}(s_i)} \\ &\le \frac{\mu + (2\mu)^{|\partial a\cap\gamma|}}{1 - \mu} \\ &\le 6\mu, \end{aligned} \tag{5.11}$$

where in the last line we use the fact that subgraphs γ have no dangling edges (i.e., $|\partial a \cap \gamma| \ge 2$) and $\mu \le 1/2$.

The second activity (4.14) is directly controlled using bounds on distributions $\widehat{\nu}$ and ν given by equations (5.10)

$$\begin{aligned} |K_i(\gamma)| &\le \frac{\sum_{s_i} \prod_{a\in\partial i\backslash\gamma} |\widehat{\nu}_{ai}(s_i)| \prod_{a\in\partial i\cap\gamma} |\nu_{ia}(-s_i)|}{|\sum_{s_i} \prod_{a\in\partial i} \widehat{\nu}_{ai}(s_i)|} \\ &\le \left(\frac{1 + (l_{\max} - 1)2\mu}{1 - 2\mu}\right)^{l_{\max}} \\ &\le (1 + 4l_{\max}\mu)^{l_{\max}}. \end{aligned} \tag{5.12}$$

The total activity of a polymerγ, given by the relation (4.13), is then bounded by

$$\begin{aligned} |K(\gamma)| &\le \prod_{i\in\gamma\cap V} |K_i| \prod_{a\in\gamma\cap C} |K_a| \\ &\le \exp\left(|\gamma\cap V|l_{\max}\ln(1 + 4l_{\max}\mu) + |\gamma\cap C|\ln(6\mu)\right). \end{aligned} \tag{5.13}$$

There are two antagonistic contributions in the loops activities. One is exponentially increasing in the number of variable nodes contained in γ. The other is exponentially decreasing in the number of check nodes contained in γ. We recall that the size of a subgraph γ, denoted by $|\gamma|$, is the total number of variable and check nodes contained in the loop

$$|\gamma| := |\gamma\cap C| + |\gamma\cap V|. \tag{5.14}$$

In order to show that the activities in (5.13) are exponentially decreasing in the polymer size, we have to show that the number of variable nodes contained in a polymer cannot

be arbitrarily larger than the number of check nodes. Consider the number of edges contained in a subgraph. We can bound from above this number counted from the check node perspective and we can find a lower bound counted from the variable node perspective. This leads to the following bound on the number of variable nodes

$$r_{\max}|\gamma \cap C| \geq 2|\gamma \cap V|. \tag{5.15}$$

Using the definition (5.14) and the bound (5.15), we find that for every non-negative numbers p and q

$$|\gamma \cap V|p - |\gamma \cap C|q = -(p+q)|\gamma \cap C| + p|\gamma|$$
$$\leq |\gamma|\frac{r_{\max}p - 2q}{2 + r_{\max}} \tag{5.16}$$

This implies the upper bound for the exponent in (5.13)

$$\frac{|\gamma|}{2 + r_{\max}} \ln\left((6e\mu)^2 (1 + 4l_{\max}\mu)^{r_{\max}l_{\max}} \right). \tag{5.17}$$

Moreover for $\mu < 1/2l_{\max}^2 r_{\max}$ we have

$$(6\mu)^2 (1 + 4r_{\max}\mu)^{r_{\max}l_{\max}} \leq (6e\mu)^2. \tag{5.18}$$

From (5.13), (5.17) and (5.18) we deduce the bound on the activities

$$|K(\gamma)| \leq (6e\mu)^{2|\gamma|/(2+r_{\max})}. \tag{5.19}$$

$$\square$$

Theorem 5.1 (Exactness of the Bethe free energy: High temperature). *Let Γ_n be a sequence of Tanner graphs, with uniformly bounded degrees and with large girth in the sense that $\mathrm{girth}(\Gamma_n) \geq C \ln |\Gamma_n|$ where $C > 0$ is a numerical constant. Consider the free energy sequences of models constructed on Γ_n. For $0 < \mu < e^{-\max(l_{\max}, r_{\max})^2 - 7}$ we have*

$$\lim_{n \to +\infty} |f - f^{\mathrm{Bethe}}(\underline{\nu}, \underline{\nu})| = 0. \tag{5.20}$$

Remark 5.1. *Even if the individual limits of f and f^{Bethe} are not well defined their difference tends to zero. As will be seen in the proof, the order of magnitude of this difference is $O((c\mu)^{2\mathrm{girth}/(2+r_{\max})})$ with $c > 0$ a constant depending only on the degrees of the nodes and the couplings J_I.*

Proof. Let n the number of variable nodes of the graph Γ_n. Proving Equ. (5.20) is equivalent to

$$\lim_{n \to +\infty} \frac{1}{n} \ln Z^{\mathrm{polymer}} = 0. \tag{5.21}$$

To show this we apply Corollary 4.1 for $\epsilon = (6e\mu)^{\frac{2}{2+r_{\max}}} e^{1+\max(l_{\max}, r_{\max})}$, Lemma 5.1 ensuring that the activities are decreasing fast enough. The graphs Γ_n have large girth, and since a polymer $\gamma \subset \Gamma_n$ containing x certainly contains at least one closed cycle, we have $|\gamma| \geq C \ln n$ (for $C > 0$ not too large). Using this fact we find

$$
\begin{aligned}
\frac{1}{n} |\ln Z^{\text{polymer}}| &\leq \frac{4}{n} \sum_{x \in V \cup C} |\epsilon|^{C \ln n} \\
&\leq 4 \left(1 + \frac{l}{r}\right) |\epsilon|^{C \ln n}
\end{aligned}
\tag{5.22}
$$

Clearly this estimate tends to zero as $n \to +\infty$ for $\mu < e^{-\max(l_{\max}, r_{\max})^2 - 7}$. In fact we have that $n^{-1} \ln \left| Z^{\text{polymer}} \right| = O(\mu^{2\text{girth}/2 + r_{\max}})$. $\qquad \square$

5.3 Application to LDGM Codes

5.3.1 Preamble

For a graph Γ from a standard ensemble $\text{LDGM}(\Lambda, P)$ [59] the fraction of variable nodes of degree $1 \leq s \leq l_{\max}$ is $\Lambda_s \geq 0$, and the fraction of check nodes with degree $1 \leq t \leq r_{\max}$ is $P_t \geq 0$. Of course $\sum_{s=1}^{l_{\max}} \Lambda_s = \sum_{t=1}^{r_{\max}} P_t = 1$. Here Γ is the Tanner graph of an LDGM code with design rate $r/l = n/m$, where l and r are the average variable and check node degrees, respectively. The large block length limit corresponds to $n, m \to \infty$ with fixed design rate. When we say that Γ is random we mean that we draw it uniformly from some specified ensemble. The corresponding expectation and probability are denoted by $\mathbb{E}_\Gamma$, $\mathbb{P}_\Gamma$.

We transmit codewords from an LDGM code with Tanner graph Γ and uniform prior over a BSC with flip probability p. Here information bits $\underline{u} = (u_i)_{i=1}^n$ are attached to variable nodes V and codewords are given by $\underline{x} = (x_a)_{a=1}^m$ with

$$
x_a = \oplus_{i \in \partial a} u_i \,.
\tag{5.23}
$$

We must have $n < m$ and $l > r$ so that the design rate r/l is well defined. We can assume without loss of generality that the all-zero codeword is transmitted. The posterior probability that $\underline{x} = (x_i)_{i=1}^n \in \{0, 1\}^n$, or equivalently $\underline{u} = (u_a)_{a=1}^m$, is transmitted given that $\underline{y} = (y_a)_{i=1}^n \in \{0, 1\}^n$ is received, reads

$$
p_{\underline{U}|\underline{Y}} \left(\underline{u}|\underline{y}\right) = \frac{1}{Z_{\text{LDGM}}} \prod_{a \in C} e^{h_a \prod_{i \in \partial a} (-1)^{u_i}} \,.
\tag{5.24}
$$

In this expression

$$
h_a = (-1)^{y_a} \frac{1}{2} \ln \frac{1-p}{p}
\tag{5.25}
$$

are the half-log-likelihood ratios and

$$Z_{\text{LDGM}} = \sum_{\underline{u} \in \{0,1\}^n} \prod_{a \in C} e^{h_a \prod_{i \in \partial a} (-1)^{u_i}} \tag{5.26}$$

is the partition function. The amplitude of h_a is set to

$$|h_a| \equiv h \equiv \frac{1}{2} \ln \frac{1-p}{p}. \tag{5.27}$$

It is good to keep in mind that the high-noise regime – p close to $1/2$ – corresponds to small h. It is equivalent to describe the channel outputs $\underline{y}$ in terms of the half-log-likelihood ratios $\underline{h} = (h_a)_{a=1}^m$ which are i.i.d. with probability distribution

$$q(h_a) = (1-p)\delta(h_a - h) + p\delta(h_a + h). \tag{5.28}$$

The expectation with respect to this distribution is called $\mathbb{E}_{\underline{h}}$.

Remark 5.2. *Equ. (5.26) is the partition function of a spin system with one coupling constant $\beta J_I \to h_a$ per check, and the high-temperature regime (5.4) simply corresponds to $h \ll 1$.*

The free energy for fixed $(\Gamma, \underline{y})$ is

$$f_{\text{LDGM}} = \frac{1}{n} \ln Z_{\text{LDGM}} \tag{5.29}$$

For communications, the importance of this quantity stems from the fact that it is intimately related to the Shannon conditional entropy by the simple formula,

$$\frac{1}{n} H_{\text{LDGM}}(\underline{U}|\underline{Y}) = \mathbb{E}_{\underline{h}}[f_{\text{LDGM}}] - \frac{l}{r} \frac{1-2p}{2} \ln \frac{1-p}{p}. \tag{5.30}$$

5.3.2 The Bethe Free Energy for LDGM Codes

We parametrize the BP messages by reals numbers

$$\zeta_{i \to a} = \sum_{\sigma_i} \sigma_i \nu_{ia}(\sigma_i) \quad \text{and} \quad \widehat{\zeta}_{a \to i} = \sum_{\sigma_i} \sigma_i \widehat{\nu}_{ai}(\sigma_i). \tag{5.31}$$

It is immediate to specialize the Bethe formulas to LDGM codes. The Bethe free energy (4.5) is given by a sum of three quantities

$$f_{\text{LDGM}}^{\text{Bethe}}(\underline{\zeta}, \widehat{\underline{\zeta}}) = \sum_{a \in C} F_a + \sum_{i \in V} F_i - \sum_{(i,a) \in E} F_{ia}, \tag{5.32}$$

77

where

$$
\begin{cases}
F_a = \ln\{1 + \tanh h_a \prod_{i \in \partial a} \tanh \zeta_{i \to a}\} + \ln \cosh h, \\[3mm]
F_i = \ln\{\prod_{a \in \partial i}(1 + \tanh \zeta_{a \to i}) + \prod_{a \in \partial i}(1 - \tanh \zeta_{a \to i})\} \\[3mm]
F_{ia} = \ln\{1 + \tanh \zeta_{i \to a} \tanh \widehat{\zeta}_{a \to i}\}.
\end{cases}
\tag{5.33}
$$

The BP equations (4.4) become

$$
\begin{cases}
\zeta_{i \to a} = \sum_{b \in \partial i \setminus a} \widehat{\zeta}_{b \to i} \\[3mm]
\widehat{\zeta}_{a \to i} = \tanh^{-1}(\tanh h_a \prod_{j \in \partial a \setminus i} \tanh \zeta_{j \to a}).
\end{cases}
\tag{5.34}
$$

If the LDGM code contains no degree-one check nodes then $(\underline{\zeta}, \underline{\widehat{\zeta}}) = (\underline{0}, \underline{0})$ is the trivial fixed point. However if there is a non-vanishing fraction of degree-one check nodes the fixed point $(\underline{\zeta}, \underline{\widehat{\zeta}})$ is non-trivial.

5.3.3   Exactness of the Bethe Free Energy for LDGM Codes

For h small enough, an instance of an LDGM code is a particular high-temperature graphical-model.

Lemma 5.2. *For $h < (4l_{\max}^2 r_{\max})^{-1}$, the polymers activity is bounded by*

$$
|K(\gamma)| \le (12eh)^{2|\gamma|/(2+r_{\max})}.
$$

Proof. The LDGM codes can be seen as a special case of the high-temperature general-models with $I \to \partial a$ and $\beta J_I \to h_a$. The high-temperature condition translates into $2\sup_a |h_a| = \mu \ll 1$. Recalling that $|h_a| = h = \frac{1}{2} \ln \frac{1-p}{p}$, we see that the high-temperature condition is equivalent to taking p close to $1/2$. The bound on the activity is obtained by applying (5.19)

$$
|K(\gamma)| \le \left(6e \ln \frac{1-p}{p}\right)^{2|\gamma|/(2+r_{\max})}.
\tag{5.35}
$$

The activities of the LDGM codes have a high-temperature bound and the high-noise regime $p \approx 1/2$ is then similar to a high-temperature regime for general models. $\square$

Remark 5.3. *There is a simplification for LDGM ensembles with no degree-one check nodes. In this case the BP equations admit a trivial fixed point where $(\underline{\zeta}, \underline{\widehat{\zeta}}) = (\underline{0}, \underline{0})$. The*

activities at the trivial fixed point can be computed exactly

$$K_a^{trivial}(g) = \begin{cases} \tanh h_a & \text{if } \partial a \cap g = \partial a \\ 0 & \text{otherwise,} \end{cases} \tag{5.36}$$

and

$$K_i^{trivial}(g) = \begin{cases} 1 & \text{if } |\partial a \cap g| \text{ is even} \\ 0 & \text{otherwise.} \end{cases} \tag{5.37}$$

Subgraphs contributing in the loop sum are only those which have check nodes with maximal induced degree and variable nodes with odd degree. Their activities admit the simple bound

$$|K^{trivial}(g)| \le (1-2p)^{|g \cap C|} \le (1-2p)^{\frac{2|g|}{2+r_{max}}} \tag{5.38}$$

Theorem 5.2 (Exactness of the Bethe free energy: LDGM at high noise)**.** *Suppose that we draw Γ uniformly at random from the ensemble $LDGM(\Lambda, P, n)$. For $h < e^{-\max(l_{max}, r_{max})^2 - 7}$ we have*

$$\lim_{n \to +\infty} \mathbb{E}_\Gamma \left[|f_{LDGM} - f_{LDGM}^{Bethe}(\underline{\zeta}, \widehat{\underline{\zeta}})| \right] = 0. \tag{5.39}$$

Proof. Set $\epsilon = (12eh)^{\frac{2|\gamma|}{2+r_{max}}} e^{-\max(l_{max}, r_{max}) - 1}$. For h small enough Corollary 4.1 implies

$$\frac{1}{n} |\ln Z_{LDGM}^{polymer}| \le \frac{4}{n} \sum_{x \in V \cup C} \epsilon^{R_x}. \tag{5.40}$$

We recall that R_x is the shortest size of a polymer containing x. Taking the expectation of this inequality,

$$\mathbb{E}_\Gamma \left[\frac{1}{n} |\ln Z_{LDGM}^{polymer}| \right] \le 4 \left(1 + \frac{l_{max}}{r_{max}} \right) \frac{1}{|V \cup C|} \sum_{x \in V \cup C} \mathbb{E}_\Gamma \left[\epsilon^{R_x} \right]. \tag{5.41}$$

Given Γ, let $N_D(x)$ be subgraphs formed by the set of nodes that are at distance less than D from x. For the moment D is a fixed number. For D fixed and n large enough, this subgraph is a tree with probability

$$1 - O(\frac{C_{l_{max}, r_{max}, D}}{n}), \tag{5.42}$$

where $C_{l_{max}, r_{max}, D} > 0$ depends only on D and the maximal degrees. This means that for n large enough the polymers $\gamma \ni x$ have a size $|\gamma| \ge D$. Thus for D fixed and n large

enough

$$\frac{1}{|V \cup C|} \sum_{x \in V \cup C} \mathbb{E}_\Gamma \left[\epsilon^{R_x} \right] \leq \epsilon^D (1 - O(\frac{C_{l_{\max}, r_{\max}, D}}{n})) + \epsilon^2 O(\frac{C_{l_{\max}, r_{\max}, D}}{n}) \qquad (5.43)$$

Replacing this estimate in (5.41) and taking the limit $n \to +\infty$,

$$\lim_{n \to +\infty} \mathbb{E}_\Gamma \left[\frac{1}{n} | \ln Z_{\text{LDGM}}^{\text{polymer}} | \right] \leq 4 \left(1 + \frac{l_{\max}}{r_{\max}} \right) \epsilon^D. \qquad (5.44)$$

Finally taking the limit $D \to \infty$ ends the proof. $\qquad \square$

6 LDPC Codes

In this chapter[1], we prove that the Bethe free energy is exact for LDPC codes over a BSC at high noise.

In Section 6.1, we set up the notation and recall the relation between the conditional entropy of a code and its free energy.

In Section 6.2, we compute the free energy of an LDPC code and we derive the associated BP equations. We show that the activity of a "big" polymer does not necessarily decrease with increasing size. We separate in the loop sum contributions from "small" and "big" polymers.

In Section 6.3, we show that, by using expander properties of typical instances from the LDPC ensembles, the polymer expansion truncated to "small" polymers is uniformly convergent (Corollary 6.1 on page 89).

In Section 6.4, we estimate the contribution from "big" polymers by using a combinatorial counting method.

In Section 6.5, we prove that the convergent truncated polymer expansion accounts for the biggest part of the corrections to the Bethe free energy, up to a remainder of order $O(e^{-n\epsilon})$, $\epsilon > 0$ (Theorem 6.1 on page 94). Finally, we prove that the Bethe free energy is with high probability asymptotically exact in the large size limit (Theorem 6.2 on page 96).

In Section 6.6, we discuss some remaining open problems.

---

[1]The material of this chapter is based on [88].

6.1 Settings

For LDPC codes we will limit ourselves to regular code ensembles with variable-node degree l and check-node degree r. Instead of using the notation $\text{LDPC}(z^l, z^r, n)$, we find it more convenient to denote the regular ensemble by $\text{LDPC}(l, r, n)$. The Tanner graph of a regular LDPC code is a bipartite (l, r) regular graph – call it Γ – connecting V and C. In other words the vertices of V have degree l, vertices of C have degree r, and there are no double edges. We recall that the design rate of a LDPC code is $1 - l/r = 1 - m/n$. The large block length limit again corresponds to $n, m \to +\infty$ with fixed design rate.

We transmit code words with uniform prior, from an LDPC code with Tanner graph Γ, over a BSC with flip probability p. Here $n > m$ and $l < r$ so that the design rate $1 - l/r$ is well defined. We assume without loss of generality that the all zero codeword is transmitted. Then the posterior probability that $\underline{x} = (x_i)_{i=1}^n \in \{0, 1\}^n$ is the transmitted word given that $\underline{y} = (y_i)_{i=1}^n \in \{0, 1\}^n$ is received, reads

$$\mathbb{P}_{\underline{X}|\underline{Y}}\left(\underline{x}|\underline{y}\right) = \frac{1}{Z_{\text{LDPC}}} \prod_{a \in C} \mathbb{I}\left(\oplus_{i \in \partial a} x_i = 0\right) \prod_{i \in V} \exp((-1)^{x_i} h_i). \tag{6.1}$$

In this formula

$$h_i = (-1)^{y_i} \frac{1}{2} \ln \frac{1-p}{p} \tag{6.2}$$

are the half-log-likelihood ratios and the normalizing factor Z is the partition function

$$Z_{\text{LDPC}} = \sum_{\underline{x} \in \{0,1\}^n} \prod_{a \in C} \mathbb{I}\left(\oplus_{i \in \partial a} x_i = 0\right) \prod_{i \in V} \exp((-1)^{x_i} h_i). \tag{6.3}$$

As before the amplitude of h_i is set to $|h_i| \equiv h \equiv \frac{1}{2} \ln \frac{1-p}{p}$ and the high-noise regime - p close to $1/2$ - corresponds to small h. The distribution of the i.i.d. half-log-likelihood ratios is $q(h_i) = (1-p)\delta(h_i - h) + p\delta(h_i + h)$.

Remark 6.1. *Equ. (6.3) is the partition function of a spin system with two types of coupling constants $\beta J_I \to h_i$ and $+\infty$. The infinite coupling constant mimics the parity-check constraints, so the high-temperature condition (5.4) is never satisfied which makes the ensuing analysis more challenging.*

The Shannon conditional entropy $H_{\text{LDPC}}(\underline{X}|\underline{Y})$ of the input word given the output word y is again directly related to the free energy

$$f_{\text{LDPC}} = \frac{1}{n} \ln Z_{\text{LDPC}} \tag{6.4}$$

through the formula

$$\frac{1}{n} H_{\text{LDPC}}\left(\underline{X}|\underline{Y}\right) = \mathbb{E}_{\underline{h}}[f_{\text{LDPC}}] - \frac{1-2p}{2} \ln \frac{1-p}{p}. \tag{6.5}$$

6.2 The Polymer Expansion for LDPC Codes

6.2.1 The Bethe Free Energy of LDPC Codes

We parametrize the BP messages by reals numbers $(\underline{\eta}, \underline{\widehat{\eta}})$, i.e.,

$$\eta_{i \to a} = \sum_{\sigma_i} \sigma_i \nu_{ia}(\sigma_i) \text{ and } \widehat{\eta}_{a \to i} = \sum_{\sigma_i} \sigma_i \widehat{\nu}_{ai}(\sigma_i). \tag{6.6}$$

The associated Bethe free energy (4.5) is

$$f_{\text{LDPC}}^{\text{Bethe}}(\underline{\eta}, \underline{\widehat{\eta}}) = \frac{1}{n}\left(\sum_{a \in C} P_a + \sum_{i \in V} P_i - \sum_{(i,a) \in E} P_{ia} \right), \tag{6.7}$$

where

$$\begin{cases} P_a = \ln\{1 + \prod_{i \in \partial a} \tanh \eta_{i \to a}\} - \ln 2, \\ \\ P_i = \ln\{e^{h_i} \prod_{a \in \partial i}(1 + \tanh \eta_{a \to i}) + e^{-h_i} \prod_{a \in \partial i}(1 - \tanh \eta_{a \to i})\} \\ \\ P_{ia} = \ln\{1 + \tanh \eta_{i \to a} \tanh \widehat{\eta}_{a \to i}\}. \end{cases} \tag{6.8}$$

According to equations (4.4), BP messages satisfy the belief propagation fixed-point equations

$$\begin{cases} \eta_{i \to a} = h_i + \sum_{b \in \partial i \setminus a} \widehat{\eta}_{b \to i} \\ \\ \widehat{\eta}_{a \to i} = \tanh^{-1}(\prod_{j \in \partial a \setminus i} \tanh \eta_{j \to a}). \end{cases} \tag{6.9}$$

The BP equations (6.9) always admit the trivial solution $\tanh \eta_{a \to i} = 1$, $\tanh \widehat{\eta}_{i \to a} = 1$ for all noise levels. Thus unlike the high-temperature cases, the BP equations of LDPC codes are not sufficient to control the BP fixed points. We need an extra requirement on the class of fixed point used in the loop expansion, called high-noise fixed-points.

Given $\epsilon > 0$, we say that a fixed point $(\underline{\eta}, \underline{\widehat{\eta}})$ is an ϵ high-noise fixed-point if for all $(i, a) \in E$

$$|\tanh \eta_{i \to a}| \leq \theta. \tag{6.10}$$

where

$$\theta = (1 + \epsilon) \tanh h. \tag{6.11}$$

In the remainder of this chapter we work under the following hypothesis

Hypothesis 6.1 (Existence of high-noise fixed-points). *We assume that for $h < h_0$ small enough independent of n and for $\epsilon > 0$ independent of n and h, there exists a solution $(\underline{\eta}, \underline{\widehat{\eta}})$ of the belief propagation equations (6.9) which satisfies the high-noise condition (6.10).*

The analysis does not require the uniqueness of this fixed point but only its existence. We call such fixed points "high-noise solutions."

The condition (6.10) can be justified by looking at the Taylor series expansion of a solution at high noise. For $h = 0$, the BP equations (6.9) admit the simple solution $\tanh \eta_{a \to i} = 0$, $\tanh \widehat{\eta}_{i \to a} = 0$. If we compute the Taylor expansion of this solution with respect to the noise parameter, we find

$$\begin{cases} \tanh \widehat{\eta}_{a \to i} = \prod_{j \in \partial a \setminus i} \tanh h_j \\ \tanh \eta_{i \to a} = \tanh h_i + \sum_{b \in \partial i \setminus a} \prod_{j \in \partial a \setminus i} \tanh h_j, \end{cases} \tag{6.12}$$

plus some terms of order $\mathcal{O}((\tanh h)^r)$. This shows that there exists a $h_0(\epsilon, n)$ such that the high-noise condition (6.10) is satisfied for $h < h_0(\epsilon, n)$. However it does not guarantee that $h_0(\epsilon, n)$ is uniform in the size of the graph.

6.2.2   Bound on the Activity

We recall that the induced degree of check and variable nodes are denoted by $d_a(g) = |\partial a \cap g|$ and $d_i(g) = |\partial i \cap g|$, respectively. The number of check nodes and variable nodes with prescribed induced degree are denoted by $n_s(g) = |\{i \in g \cap V | d_i(g) = s\}|$ and $m_t(g) = |\{a \in g \cap C | d_a(g) = t\}|$, respectively.

Lemma 6.1 (High-noise activity bound for LDPC codes). *For fixed numerical constants α, β that we can take close to one, there exist $\theta_0 > 0$ such that for all $\theta < \theta_0$*

$$|K(g)| \leq \overline{K}(\underline{n}(g), \underline{m}(g)), \tag{6.13}$$

where

$$\begin{aligned} \overline{K}(\underline{n}(g), \underline{m}(g)) = {} & (1 + \alpha \theta^r)^{m_r(g)} \prod_{t=2}^{r-1} \left(\alpha \theta^{r-t} \right)^{m_t(g)} \\ & \times \prod_{\substack{s=2, \\ \text{even}}}^{l-1} \left(1 + \frac{\beta}{2}(1 + 4s + s^2)\theta^2 \right)^{n_s(g)} \\ & \times \prod_{\substack{s=3, \\ \text{odd}}}^{l} (\beta(1 + s)\theta)^{n_s(g)}. \end{aligned} \tag{6.14}$$

Proof. The LDPC codes cannot be seen as high-temperature models. The LDPC partition function (6.3) is composed of two type of factors. The variable-node factors, coming from channel observations, satisfy the high-temperature condition at high-noise. However the check-node factors are equal to zero whenever neighboring bits does not satisfy the parity-check constraints. Thus the check-node factors cannot satisfy the high-temperature condition (5.4).

By using the high-noise condition (6.10) along with the BP equations (6.9), we find the reciprocal bound on messages from check nodes to variable nodes

$$|\tanh \widehat{\eta}_{a \to i}| \leq \theta^{r-1}. \tag{6.15}$$

For LDPC codes, the activities associated with check nodes (4.15) are

$$K_a(g) = \frac{u_a + (-1)^{d_a(g)} v_a}{1 + u_a w_a}, \tag{6.16}$$

where

$$\begin{cases} u_a &= \prod_{i \in \partial a \setminus g} \tanh \eta_{i \to a} \\ v_a &= \prod_{i \in \partial a \cap g} \tanh \widehat{\eta}_{a \to i} \\ w_a &= \prod_{i \in \partial a \cap g} \tanh \eta_{i \to a}. \end{cases} \tag{6.17}$$

Using inequalities (6.10), (6.15), it is straightforward to bound the check activities

$$\begin{aligned} |K_a(g)| &\leq \frac{|u_a| + |v_a|}{1 - |u_a||w_a|} \\ &\leq \frac{\theta^{r-d_a(g)} + \theta^{(r-1)d_a(g)}}{1 - \theta^r}. \end{aligned} \tag{6.18}$$

Thus for a fixed numerical constant α that we can take close to one

$$|K_a(g)| \leq \begin{cases} 1 + \alpha \theta^r & \text{if} \quad d_a(g) = r \\ \alpha \theta^{r-d_a(g)} & \text{if} \quad d_a(g) \neq r. \end{cases} \tag{6.19}$$

The activities associated with variable nodes (4.14) read

$$K_i(g) = \frac{e^{(u_i - v_i)} + (-1)^{d_i(g)} e^{-(u_i - v_i)}}{e^{(u_i + w_i)} + e^{-(u_i + w_i)}} \prod_{a \in \partial i \cap g} \frac{\cosh \widehat{\eta}_{a \to i}}{\cosh \eta_{i \to a}}, \tag{6.20}$$

where

$$
\begin{cases}
u_i & = h_i + \sum_{a \in \partial i \setminus g} \widehat{\eta}_{a \to i} \\
v_i & = \sum_{a \in \partial i \cap g} \eta_{i \to a} \\
w_i & = \sum_{a \in \partial i \cap g} \widehat{\eta}_{a \to i}.
\end{cases}
\tag{6.21}
$$

Again by a direct application of inequalities (6.10) and (6.15), we find

$$
\begin{aligned}
|K_i(g)| \leq{} & \frac{e^{(d_i(g)+1)\tanh^{-1}\theta} + (-1)^{d_i(g)} e^{-(d_i(g)+1)\tanh^{-1}\theta}}{e^{(d_i(g)+1)\tanh^{-1}\theta} + e^{-(d_i(g)+1)\tanh^{-1}\theta}} \\
& \times \left(\frac{1+\theta^2}{1-\theta^2} \right)^{d_i(g)/2}.
\end{aligned}
\tag{6.22}
$$

For a fixed constant β close to one we have the following bound

$$
|K_i(g)| \leq \begin{cases}
1 + \frac{\beta}{2}(1 + 4d_i(g) + d_i(g)^2)\theta^2 & \text{if} \quad d_i(g) \text{ even} \\
\beta(1 + d_i(g))\theta & \text{if} \quad d_i(g) \text{ odd.}
\end{cases}
\tag{6.23}
$$

Using the formulas (4.13) we derive the estimate (6.14) of subgraphs activities for $\theta < \theta_0$ small enough. Estimate (6.14) is essentially optimal for small θ as can be checked by Taylor expanding $K(g)$ in powers of θ. $\qquad\square$

6.2.3 Splitting of the Loop Sum

Estimate (6.13), (6.14) is quite cumbersome, so let us begin with a few remarks to understand its main qualitative features. The activity $K(\gamma)$ is not necessarily very small for graphs containing too many check nodes of maximal induced degree and too many variable nodes of even induced degree. More precisely for these "bad graphs" the rate of decay as $|\gamma|$ grows is too slow even for θ small, and it is not clear that it counterbalances the exponentially large entropic terms. However the rate of decay as $|\partial\gamma \cap C|$ grows is large for θ small. Here the boundary $\partial\gamma \cap C$ is by definition the set of check nodes in γ of non-maximal induced degree. An example is shown in Figure 6.1. For this reason we will analyze the contributions from small and large polymers separately.

The size under which a polymer is small is chosen based on expansion arguments. For the convenience of the reader we briefly review the necessary tools [59]. We will say that Γ is a (λ, κ) expander if for every subset $\mathcal{V} \subset V$ such that $|\mathcal{V}| < \lambda n$ we have

$$
|\partial\mathcal{V}| \geq \kappa l \, |\mathcal{V}|,
\tag{6.24}
$$

where $|\partial\mathcal{V}|$ is the number of check nodes that are connected to $\mathcal{V}$, and λ, κ are two positive numerical constants. Take a random $\Gamma \subset \mathcal{B}(l, r, n)$. Fix $0 < \kappa < 1 - \frac{1}{l}$ and

$$\overline{K} = (\alpha\theta^2)^2 \qquad\qquad \overline{K} = (1 + \alpha\theta^4)^3(\alpha\theta)$$
$$(1 + \tfrac{13}{2}\beta\theta^2)^2 \qquad\qquad (4\beta\theta)(1 + \tfrac{13}{2}\beta\theta^2)^6$$
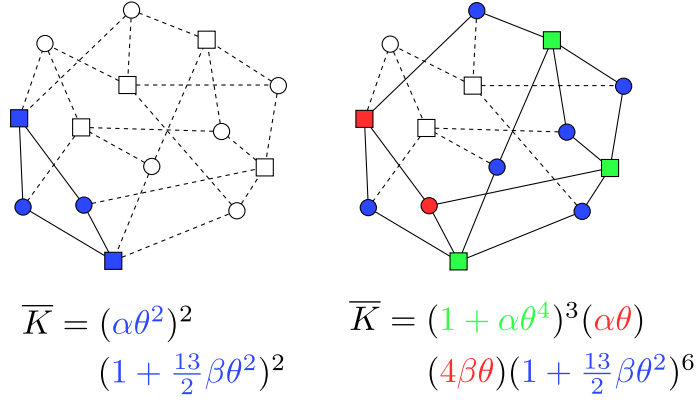
Figure 6.1: Example for a $\Gamma \in \mathrm{LDPC}\,(3,4,8)$ of polymers and their associated activities. On the left a small polymer and on the right a large polymer.

$0 < \lambda < \lambda_0$ where λ_0 is the (only) positive solution of the equation[2]

$$\frac{l-1}{l}h_2(\lambda_0) - \frac{1}{r}h_2(\lambda_0\kappa r) - \lambda_0\kappa r h_2(\frac{1}{\kappa r}) = 0\,. \tag{6.25}$$

Then we have

$$\mathbb{P}[\Gamma \text{ is a } (\lambda,\kappa) \text{ expander}] = 1 - O(\frac{1}{n^{l(1-\kappa)-1}})\,. \tag{6.26}$$

Later on we need to take $\kappa \in]1 - \frac{2(r-1)}{lr}, 1 - \frac{1}{l}[$, which is always possible for $r > 2$. In the rest of the chapter κ is always a constant in this interval, and $0 < \lambda < \lambda_0$. For concreteness, one can think of the example $(l,r) = (3,6)$, $\kappa = 0.5$ and $\lambda_0 = 7.7 \times 10^{-4}$.

We say that a polymer is small if $|\gamma| < \lambda n$. We define the partition function (with activities computed at the high-noise fixed-point) of a gas of small polymers

$$Z_{\text{small}} = \sum_{M \geq 0}\frac{1}{M!} \sum_{\gamma_1,\ldots,\gamma_M \text{ s.t } |\gamma_k|<\lambda n} \prod_{k=1}^{M} K\,(\gamma_k) \prod_{k<k'} \mathbb{I}\,(\gamma_k \cap \gamma_{k'} = \emptyset)\,. \tag{6.27}$$

The loop sum is equal to the partition function of the gas of small polymers plus a contribution containing at least one polymer of large size $|\gamma| > \lambda n$. We call the latter contribution R_{large}. More precisely

$$1 + \sum_{g \subset \Gamma} K(g) = Z_{\text{small}} + R_{\text{large}}, \tag{6.28}$$

where

$$R_{\text{large}} = \sum_{g \subset \Gamma \text{ s.t } \exists\gamma\subset g \text{ with } |\gamma|\geq\lambda n} K(g), \tag{6.29}$$

---

[2]Here $h_2(x) = -x\ln x - (1-x)\ln(1-x)$ is the binary entropy function.

6.3 Contribution of Small Polymers

For $\Gamma \subset \mathrm{LDPC}(l, r, n)$ that are expanders, if γ is "small" then $|\partial\gamma \cap C|$ is of the order of $|\gamma|$ and the activity is exponentially small in the size of the polymer. This is the meaning of the following lemma.

Lemma 6.2 (Activity bound for expander graphs). *Assume that Γ is a (λ, κ) expander with $\kappa \in]1 - \frac{2(r-1)}{lr}, 1 - \frac{1}{l}[$. For $|\gamma| < \lambda n$ we have for θ small enough*

$$|K(\gamma)| \leq \theta^{\frac{c}{2}|\gamma|}, \tag{6.30}$$

with

$$c = r - \frac{2 + r}{3 - l(1 - \kappa)}. \tag{6.31}$$

Remark 6.2. *In the process of this derivation one has to require $3 - l(1 - \kappa) > 0$ and $c > 0$. This imposes the condition on the expansion constant $\kappa > 1 - \frac{2(r-1)}{lr}$. Note that an expansion constant cannot be greater than $1 - 1/l$, so it is fortunate that we have $1 - \frac{1}{l} > 1 - \frac{2(r-1)}{lr}$ (for any $r > 2$).*

Proof. Recall that $d_i(\gamma)$ (resp. $d_a(\gamma)$) is the induced degree of node i (resp. a) in γ. The type of γ is given by two vectors $\underline{n} = (n_s(\gamma))_{s=2}^l$ and $\underline{m} = (m_t(\gamma))_{t=2}^r$ defined as $n_s(\gamma) := |\{i \in \gamma \cap V | d_i(\gamma) = s\}|$ and $m_t(\gamma) := |\{a \in \gamma \cap C | d_a(\gamma) = t\}|$. In words, $n_s(\gamma)$ and $m_t(\gamma)$ count the number of variable and check nodes with induced degrees s and t in γ. Note that we have the constraints

$$\begin{cases} |\gamma| = \sum_{s=2}^l n_s(\gamma) + \sum_{t=2}^r m_t(\gamma) \\ \sum_{s=2}^l s n_s(\gamma) = \sum_{t=2}^r t m_t(\gamma) \end{cases} \tag{6.32}$$

We apply the expander property to the set $\mathcal{V} = \{i \in \gamma \cap V\}$. This reads

$$|\partial\mathcal{V}| \geq \kappa l \sum_{s=2}^l n_s(\gamma). \tag{6.33}$$

On the other hand $|\partial\mathcal{V}| \leq \sum_{t=2}^r m_t(\gamma) + \sum_{s=2}^l (l - s) n_s(\gamma)$. With (6.32) and (6.33) this yields the constraint

$$\sum_{t=2}^r (r - t) m_t(\gamma) \geq -|\gamma| l(1 - \kappa) + (l(1 - \kappa) + r - 1) \sum_{t=2}^r m_t(\gamma). \tag{6.34}$$

Relaxing the second constraint in (6.32) gives

$$\sum_{t=2}^r t m_t(\gamma) \geq 2 \sum_{s=2}^l n_s(\gamma). \tag{6.35}$$

Combined with the first constraint of (6.32) this yields

$$(r+2)\sum_{t=2}^{r}m_t(\gamma) \geq 2|\gamma| + \sum_{t=2}^{r}(r-t)m_t(\gamma). \tag{6.36}$$

We have by use of inequalities (6.34) and (6.36)

$$\sum_{t=2}^{r-1}(r-t)m_t(\gamma) \geq \left(r - \frac{2+r}{3-l(1-\kappa)}\right)|\gamma|. \tag{6.37}$$

Finally, by bounding the product over $t = 2, \cdots, r-1$ in the activity bound (6.14), we obtain (6.30). $\qquad\square$

The free energy of the gas of small polymers $n^{-1}\ln Z_{\text{small}}$ has a polymer expansion (4.24) with the second sum replaced by a sum over $\gamma_1, \ldots, \gamma_M$ s.t $|\gamma_i| < \lambda n$.

Lemma 6.3 (Convergence of the small polymer expansion)**.** *Suppose $r > 2$ and assume that Γ is a (λ, κ) expander with $\kappa \in]1 - \frac{2(r-1)}{lr}, 1 - \frac{1}{l}[$. One can find $\theta_0 > 0$ such that for $|\theta| < \theta_0$*

$$\left|\frac{1}{n}\ln Z_{\text{small}}\right| \leq \frac{4}{n}\sum_{x \in V \cup C}\theta^{\frac{c}{2}R_x}e^{AR_x}, \tag{6.38}$$

where R_x is the smallest size of a polymer containing x.

Proof. When Γ is an expander we can use the bound (6.30) on the activities of the small polymers and Corollary 4.1. $\qquad\square$

Lemma 6.3 has the following consequence:

Corollary 6.1 (Control of the small polymer expansion for expander graphs)**.** *Suppose $r > 2$. Let E be the event that Γ is (λ, κ) expander. For $|\theta| < \theta_0$,*

$$\mathbb{E}_{\Gamma}\left[\frac{1}{n}|\ln Z_{\text{small}}|\Big| E\right] = O(n^{-(1-\chi)}) \tag{6.39}$$

for any $0 < \chi < 1$.

Remark 6.3. *We stress that Corollary 6.1 and Lemma 6.3 hold for any (l, r) with $r > 2$. The restriction to odd l will come only when we estimate the contribution of large polymers.*

Proof. Taking the conditional expectation over expander graphs (6.38) implies

$$\frac{1}{n}\mathbb{E}_{\Gamma}\left[|\ln Z_{\text{small}}|\Big| E\right] \leq 4\left(1 + \frac{l}{r}\right)\mathbb{E}_{\Gamma}\left[\theta^{\frac{c}{2}R_o}e^{(1+\max(l_{\max}, r_{\max}))R_o}\Big| E\right], \tag{6.40}$$

were $o \in C \cup V$ is a fixed vertex. We can compute this expectation by conditioning on the first event that Γ is tree-like in a neighborhood of size $O(\ln n)$ around this vertex, and on the second complementary event. The second event has small probability $O(n^{-(1-\chi)})$ for any $0 < \chi < 1$. Besides, from (6.38) it is easy to show that $n^{-1}|\ln Z_{\text{small}}|$ is bounded uniformly in n (put $2 \leq R_x$). Thus the second event contributes only $O(n^{-(1-\chi)})$ to the expectation. For the first event we have that the smallest polymer is a cycle with $|\gamma| = O(\ln n)$. This implies that this event contributes $O((\theta^{\frac{c}{2}} e^A)^{\ln n}))$ to the expectation. For small θ it is $O(n^{-(1-\chi)})$ that dominates. $\qquad\square$

6.4   Probability Estimates on Graphs

The next lemma shows that the contribution from large polymers is exponentially small, with high probability with respect to the graph ensemble.

Lemma 6.4 (Control of the large polymer loop sum). *Fix $\delta > 0$. Assume $l \geq 3$ odd and $l < r$. There exists a constant $C > 0$ depending only on l and r such that for θ small enough*

$$\mathbb{P}_\Gamma \left[|R_{\text{large}}| \geq \delta \right] \leq \frac{1}{\delta} e^{-Cn}. \tag{6.41}$$

Proof. Let $\Omega_\Gamma(\underline{n}, \underline{m})$ be the set of all $g \subset \Gamma$ with prescribed type $(\underline{n}(g), \underline{m}(g))$. By (6.14) and the Markov bound

$$\mathbb{P}\left[|R_{\text{large}}| \geq \delta \right] \leq \frac{1}{\delta} \sum_{\underline{n}, \underline{m} \in \Delta} \overline{K}(\underline{n}, \underline{m}) \, \mathbb{E}_\Gamma \left[|\Omega_\Gamma(\underline{n}, \underline{m})| \right], \tag{6.42}$$

where

$$\Delta \equiv \left\{ (\underline{n}, \underline{m}) \mid \lambda n \leq \sum_{s=2}^{l} n_s + \sum_{t=2}^{r} m_t, \ \sum_{s=2}^{l} s n_s = \sum_{t=2}^{r} t m_t, \ \sum_{s=2}^{l} n_s < n, \sum_{t=2}^{r} m_t < nl/r \right\}. \tag{6.43}$$

The expectation of the number of $g \subset \Gamma$ with prescribed type can be estimated by combinatorial bounds provided by McKay [102]. It turns out that these subgraphs proliferate exponentially in n only for a subdomain of Δ where $\overline{K}(\underline{n}, \underline{m})$ is exponentially much smaller in n. In the subdomain where $\overline{K}(\underline{n}, \underline{m})$ is not small (but it is always bounded) the number of subgraphs is sub-exponential when l is odd and $l < r$. As a consequence for l odd and $l < r$, we are able to prove that the sum on the right-hand side of (6.42) is smaller than e^{-Cn}.

Let us now give the details of this calculation. Let $\omega = 4r^2 - 2r + 2$, a number independent of n. The combinatorial bound is only valid for subgraphs g with number of edges at

most equal to $nl - \omega$. Thus we have to separate the domain of summation (6.43) into

$$\Delta_\omega = \Delta \cap \left\{ (\underline{n}, \underline{m}) \mid \sum_{s=2}^{l} s n_s \leq nl - \omega \right\} \text{ and } \Delta_\omega^c = \Delta \setminus \Delta_\omega, \tag{6.44}$$

and handle each part separately.

For $(\underline{n}, \underline{m}) \in \Delta_\omega^c$ a trivial bound of the expected size of $\Omega_\Gamma(\underline{n}, \underline{m})$ is given by

$$\mathbb{E}_\Gamma \left[|\Omega_\Gamma(\underline{n}, \underline{m})| \right] \leq \binom{nl}{\omega}$$
$$= O(n^\omega). \tag{6.45}$$

This is nothing but simply counting the possible subgraphs obtained by removing ω edges from Γ. For the same reason the activity (6.14) is upper-bounded by

$$\left| \overline{K}(\underline{n}, \underline{m}) \right| \leq (1 + \alpha \theta^r)^{n\frac{l}{r}} \left(1 + \frac{\beta}{2}(1 + 4l + l^2)\theta^2 \right)^\omega (\beta(1 + l)\theta)^{n - \omega}$$
$$= O\left(\left(\beta(1 + l)(1 + \alpha)^{\frac{l}{r}} \theta \right)^{n - \omega} \right). \tag{6.46}$$

Indeed the activity of the total graph is upper-bounded by

$$\overline{K}(\Gamma) = (1 + \alpha \theta^r)^{n\frac{l}{r}} (\beta(1 + l)\theta)^n. \tag{6.47}$$

The worst case scenario for the activity of subgraphs obtained by removing ω edges from Γ is then bounded by (6.46). Therefore for every $\theta < \theta_1 = \left(\beta(1 + l)(1 + \alpha)^{\frac{l}{r}} \right)^{-1}$ and n large enough, there exists a constant $C_1 > 0$ depending on l and r such that

$$\sum_{\underline{n}, \underline{m} \in \Delta_\omega^c} \overline{K}(\underline{n}, \underline{m}) \mathbb{E}_\Gamma \left[|\Omega_\Gamma(\underline{n}, \underline{m})| \right] \leq e^{C_1(n - \omega) \ln \left(\frac{\theta}{\theta_1} \right)}. \tag{6.48}$$

For $(\underline{n}, \underline{m}) \in \Delta_\omega$, the probability that a graph g with prescribed type $(\underline{n}(g), \underline{m}(g))$ belongs to $\Omega_\Gamma(\underline{n}, \underline{m})$ is upper-bounded by McKay's estimate

$$\mathbb{P}_\Gamma \left[g \in \Omega_\Gamma(\underline{n}, \underline{m}) \right] \leq \frac{\prod_{s=2}^{l} \left(\frac{l!}{(l-s)!} \right)^{n_s} \prod_{t=2}^{r} \left(\frac{r!}{(r-t)!} \right)^{m_t}}{\frac{(nl - \omega)!}{\left(nl - \sum_{s=2}^{l} s n_s - \omega \right)!}}. \tag{6.49}$$

By counting the number of graphs g with prescribed degrees $(\underline{n}(g), \underline{m}(g))$, we deduce

$$\mathbb{E}_\Gamma \left[|\Omega_\Gamma(\underline{n}, \underline{m})| \right] \leq \binom{n}{n_2, ..., n_l} \binom{n\frac{l}{r}}{m_2, ..., m_r} \frac{\left(\sum_{s=2}^l sn_s \right)!}{\prod_{s=2}^l s!^{n_s} \prod_{t=2}^r t!^{m_t}} \mathbb{P}_\Gamma \left[g \in \Omega_\Gamma(\underline{n}, \underline{m}) \right].$$

(6.50)

Setting $x_s = \frac{n_s}{n}$, $y_t = \frac{r}{l} \frac{m_t}{n}$, we perform an asymptotic analysis for n large of the bound (6.50). We first estimate factorials using Stirling approximation valid for $k > 0$

$$e^{\frac{1}{12k+1}} \leq \frac{k!}{\sqrt{2\pi k}e^{-k}k^k} \leq e^{\frac{1}{12k}}.$$

(6.51)

In order to simplify the terms in ω we also use the following inequality valid for $n > l\omega$ and $0 \leq z \leq 1 - \frac{\omega}{nl}$

$$(1-z)\ln(1-z) - \frac{\omega}{nl}\ln\left(\frac{\omega}{nl}\right) \geq \left(1 - z - \frac{\omega}{nl}\right)\ln\left(1 - z - \frac{\omega}{nl}\right).$$

(6.52)

This could be easily proven by considering a joint probability distribution $p(A = 0, B = 0) = 0$, $p(A = 0, B = 1) = z$, $p(A = 1, B = 0) = \frac{\omega}{nl}$, $p(A = 1, B = 1) = 1 - z - \frac{\omega}{nl}$ and applying the inequality

$$H(A) \leq H(A, B),$$

(6.53)

where H is the Shannon entropy in nat.

Observe that

$$-\left(1 - \frac{\omega}{nl}\right)\ln\left(1 - \frac{\omega}{nl}\right) \leq \frac{\omega}{nl}.$$

(6.54)

Using the relations (6.50), (6.51), (6.52) along with (6.54) gives the following bound on the number of subgraphs of Γ

$$\mathbb{E}_\Gamma \left[|\Omega_\Gamma(\underline{n}, \underline{m})| \right] \leq C_{l,r} n^{\frac{\omega}{l}+2} \exp\left(nl f\left(\underline{x}, \underline{y}\right)\right),$$

(6.55)

where $C_{l,r}$ is a constant that depends only on l and r and

$$\begin{aligned}
f\left(\underline{x}, \underline{y}\right) = {} & \left(1 - \sum_{s=2}^l \frac{s}{l}x_s\right)\ln\left(1 - \sum_{s=2}^l \frac{s}{l}x_s\right) + \left(\sum_{s=2}^l \frac{s}{l}x_s\right)\ln\left(\sum_{s=2}^l \frac{s}{l}x_s\right) \\
& + \frac{1}{l}\left(\sum_{s=2}^l x_s \ln\binom{l}{s}\right) + \frac{1}{r}\left(\sum_{t=2}^r y_t \ln\binom{r}{t}\right) \\
& - \frac{1}{r}\left(\left(1 - \sum_{t=2}^r y_t\right)\ln\left(1 - \sum_{t=2}^r y_t\right) + \sum_{t=2}^r y_t \ln y_t\right) \\
& - \frac{1}{l}\left(\left(1 - \sum_{s=2}^l x_s\right)\ln\left(1 - \sum_{s=2}^l x_s\right) + \sum_{s=2}^l x_s \ln x_s\right).
\end{aligned}$$

(6.56)

The bound on the activity (6.14) can also be put in a form where the growth rate in n is explicit

$$\overline{K}\left(\underline{n},\underline{m}\right) = \exp\left(nlk_\theta\left(\underline{x},\underline{y}\right)\right), \tag{6.57}$$

where

$$
\begin{aligned}
k_\theta\left(\underline{x},\underline{y}\right) =& \frac{y_r}{r}\ln\left(1+\alpha\theta^r\right) + \sum_{t=2}^{r-1}\frac{y_t}{r}\ln\left(\alpha\theta^{r-t}\right)\\
&+ \sum_{\substack{s=2,\\\text{even}}}^{l-1}\frac{x_s}{l}\ln\left(1+\frac{\beta}{2}\left(1+4s+s^2\right)\theta^2\right)\\
&+ \sum_{\substack{s=3,\\\text{odd}}}^{l}\frac{x_s}{l}\ln\left(\beta\left(1+s\right)\theta\right).
\end{aligned}
\tag{6.58}
$$

Define the ensemble

$$
\begin{aligned}
\Delta' \equiv \Bigg\{&\left(\underline{x},\underline{y}\right)\in\mathbb{R}_+^{l-1}\times\mathbb{R}_+^{r-1}\mid \lambda\leq\frac{1}{l}\sum_{s=2}^{l}x_s+\frac{1}{r}\sum_{t=2}^{r}y_t,\\
&\sum_{s=2}^{l}\frac{s}{l}x_s=\sum_{t=2}^{r}\frac{t}{r}y_t,\sum_{s=2}^{l}x_s<1,\sum_{t=2}^{r}y_t<1\Bigg\}.
\end{aligned}
\tag{6.59}
$$

It is easy to verify that if $\left(\underline{n},\underline{m}\right)\in\Delta_\omega$ then $\left(\underline{x},\underline{y}\right)\in\Delta'$. Combining (6.48), (6.55) and (6.57) gives finally

$$\sum_{\underline{n},\underline{m}\in\Delta_\omega}\overline{K}\left(\underline{n},\underline{m}\right)\mathbb{E}_\Gamma\left[|\Omega_\Gamma\left(\underline{n},\underline{m}\right)|\right]\leq C'_{l,r}n^{\frac{\omega}{l}+l+r}\exp\left(nl\Lambda\right), \tag{6.60}$$

where

$$\Lambda\left(\theta\right) = \max_{\left(\underline{x},\underline{y}\right)\in\Delta'}\left\{f\left(\underline{x},\underline{y}\right)+k_\theta\left(\underline{x},\underline{y}\right)\right\}. \tag{6.61}$$

In (6.48) we estimate the sum over $\left(\underline{n},\underline{m}\right)\in\Delta_\omega$ by the crude bound $|\Delta_\omega|\leq n^{l-1}\left(\frac{nl}{r}\right)^{r-1}$.

It remains now to prove that $\Lambda\left(\theta\right)$ is strictly negative for θ small enough. In the subspace $\Delta_0\subset\Delta'$ defined by having all coordinates x_s for s odd and y_t for $t<r$ equal to zero, the function $k_\theta\left(\underline{x},\underline{y}\right)$ can be made arbitrarily close to zero as θ is small. Notice also that in the complementary subspace $\Delta'\setminus\Delta_0$, the function $k_\theta\left(\underline{x},\underline{y}\right)$ can be made arbitrarily negative for small θ due to the presence of the terms $\ln\theta$. It is therefore sufficient to show that the restriction of $f\left(\underline{x},\underline{y}\right)$ to Δ_0 is strictly negative. Call $z_s=x_{2s}$ and define the set

$$\Delta_0' \equiv \left\{\underline{z}\in\mathbb{R}_+^{\frac{l-1}{2}}\mid l\lambda\leq\sum_{s=1}^{\frac{l-1}{2}}z_s<1\right\}. \tag{6.62}$$

If $\underline{z} \in \Delta'_0$ then $\left(\underline{x}, \underline{y}\right) \in \Delta_0$, as we can express the variable $y_r = \sum_{s=1}^{\frac{l-1}{2}} \frac{2s}{l} z_s$ with the second constraint in (6.59). The restriction to Δ_0 of $f\left(\underline{x}, \underline{y}\right)$ can be recast into the form

$$l f\left(\underline{x}, \underline{y}\right) = f_0\left(\underline{z}\right) - \left(1 - \frac{l}{r}\right) h_2 \left(\sum_{s=1}^{\frac{l-1}{2}} \frac{2s}{l} z_s \right), \tag{6.63}$$

where

$$f_0\left(\underline{z}\right) = -\left(l - 1\right) h_2 \left(\sum_{s=1}^{\frac{l-1}{2}} \frac{2s}{l} z_s \right) + \sum_{s=1}^{\frac{l-1}{2}} z_s \ln \binom{l}{2s}$$

$$- \left(1 - \sum_{s=1}^{\frac{l-1}{2}} z_s \right) \ln \left(1 - \sum_{s=1}^{\frac{l-1}{2}} z_s \right) - \sum_{s=1}^{\frac{l-1}{2}} z_s \ln z_s. \tag{6.64}$$

The function f_0 takes its maximum in Δ'_0 at $\underline{z}^* = \frac{1}{2^{l-1}} \binom{l}{2s}$ and $f_0\left(\underline{z}^*\right) = 0$. Thus, since $2\lambda < \sum_{s=1}^{\frac{l-1}{2}} \frac{2s}{l} z_s < 1 - \frac{1}{l}$, for $\left(\underline{x}, \underline{y}\right) \in \Delta_0$ we have

$$l f\left(\underline{x}, \underline{y}\right) < -\left(1 - \frac{l}{r}\right) \min \left\{ h_2\left(2\lambda\right), h_2\left(\frac{1}{l}\right) \right\} < 0. \tag{6.65}$$

Therefore for θ small enough $\Lambda\left(\theta\right) < 0$ and there exist for large n a constant $C_2 > 0$ depending on l and r such that

$$\sum_{\underline{n}, \underline{m} \in \Delta_\omega} \overline{K}\left(\underline{n}, \underline{m}\right) \mathbb{E}_\Gamma\left[\left| \Omega_\Gamma\left(\underline{n}, \underline{m}\right) \right| \right] \leq e^{-nC_2}. \tag{6.66}$$

Combining Markov's inequality (6.42) and inequalities (6.48), (6.66) ends the proof.

Notice that the condition $\frac{l}{r} < 1$ appears naturally in (6.65). It is thus necessary that the graph Γ describes a code (i.e., with positive rate). $\qquad \square$

6.5 Exactness of the Bethe Free Energy for LDPC Codes

The results of sections 6.3 and 6.4 allow first to prove

Theorem 6.1 (Leading correction terms are given by the polymer expansion)**.** *Suppose l is odd and $3 \leq l < r$. Take Γ at random in $\mathrm{LDPC}(l, r, n)$. There exist a small θ_0 independent of n such that for $\theta < \theta_0$, and any high-noise-solution $\left(\underline{\eta}, \underline{\widehat{\eta}}\right)$ of the BP equations, with probability $1 - O(n^{-(l(1-\kappa)-1)})$ we have*

$$\left| \frac{1}{n} \ln Z_{\mathrm{LDPC}} - (f_{\mathrm{LDPC}}^{\mathrm{Bethe}}\left(\underline{\eta}, \underline{\widehat{\eta}}\right) + \frac{1}{n} \ln Z_{\mathrm{small}}) \right| = O(e^{-n^{l(1-\kappa)-1}}). \tag{6.67}$$

Remark 6.4. *We recall that $0 < l(1 - \kappa) - 1 < (r - 2)/r$. This proposition shows that large polymers contribute only with exponentially small corrections to the Bethe free energy. Inverse power in n corrections can be computed systematically from the polymer expansion of $n^{-1} \ln Z_{\text{small}}$.*

Proof. Note that

$$\frac{1}{n} \ln \left(\sum_{g \subset \Gamma} K(g) \right) = \frac{1}{n} \ln Z_{\text{small}} + \frac{1}{n} \ln \left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}} \right), \tag{6.68}$$

which means that the term on the left hand side of (6.67) is equal to

$$\frac{1}{n} \left| \ln \left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}} \right) \right|. \tag{6.69}$$

On one hand, from corollary 6.1 and the Markov bound, we have for any $\epsilon > 0$,

$$\mathbb{P} \left[e^{-n\epsilon} \leq Z_{\text{small}}^{-1} \leq e^{n\epsilon} \middle| E \right] = 1 - \frac{1}{\epsilon} O(n^{-(1-\chi)}). \tag{6.70}$$

On the other hand, from lemma 6.4

$$\mathbb{P} \left[|R_{\text{large}}| \geq \delta \middle| E \right] \mathbb{P}[E] \leq \mathbb{P} \left[|R_{\text{large}}| \geq \delta \right]$$
$$\leq \frac{1}{\delta} e^{-Cn} \tag{6.71}$$

and since $\mathbb{P}[E] = 1 - O(n^{-(l(1-\kappa)-1)})$,

$$P \left[|R_{\text{large}}| \geq \delta \middle| E \right] \leq \frac{2}{\delta} e^{-Cn}. \tag{6.72}$$

Using (6.70) and (6.72), and choosing $\frac{\delta}{2} = e^{-2n\epsilon}$ it is not difficult to show that (at this point one must take $0 < 2\epsilon < C$)

$$\mathbb{P} \left[\left| \frac{R_{\text{large}}}{Z_{\text{small}}} \right| \geq e^{-n\epsilon} \middle| E \right] \leq \frac{1}{\epsilon} O(n^{-(1-\chi)}) + e^{-n(C-2\epsilon)}. \tag{6.73}$$

Thus

$$\mathbb{P} \left[\left| \frac{R_{\text{large}}}{Z_{\text{small}}} \right| \leq e^{-n\epsilon} \middle| E \right] \geq 1 - \frac{1}{\epsilon} O(n^{-(1-\chi)}). \tag{6.74}$$

This implies for n large

$$\mathbb{P}\left[\left|\ln\left(1+\frac{R_{\text{large}}}{Z_{\text{small}}}\right)\right|\leq 2e^{-n\epsilon}\Big|E\right]\geq 1-\frac{1}{\epsilon}O(n^{-(1-\chi)}). \tag{6.75}$$

Furthermore

$$\mathbb{P}\left[\left|\ln\left(1+\frac{R_{\text{large}}}{Z_{\text{small}}}\right)\right|\leq 2e^{-n\epsilon}\right]\geq\left(1-\frac{1}{\epsilon}O(n^{-(1-\chi)})\right)\left(1-O(n^{-(l(1-\kappa)-1)})\right)$$
$$\geq 1-O(n^{-(l(1-\kappa)-1)}). \tag{6.76}$$

The last line is obtained by choosing

$$\epsilon=\frac{n^{l(1-\kappa)-1}}{n^{1-\chi}}\leq n^{\frac{r-2}{r}-1+\chi}<\frac{C}{2}, \tag{6.77}$$

which is possible since $\kappa\in]1-\frac{2(r-1)}{lr},1-\frac{1}{l}[$ and we can take $\chi>0$ as small as we wish.

Finally from

$$n\epsilon=n^{l(1-\kappa)-1+\chi}\geq n^{l(1-\kappa)-1} \tag{6.78}$$

and (6.76) we deduce the statement of the proposition. $\qquad\square$

Theorem 6.2 (Exactness of the Bethe free energy: LDPC high noise)**.** *Suppose l is odd and $3\leq l\leq r$. There exists $\theta_0>0$ (small) independent of n, such that for $\theta\leq\theta_0$ and any high-noise-solution*

$$\mathbb{E}_\Gamma\left[\left|\frac{1}{n}\ln Z_{\text{LDPC}}-f_{\text{LDPC}}^{\text{Bethe}}\left(\underline{\eta},\underline{\widehat{\eta}}\right)\right|\right]=O(n^{-(l(1-\kappa)-1)}). \tag{6.79}$$

The $O(\cdot)$ is uniform in the channel output realizations $\underline{h}$.

Remark 6.5. *We recall that $\kappa\in]1-\frac{2(r-1)}{lr},1-\frac{1}{l}[$ which implies that the expansion constant κ is such that, for $r>2$, $0<l(1-\kappa)-1<(r-2)/r$.*

Proof. Consider the difference

$$\left|\frac{1}{n}\ln Z_{\text{LDPC}}-f_{\text{LDPC}}^{\text{Bethe}}(\underline{\eta},\underline{\widehat{\eta}})\right| \tag{6.80}$$

We first remark that this quantity is bounded uniformly in n because each term $n^{-1}|\ln Z_{\text{LDPC}}|$ and $|f_{\text{LDPC}}^{\text{Bethe}}|$ is bounded, as can be checked directly from their definition.

Now consider the event S - or the set of graphs - such that

$$\left|\ln\left(1+\frac{R_{\text{large}}}{Z_{\text{small}}}\right)\right|\leq e^{-n^{l(1-\kappa)-1}}. \tag{6.81}$$

Proposition (6.1) says that

$$\mathbb{P}[S^c] = O(n^{-(l(1-\kappa)-1)}). \tag{6.82}$$

Thus we have

$$\mathbb{E}\left[\left|\frac{1}{n}\ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}}\right|\Big|S^c\right]\mathbb{P}[S^c] = O(n^{-(l(1-\kappa)-1)}). \tag{6.83}$$

We will now estimate

$$\mathbb{E}\left[\left|\frac{1}{n}\ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}}\right|\Big|S\right]\mathbb{P}[S]. \tag{6.84}$$

Since $\mathbb{P}[S] = 1 - O(n^{-(l(1-\kappa)-1)})$ we have to show that the expectation conditioned over S is small.

$$\mathbb{E}\left[\left|\frac{1}{n}\ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}}\right|\Big|S\right] = \mathbb{E}\left[\left|\frac{1}{n}\ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}}\right|\Big|S\cap E\right]\mathbb{P}[E]$$

$$+ \mathbb{E}\left[\left|\frac{1}{n}\ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}}\right|\Big|S\cap E^c\right]\mathbb{P}[E^c]. \tag{6.85}$$

Since, as remarked before, (6.80) is bounded and $\mathbb{P}[E^c] = O(n^{-(l(1-\kappa)-1)})$ the second term on the right-hand side is $O(n^{-(l(1-\kappa)-1)})$. It remains to show that the conditional expectation in the first term of the right-hand side is small. This is bounded above by two contributions. The first one is

$$\mathbb{E}\left[\frac{1}{n}\left|\ln\left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}}\right)\right|\Big|S\cap E\right] \le e^{-n^{l(1-\kappa)-1}}, \tag{6.86}$$

and the second (recall corollary 6.1)

$$\mathbb{E}\left[\frac{1}{n}|\ln Z_{\text{small}}|\Big|S\cap E\right] = O(n^{-(1-\chi)}). \tag{6.87}$$

Putting all contributions (6.83), (6.85), (6.86), (6.87) together we obtain the desired result

$$\mathbb{E}\left[\left|\frac{1}{n}\ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}}\right|\right] = O(n^{-(1-\chi)}) + O(n^{-(l(1-\kappa)-1)})$$

$$= O(n^{-(l(1-\kappa)-1)}). \tag{6.88}$$

In the last step we have taken $0 < \chi < l(1-\kappa) - 1$. $\qquad\square$

6.6 Discussion

6.6.1 LDPC: The Case l Even

When l is even, the point $\theta = 0$ has a singular behavior. As the channel realization is trivial $\underline{h} = \underline{0}$, the low-noise fixed-point is simply the all zero messages $(\eta_{i \to a}, \widehat{\eta}_{a \to i}) = (0,0)$. The activities can be exactly computed for this BP fixed-point

$$K_a(\gamma) = \begin{cases} 1 & \text{if } |\partial a \cap \gamma| = r \\ 0 & \text{otherwise} \end{cases}, \ K_i(\gamma) = \frac{1 + (-1)^{|\partial i \cap \gamma|}}{2}. \tag{6.89}$$

When the graph Γ is an expander, every small polymer $|\gamma| < n\lambda$ contains at least one check node with an induced degree less than r (see Lemma 6.2). Thus $K(\gamma) = 0$ and

$$Z_{\text{small}} = 1. \tag{6.90}$$

The contribution of the small polymer vanishes, as predicted already by the polymer expansion (see Lemma 6.3).

However for the total graph, and unlike the case l odd, we have

$$K(\Gamma) = 1. \tag{6.91}$$

More generally, polymers with a size on the order of the total graph have an activity close to one. This implies that the contribution from large polymers is non-vanishing but growing linearly

$$1 < \mathbb{E}_\Gamma(|R_{\text{large}}|) < C_{l,r} n^{4r^2}. \tag{6.92}$$

This can be shown using the same counting arguments as in Section 6.4. As a consequence, we find that similarly to Theorem 6.1 the Bethe free energy is asymptotically exact with high probability. More precisely, with probability $1 - O(n^{-(l(1-\kappa)-1)})$

$$|\frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}}| = O\left(\frac{1}{n} \ln n\right). \tag{6.93}$$

The notable difference with Theorem 6.1 is that the decay rate of the difference is not exponential.

When l is even and $\theta > 0$, the bound on the activity of the total graph predict an exponential growth

$$\overline{K}(\Gamma) = (1 + \alpha \theta^r)^{\frac{l}{r}n} \left(1 + \beta \theta^2\right)^n. \tag{6.94}$$

The contribution of the large polymer can no longer be estimated as in Section 6.4. To tackle this problem, we believe that it is necessary to have a precise control of sign cancellations in the sum R_{large}. Such a control is out of reach of the method presented in this thesis.

6.6.2 The Case $l > r$

The constraint $r > l$ appears naturally in the course of proving Theorem 6.4. If $r < l$, large polymers with an activity exponentially increasing with their size are also exponentially numerous. Therefore we found by using counting arguments that the contribution of R_{large} is not negligible.

When $r < l$, the graphical model no longer describes a code. In fact at exactly $h = 0$, the partition function counts the number of solutions of a random linear system of equations that is overdetermined. This corresponds to an UNSAT phase of a linear-constraints satisfaction-problem. In this phase, it is expected that the Bethe free energy is not a good approximation of the free energy (instead one should use the RSB free energy [44]). Then it seems reasonable to think that the corrections to the Bethe free energy are non-vanishing, even with a precise control of the sign cancellation in the activities.

6.6.3 LDPC at Low Noise

The low-noise regime is characterized by half-log likelihood ratios with high magnitudes $h \approx \infty$. The low-noise fixed-point of the belief-propagation equations is the trivial solution

$$(\eta_{i \to a}, \widehat{\eta}_{a \to i}) = (+\infty, +\infty). \tag{6.95}$$

The activities can be exactly computed at the low-noise fixed-point

$$K(\gamma) = \prod_{a \in \gamma \cap C} \frac{1 + (-1)^{|\partial a \cap \gamma|}}{2} \prod_{i \in \gamma \cap V} (-1)^l \, e^{-2h_i} \mathbb{I}\left(|\partial i \cap \gamma| = l\right). \tag{6.96}$$

According to (6.96), polymers are subgraphs that have check nodes with even induced degree and variable nodes with induced degree equal to l. The particularity of the low-noise activities is that their intensity depends on the sign of the half-log likelihood ratios h_i, and a fortiori on the distribution of $\underline{h}$. Using Hoeffding's inequality, we see that a polymer has, with large probability, a small activity

$$\mathbb{P}_{\underline{h}}\left(|K(\gamma)| \le e^{-2h(\tanh h - 2\epsilon)|\gamma \cap V|}\right) \ge 1 - e^{-2\epsilon^2|\gamma \cap V|}. \tag{6.97}$$

However, the expected activity is dominated by rare events

$$\mathbb{E}_{\underline{h}}\left[|K(\gamma)|\right] = 1. \tag{6.98}$$

This prevents the use of the same counting arguments as in Section 6.4.

Approaching the Rate-Distortion Bound by Spatial Coupling, a Perspective from the Cavity Method

Part III

7 Approaching the Rate-Distortion Bound by Spatial Coupling

In this chapter[1], we show that, under a low-complexity BPGD-h and BPGD-r algorithm, spatially-coupled LDGM codes approach the rate distortion curve for any compression rate.

In Section 7.1, we recall briefly the problem of lossy source coding. We describe the LDGM and the spatially-coupled LDGM ensemble in terms of their graphical representations. We also recall the spin glass formulation of the problem and its associated Gibbs measure.

In Section 7.2, we define the BPGD-h and BPGD-r algorithms (Algorithm 7.1 on page 113).

In Section 7.3, we show the simulations results (Figure 7.3 on page 114 and Figure 7.4 on page 115). We also discuss the dependence of the algorithmic performances when we change the temperature of the Gibbs measure.

7.1 Spatially-Coupled LDGM Codes

7.1.1 Lossy Compression of Symmetric Bernoulli Sources

Let $\underline{X} = \{X_1, X_2, \ldots, X_N\}$ represent a source of length N, where X_a, $a = 1, \ldots, N$ are i.i.d. Bernoulli($1/2$) random variables. We desire to encode this information into a memory which possesses only $M = RN$ bits, where $R \in [0, 1]$ is the compression rate. The encoder maps a source word $\underline{x} \in \{0, 1\}^N$ to one of 2^{NR} index words $\underline{u} \in \{0, 1\}^{NR}$. The decoding operation maps the stored sequence $\underline{u}$ to a reconstructed sequence $\widehat{\underline{x}}(\underline{u}) \in \{0, 1\}^N$. For a given pair $(\underline{x}, \widehat{\underline{x}})$, the quality of the reconstruction is measured by the Hamming distance

---

[1]The material and the presentation of this chapter are based on [89].

between the original sourceword and its reproduction

$$d_N\left(\underline{x}, \widehat{\underline{x}}\right) = \frac{1}{N}\sum_{a=1}^{N}|x_a - \widehat{x}_a|. \tag{7.1}$$

The quality of the reconstruction for the encoding/decoding scheme is the average of (7.1) over the source realizations and over the encoding/decoding output realization

$$D_N = \mathbb{E}_{\underline{X}, \widehat{\underline{X}}}\left[d\left(\underline{x}, \widehat{\underline{x}}\right)\right]. \tag{7.2}$$

For the symmetric Bernoulli source considered here, the average distortion of any encoding-decoding scheme is lower bounded by the Shannon rate-distortion curve (see Section 2.2)

$$D_{\mathrm{sh}}(R) = h_2^{-1}(1 - R) \tag{7.3}$$

where $h_2(q) = -q\log_2 q - (1 - q)\log_2(1 - q)$ is the binary entropy function. The rate-distortion curve is convex decreasing with $D_{\mathrm{sh}}(0) = 1/2$ and $D_{\mathrm{sh}}(1) = 0$.

7.1.2 Spatially-Coupled Low-Density Generator-Matrix Constructions

Our lossy source coding scheme is based on a spatially-coupled LDGM code ensemble. We first describe the *underlying* ensemble in terms of its graphical representation.

Underlying Poisson LDGM(p, R, N) Ensemble

These are bipartite graphs with a set C of N check nodes of constant degree p, a set V of M code-bit nodes with random degree, and a set E of edges connecting C and V. The ensemble of graphs is generated as follows: each edge emanating from a check node is connected uniformly at random to one of the code-bit nodes. The degree of code-bit nodes is a random variable with Binomial distribution $\mathrm{Bi}(pN, 1/NR)$. In the asymptotic regime of large N, M with $M/N = R$ the code-bit node degrees are i.i.d. Poisson distributed with an average degree p/R.

Spatially-Coupled SCLDGM(p, R, L, w, ρ) Ensemble

We first lay out a set of positions indexed by integers $z \in \mathbb{Z}$ on a one-dimensional line. This line represents a "spatial dimension". We fix a "window size" which is an integer $w \geq 1$. Consider L sets of check nodes each having ρ nodes, and locate the sets in positions 1 to L. Similarly, locate $L + w - 1$ sets of ρR code-bit nodes each, in positions 1 to $L + w - 1$. All checks have constant degree p, and each of the p edges emanating from a check at position $z \in \{1, \ldots, L\}$ is connected uniformly at
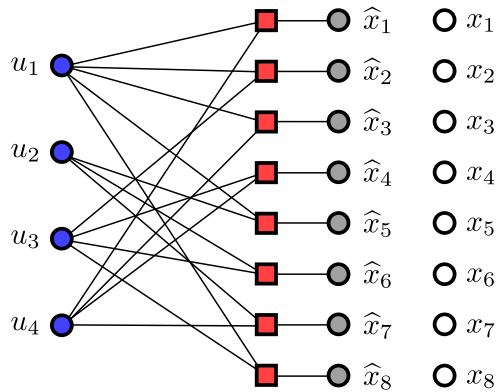
Figure 7.1: A bipartite graph from the underlying $\mathrm{LDGM}(2, 0.5, 8)$ ensemble. Here $N = 8$, $M = 4$ and $p = 2$. Labels represent code-bits u_i, reconstructed bits $\hat{x}_i$ and source bits x_i.
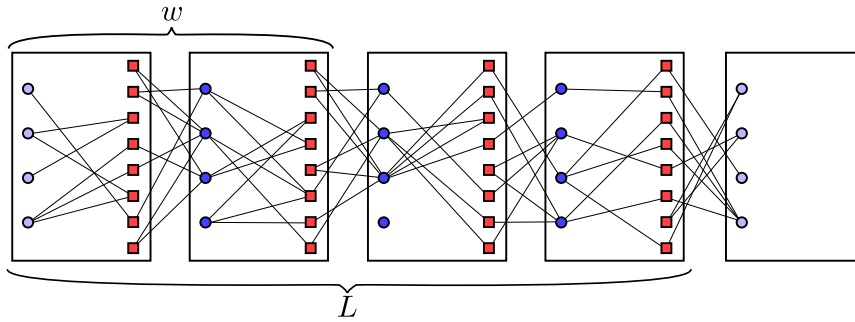
Figure 7.2: A graph from the $\mathrm{SCLDGM}(p = 2, R = 0.5, L = 4, w = 2, \rho = 8)$ ensemble. The code-bit nodes in boundary sets have smaller degree than the code-bit nodes in the other sets (in light blue in the figure).

random to code-bit nodes within the range $\{z, \ldots, z + w - 1\}$. It is easy to see that for $z \in \{w, \ldots, L - w + 1\}$, in the asymptotic limit $n \to +\infty$, the code-bit nodes have Poisson-distributed degrees with average p/R. For the remaining positions close to the boundary the average degree is reduced. More precisely for positions on the left side $z \in \{1, \ldots, w - 1\}$ the degree is asymptotically i.i.d. Poisson with average $p/R \times z/w$. For positions on the right side $z \in \{L + 1, \ldots, L + w - 1\}$ the degree is asymptotically Poisson with average $p/R \times (L + w - z)/w$. Figures 7.1 and 7.2 give a example of an underlying and a spatially-coupled graph.

Notation

Generic graphs from the ensembles will be denoted by Γ or $\Gamma(C, V, E)$. We will use letters a, b, c for check nodes and letters i, j, k for code-bit nodes of a given graph (from underlying or coupled ensembles). Following our convention we denote the total number of check nodes by $m = |C|$ and the total number of variable nodes by $n = |V|$. We

will often make use of the notation ∂a for the set of all code-bit nodes connected to $a \in C$, i.e.,$\partial a = \{i \in V \,|\, (i,a) \in E\}$. Similarly, for $i \in V$, $\partial i = \{a \in C \,|\, (i,a) \in E\}$. For spatially-coupled graphs the sets of nodes at a specified position z are C_z and V_z.

7.1.3 Decoding Rule and Optimal Encoding

A code-bit node $i \in V$ is associated with a code bit u_i. A check node $a \in C$ is associated with two types of bits: the reconstructed bit $\widehat{x}_a$ and the source bit x_a. By definition the source sequence has length N. So we have $|C| = N$ for the underlying ensembles, and $|C| = \rho L = N$ for the coupled ensembles. A compressed word $\underline{u}$ has length NR for the underlying ensemble, and $\rho R(L + w - 1)$ for the coupled ensemble. Thus the compression design rate is $R = M/N$ for the underlying ensemble, and it is $R_{\text{cou}} = \rho R(L + w - 1)/\rho L = R(1 + \frac{w-1}{L})$ for the coupled ensemble. The compression design rate of the coupled ensembles is slightly higher, due to the code-bit nodes at the boundary, but in the asymptotic regime $N, M \gg L \gg w$ the difference with the design rate R of the underlying ensemble vanishes.

Decoding Rule

The reconstruction mapping is given by the linear operation (modulo 2 sum)

$$\widehat{x}_a(\underline{u}) = \oplus_{i \in \partial a} u_i. \tag{7.4}$$

We do not investigate non-linear decoding rules, although the whole analysis developed here can be adapted to such rules. Source coding with such "non-linear check nodes" have been investigated for underlying $\text{LDGM}(p, R, N)$ ensembles [103].

Optimal Encoding

Given a source word $\underline{x}$, the optimal encoder seeks to minimize the Hamming distortion (7.1), and so searches among all $\underline{u} \in \{0,1\}^{NR}$ to find a configuration $\underline{u}^{\text{opt}}$ such that

$$\underline{u}^{\text{opt}}(\underline{x}) = \arg\min_{\underline{u}} d_N(\underline{x}, \widehat{\underline{x}}(\underline{u})). \tag{7.5}$$

The resulting minimal distortion is

$$d_{N,\text{opt}}(\underline{x}) = \min_{\underline{u}} d_N(\underline{x}, \widehat{\underline{x}}(\underline{u})). \tag{7.6}$$

Optimal Distortion of the Ensemble

A performance measure for the code ensemble is given by the *optimal distortion of the ensemble* (not to be confused with Shannon's optimal distortion)

$$D_{N,\text{opt}} = \mathbb{E}_{\Gamma,\underline{X}}[d_{N,\text{opt}}(\underline{x})] \tag{7.7}$$

where $\mathbb{E}_{\Gamma,\underline{X}}$ is an expectation over the graphical ensemble at hand and the symmetric Bernoulli source $\underline{X}$.

Finding the minimizers in (7.5) by exhaustive search takes exponential time in N; and there is no known efficient algorithmic procedure to solve the minimization problem. Nevertheless, the cavity method proposes a formula for the asymptotic value of (7.7) as $N \to +\infty$. It is conjectured that this formula is exact. We come back to this point at the end of Subsection 7.1.4.

7.1.4 Statistical Mechanics Formulation

We equip the configuration space $\{0,1\}^{NR}$ with the Gibbs measure (over $\underline{u} \in \{0,1\}^{NR}$)

$$
\begin{aligned}
\mu_\beta(\underline{u} \mid \underline{x}) &= \frac{1}{Z_\beta(\underline{x})} e^{-2\beta N d_N(\underline{x},\widehat{\underline{x}}(\underline{u}))} \\
&= \frac{1}{Z_\beta(\underline{x})} \prod_{a \in C} e^{-2\beta |x_a - \bigoplus_{i \in \partial a} u_i|}
\end{aligned}
\tag{7.8}
$$

where the "inverse temperature" $\beta > 0$ is a real parameter and

$$Z_\beta(\underline{x}) = \sum_{\underline{u}} e^{-2\beta N d_N(\underline{x},\widehat{\underline{x}}(\underline{u}))} \tag{7.9}$$

is the partition function. The expectation with respect to the Gibbs measure (7.8) is denoted by the bracket $\langle \cdot \rangle$. More precisely the average of a function $A(\underline{u})$ is

$$\langle A(\underline{u}) \rangle = \frac{1}{Z} \sum_{\underline{u} \in \{-1,+1\}^{RN}} A(\underline{u}) e^{-2\beta N d_N(\underline{x},\widehat{\underline{x}}(\underline{u}))}. \tag{7.10}$$

The function for which we are interested in the average value is the Hamming distortion between the pair $(\underline{x}, \widehat{\underline{x}}(\underline{u}))$.

Note that the minimizer $\underline{u}^{\text{opt}}$ in (7.5) maximizes this conditional distribution,

$$\underline{u}^{\text{opt}}(\underline{x}) = \text{argmax}_{\underline{u}} \mu_\beta(\underline{u} \mid \underline{x}). \tag{7.11}$$

The source coding problem can thus be interpreted as an estimation problem where $\underline{x}$ is an observation and $\underline{u}$ has to be estimated.

In the statistical mechanics interpretation, the Gibbs measure (7.8) is the distribution associated with a spin glass that is defined by the Hamiltonian $2Nd_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))$. This Hamiltonian can be seen as a cost-function for assignments of "dynamical" variables, the bits $u_i \in \{0, 1\}$. The Hamiltonian is random: for each realization of the source sequence $\underline{x}$ and the graph instance we have a different realization of the cost-function. The source and graph instance are qualified as "quenched" or "frozen" random variables, to distinguish them from dynamical variables.

Finding $\underline{u}^{\mathrm{opt}}$ amounts to finding the "minimum energy configuration". The minimum energy per node is equal to $2d_{N,\mathrm{opt}}$, and it is easy to check the identity (use 7.6 and 7.9)

$$2d_{N,\mathrm{opt}}(\underline{x}) = - \lim_{\beta \to \infty} \frac{1}{\beta N} \ln Z_\beta(\underline{x}). \tag{7.12}$$

As this identity already shows, a fundamental role is played by the quenched free energy

$$\overline{f}_N(\beta) = -\frac{1}{\beta N} \mathbb{E}_{\Gamma, \underline{X}} [\ln Z_\beta(\underline{x})]. \tag{7.13}$$

For example the average free energy enables us to compute the optimal distortion of the ensemble

$$2D_{N,\mathrm{opt}} = \lim_{\beta \to +\infty} \overline{f}_N(\beta). \tag{7.14}$$

There exists also another useful relationship that we will use between average distortion and free energy. Consider the "internal energy" defined as

$$\overline{u}_N(\beta) = 2\mathbb{E}_{\Gamma, \underline{X}} [\langle d_N(\underline{x}, \widehat{\underline{x}}(\underline{u})) \rangle] \tag{7.15}$$

It is straightforward to check that the internal energy can be computed from the quenched free energy (use (7.9), (7.14), (7.15))

$$\overline{u}_N(\beta) = \frac{\partial}{\partial \beta}(\beta \overline{f}_N(\beta)) \tag{7.16}$$

and that in the zero temperature limit it reduces to the average minimum energy or optimal distortion (use (7.6), (7.7), (7.15))

$$2D_{N,\mathrm{opt}} = \lim_{\beta \to +\infty} \overline{u}_N(\beta). \tag{7.17}$$

What is the relation between the quantities $\overline{f}_N(\beta)$, $\overline{u}_N(\beta)$, and $D_{N,\mathrm{opt}}$ for the underlying and coupled ensembles? The following theorem states that they are equal in the infinite block length limit. This limit is defined as

$$\lim_{N \to +\infty}$$

Table 7.1: Optimal distortion for $LDGM(p, R = 0.5)$ ensembles computed in [103]; The Shannon bound for $R = 0.5$ is $D_{\text{sh}} \approx 0.1100$.

p	3	4	5	6
D_{opt}	0.1179	0.1126	0.1110	0.1104

with M/N fixed for the underlying ensemble; and as

$$\lim_{N \to +\infty} = \lim_{L \to +\infty} \lim_{\rho \to +\infty}$$

with M/N fixed for the coupled ensemble. We stress that for the coupled ensemble the order of limits is important.

Theorem 7.1. *Consider the two ensembles*

$$LDGM(p, R, N) \quad \text{and} \quad SCLDGM(p, R, L, w, \rho)$$

for an even p and R. Then the respective limits

$$\lim_{N \to +\infty} \overline{f}_N(\beta), \quad \lim_{N \to +\infty} \overline{u}_N(\beta) \quad \text{and} \quad \lim_{N \to +\infty} D_{N,\text{opt}}$$

exist for all $\beta \geq 0$ and have identical values for the two ensembles.

This theorem is proved in [83] for the max-XORSAT problem. The proof in [83] does not depend on the check nodes density N/M, so that it applies verbatim to the present setting. We conjecture that this theorem is valid for a wider class of graph ensembles. In particular we expect that it is valid for odd p and also for the regular LDGM ensembles (see [104] for similar results concerning LDPC codes).

It is conjectured that the one-step-replica-symmetry-breaking-formulas (1-RSB), obtained from the cavity method [105], for the $N \to +\infty$ limit of the free, internal and ground state energies are exact. Remarkably, it has been proven [106], using an extension of the Guerra-Toninelli interpolation bounds [40], that these formulas are upper bounds. The 1-RSB formulas allow to numerically compute [103], using population dynamics, $D_{\text{opt}} \equiv \lim_{N \to +\infty} D_{N,\text{opt}}$. As an illustration, Table 7.1 reproduces D_{opt} for increasing check degrees. Note that D_{opt} approaches D_{sh} as the degrees increase. One observes that with increasing degrees the optimal distortion of the ensemble attains Shannon's rate-distortion limit.

7.2 Belief Propagation Guided Decimation

Since the optimal encoder (7.5) is intractable, we investigate suboptimal low-complexity encoders. We focus on two encoding algorithms based on the belief propagation (BP)

equations supplemented with a decimation process.

7.2.1   Belief Propagation Equations

Instead of estimating the block $\underline{u}$ (as in (7.5)) we would like to estimate bits u_i with the help of the marginals

$$\mu_i(u_i \mid \underline{x}) = \sum_{\underline{u} \backslash u_i} \mu_\beta(\underline{u} \mid \underline{x}) \tag{7.18}$$

where the sum is over $u_1, \ldots u_N$ with u_i omitted. However computing the exact marginals involves a sum with an exponential number of terms and is also intractable. For sparse random graphs, when the size of the graph is large, any finite neighborhood of a node i is a tree with high probability. Computing the marginals on a tree-graph can be done exactly and leads to the BP equations. Therefore it seems reasonable to compute the BP marginal distribution in place of (7.18),

$$\mu_i^{\mathrm{BP}}(u_i \mid \underline{x}) = \frac{1}{2 \cosh \beta \eta_i} e^{\beta(-1)^{u_i} \eta_i} \tag{7.19}$$

where the biases η_i are computed from solutions of the BP equations. The latter are a set of fixed-point equations involving $2 \, |E|$ real valued messages $\eta_{i \to a}$ and $\widehat{\eta}_{a \to i}$ associated with the edges $(i, a) \in E$ of the graph. We have

$$\begin{cases} \widehat{\eta}_{a \to i} & = (-1)^{x_a} \beta^{-1} \tanh^{-1}\left(\tanh \beta \prod_{j \in \partial a \backslash i} \tanh \beta \eta_{j \to a}\right) \\ \eta_{i \to a} & = \sum_{b \in \partial i \backslash a} \widehat{\eta}_{b \to i} \end{cases} \tag{7.20}$$

and

$$\eta_i = \sum_{a \in \partial i} \widehat{\eta}_{a \to i}. \tag{7.21}$$

For any solution of the BP equations one could consider the estimator

$$\widehat{u}_i^{\mathrm{BP}} = \mathrm{argmax} \mu_i^{\mathrm{BP}}(u_i \mid \underline{x}) = \begin{cases} \frac{1}{2}(1 + \mathrm{sign} \tanh \beta \eta_i), & \text{if } \eta_i \neq 0 \\ \mathrm{Bernoulli}(\frac{1}{2}), & \text{if } \eta_i = 0. \end{cases} \tag{7.22}$$

One could then use the decoding rule (7.4) to determine a reconstructed word and the corresponding distortion. Unfortunately, given $\underline{x}$, the number of solutions of the BP equations which lead to a roughly identical distortion grows exponentially large in N. This has an undesirable consequence: it is not possible to pick the relevant solution by a plain iterative method. To get around this problem, the BP iterations are equipped with a heuristic decimation process.

7.2.2   Decimation Process

We start with a description of the *first round* of the decimation process. Let Γ, $\underline{x}$ be a graph and source instance. Fix an initial set of messages $\eta_{i\rightarrow a}^{(0)}$ at time $t = 0$. Iterate the BP equations (7.20) to get a set of messages $\eta_{i\rightarrow a}^{(t)}$ and $\widehat{\eta}_{a\rightarrow i}^{(t)}$ at time $t \geq 0$. Let $\epsilon > 0$ be some small positive number and T some large time. Define a *decimation instant* t_{dec} as follows:

- i) If the total variation of messages does not change significantly in two successive iterations,

$$\frac{1}{|E|} \sum_{(i,a)\in E} |\widehat{\eta}_{a\rightarrow i}^{(t)} - \widehat{\eta}_{a\rightarrow i}^{(t-1)}| < \epsilon \qquad (7.23)$$

  for some $t < T$, then $t_{\mathrm{dec}} = t$.

- ii) If (7.23) does not occur for all $t \leq T$ then $t_{\mathrm{dec}} = T$.

At instant t_{dec} each code-bit has a bias given by $\eta_i^{(t_{\mathrm{dec}})}$. *Select* and *fix one particular code-bit* i_{dec} according to a decision rule

$$(i_{\mathrm{dec}}, u_{i_{\mathrm{dec}}}) \leftarrow \mathcal{D}(\underline{\eta}^{(t_{\mathrm{dec}})}). \qquad (7.24)$$

The precise decision rules that we investigate are described in the next paragraph. At this point, update $x_a \leftarrow x_a \oplus u_{i_{\mathrm{dec}}}$ for all $a \in \partial i_{\mathrm{dec}}$, and decimate the graph $\Gamma \leftarrow \Gamma \setminus i_{\mathrm{dec}}$. This defines a new graph and source instance, on which we repeat a *new round*. The initial set of messages of the new round is the one obtained at time t_{dec} of the previous round.

7.2.3   Belief Propagation Guided Decimation

The decision rule (7.24) involves two choices. One has to choose i_{dec} and then set $u_{i_{\mathrm{dec}}}$ to some value. Let us first describe the choice of i_{dec}.

We evaluate the maximum bias

$$B_{t_{\mathrm{dec}}} = \max_{i\in V} |\eta_i^{(t_{\mathrm{dec}})}| \qquad (7.25)$$

at each decimation instant. If $B_{t_{\mathrm{dec}}} > 0$, we consider the set of nodes that maximize (7.25), we choose one of them uniformly at random, and call it i_{dec}. If $B_{t_{\mathrm{dec}}} = 0$ and we have a graph of the *underlying ensemble*, we choose a node uniformly at random from $\{1, \ldots m\}$, and call it i_{dec}. If $B_{t_{\mathrm{dec}}} = 0$ and we have a graph of the *coupled ensemble*, we choose a node uniformly at random from the w left-most positions of the current graph, and call it i_{dec}. Note that because the graph gets decimated, the w left-most positions of the current graph form a moving boundary.

With the above choice of decimation node the encoding process is seeded at the boundary each time the BP biases fail to guide the decimation process. We have checked empirically that if we choose i_{dec} uniformly at random from the whole chain (for coupled graphs) the performance is not improved by coupling. In [107] we adopted periodic boundary conditions and the seeding region was set to an arbitrary window of length w at the beginning of the process, which then generated its own boundary at a later stage of the iterations.

We now describe two decision rules for setting the value of $u_{i_{\text{dec}}}$ in (7.24).

1. **Hard decision**
$$u_{i_{\text{dec}}} = \begin{cases} \theta(\eta_{i_{\text{dec}}}^{(t_{\text{dec}})}), & \text{if } B_{t_{\text{dec}}} > 0 \\ \text{Bernoulli}(\frac{1}{2}), & \text{if } B_{t_{\text{dec}}} = 0 \end{cases} \qquad (7.26)$$

   where $\theta(.)$ is the Heaviside step function. We call this rule and the associated algorithm BPGD-h.

2. **Randomized decision**

$$u_{i_{\text{dec}}} = \begin{cases} 0, & \text{with prob } \frac{1}{2}(1 + \tanh \beta \eta_{i_{\text{dec}}}^{(t_{\text{dec}})}) \\ 1, & \text{with prob } \frac{1}{2}(1 - \tanh \beta \eta_{i_{\text{dec}}}^{(t_{\text{dec}})}). \end{cases} \qquad (7.27)$$

   In other words, we fix a code-bit randomly with a probability given by its BP marginal (7.19). We call this rule and the associated algorithm BPGD-r.

Algorithm 7.1 summarizes the BPGD algorithms for all situations.

7.2.4 Initialization and Choice of Parameters ϵ, T and β

We initialize $\eta_{i \to a}^{(0)}$ to zero just at the beginning of the algorithm. After each decimation step, rather than resetting messages to zero we continue with the previous messages. We have observed that resetting the messages to zero does not lead to very good results.

The parameters ϵ and T are in practice set to $\epsilon = 0.01$ and $T = 10$. The simulation results do not seem to change significantly when we take ϵ smaller and T larger.

The performance of the BPGD algorithm does depend on the choice of β which enters in the BP equations (7.20) and in the randomized decision rule (7.27). It is possible to optimize over β. This is important in order to approach (with coupled codes) the optimal distortion of the ensemble, and furthermore to approach the Shannon bound in the large degree limit. While we do not have a first principle theory for the optimal choice of β, we provide empirical observations in Section 7.3. We observe that knowing the dynamical and condensation (inverse) temperatures predicted by the cavity method enables us to

---

Algorithm 7.1 BP Guided Decimation Algorithm

Generate a graph instance $\Gamma(C, V, E)$ from the underlying or coupled ensembles;

Generate a Bernoulli symmetric source word $\underline{x}$;

Set $\eta_{i \to a}^{(0)} = 0$ for all $(i, a) \in E$;

while $V \neq \emptyset$ **do**

   Set $t = 0$; **while** *Convergence* (7.23) *is not satisfied and $t < T$* **do**

      Update $\hat{\eta}_{a \to i}^{(t)}$ according to (7.20) for all $(a, i) \in E$;

      Update $\eta_{i \to a}^{(t+1)}$ according to (7.20) for all $(i, a) \in E$;

      $t \leftarrow t + 1$.

   Compute bias $\eta_i^{(t)} = \sum_{a \in \partial i} \hat{\eta}_{a \to i}^{(t)}$ for all $i \in V$

   Find $B = \max_{i \in V} |\eta_i^{(t)}|$;

   if $B = 0$ **then** For an instance from the underlying ensemble randomly pick a code-bit i from V. For a graph from the coupled ensemble randomly pick a code-bit from the w left-most positions of Γ and fix it randomly to 0 or 1; ;

   else

      Select $i = \arg\max_{i \in V} |\eta_i^{(t)}|$; Fix a value for u_i according to rule (7.26) or (7.27);

   Update $x_a \leftarrow x_a \oplus u_i$ for all $a \in \partial i$;

   Reduce the graph $\Gamma \leftarrow \Gamma \setminus \{i\}$;

---

make an educated guess for an estimate of the optimal β. In particular, it turns out that for coupled codes with large degrees the best β approaches the information theoretic test-channel value.

7.2.5 Computational Complexity

It is not difficult to see that the complexity of the plain BPGD algorithm 7.1 is $O(N^2)$ for both the underlying and coupled ensembles. The reason is that the BP algorithm requires to update a number of BP messages proportional to N and the decimation process requires to run the BP algorithm after each decimation. The complexity for the coupled ensemble can be rewritten as $O(\rho^2 L^2)$. By employing window decoding [108, 109], one can reduce the complexity of the BP algorithm and the complexity of the coupled ensemble becomes $O(\rho^2 L w)$ with almost the same performance. This can be further reduced by noticing that the BP messages do not change significantly between two decimation steps. As a result, we could decimate $\delta\rho$ code-bits at each step for some small δ, so that the complexity of the BPGD algorithm becomes $O(\rho L w / \delta)$. To summarize, it is possible to get a linear in block length complexity without significant loss in performance.
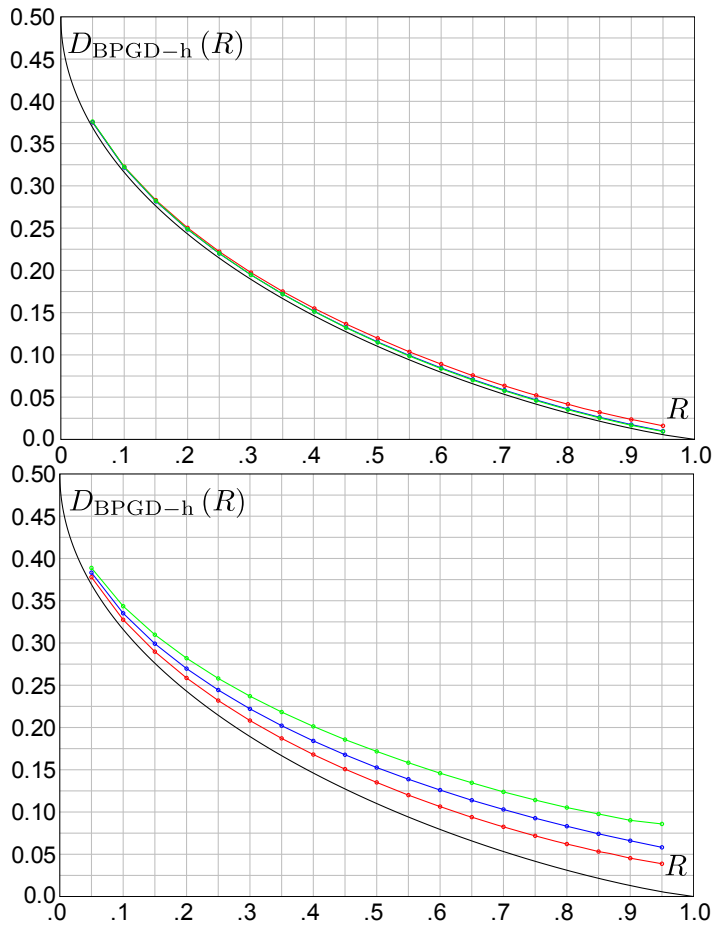
Figure 7.3: The BPGD-h algorithmic distortion versus compression rate R compared to the Shannon rate-distortion curve at the bottom. Points are obtained by optimizing over β and averaging over 50 instances. Top: spatially-coupled SCLDGM($p, R, L = 64, w = 3, \rho = 2000$) ensembles for $p = 3, 4, 5$ (top to bottom). Bottom: LDGM($p, R, N = 128000$) ensembles for $p = 3, 4, 5$ (bottom to top).

7.3   Simulations

In this section we discuss the performance of the BPGD algorithms. The comparison between underlying ensembles, coupled ensembles and the Shannon rate-distortion curve is illustrated. The role played by the parameter β is investigated.

7.3.1   BPGD Performance and Comparison to the Shannon Limit

Figures 7.3 and 7.4 display the average distortion $D_{\mathrm{BPGD}}(R)$ obtained by the BPGD algorithms (with hard and randomized decision rules) as a function of R, and compares it to the Shannon limit $D_{\mathrm{sh}}(R)$ given by the lowest curve. The distortion is computed for fixed R and for 50 instances, and the empirical average is taken. This average is then
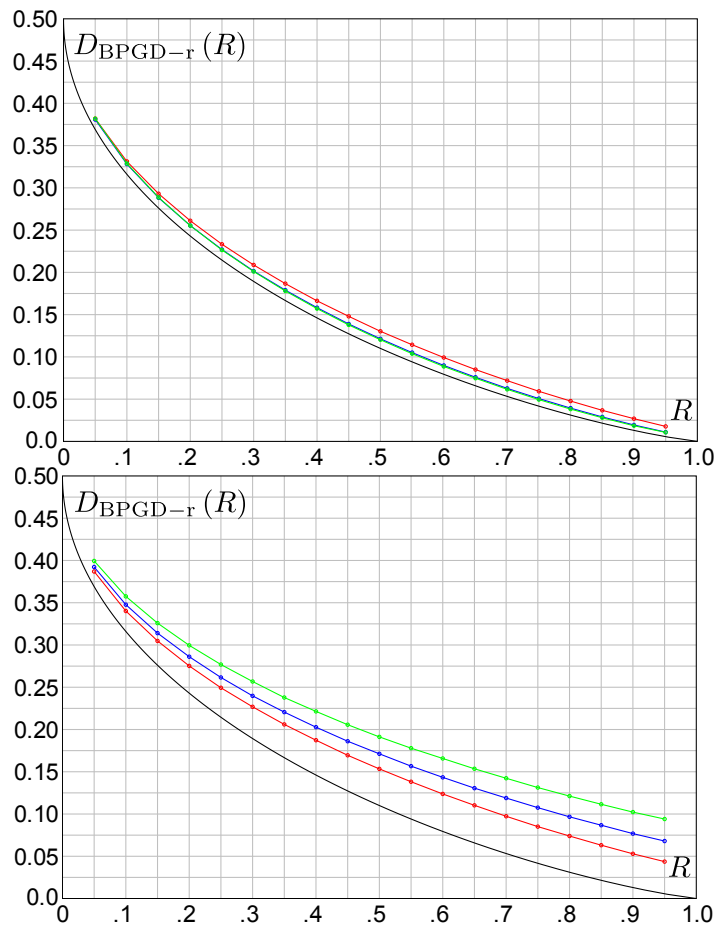
Figure 7.4: The BPGD-r algorithmic distortion versus compression rate R compared to the Shannon rate-distortion curve at the bottom. Points are obtained by choosing $\beta = \beta_{sh} = \frac{1}{2}\log(\frac{1-D_{sh}}{D_{sh}})$ and averaging over 50 instances. Continuous lines are a guide to the eye. Top: spatially-coupled SCLDGM$(p, R, L = 64, w = 3, \rho = 2000)$ ensembles for $p = 3, 4, 5$ (top to bottom). Bottom: LDGM$(p, R, N = 128000)$ ensembles for $p = 3, 4, 5$ (bottom to top).

optimized over β, giving one dot on the curves (continuous curves are a guide to the eye).

The plots at the bottom of Figures 7.3 and 7.4 are for the underlying ensembles with $p = 3, 4, 5$ and $N = 128000$. We observe that as the check degree increases the BPGD performance gets worse. But recall from Table 7.1 that with increasing degrees the optimal distortion of the ensemble (not shown explicitly on the plots) gets better and approaches the Shannon limit. Thus the situation is similar to the case of LDPC codes where the BP threshold gets worse with increasing degrees, while the MAP threshold approaches Shannon capacity.

The plots at the top of Figures 7.3 and 7.4 show the algorithmic performance for the coupled ensembles with $p = 3, 4, 5$, $\rho = 2000$, $w = 3$, and $L = 64$ (so again a total length

115
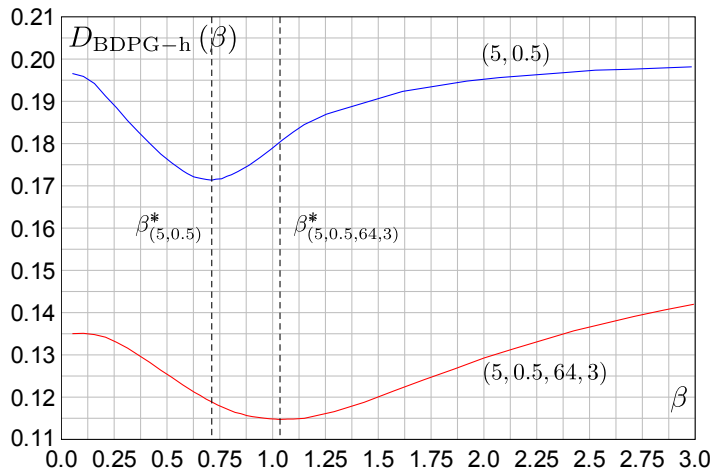
Figure 7.5:   The BPGD-h algorithmic distortion versus β. Results are obtained for the SCLDGM($p = 5, R = 0.5, L = 64, w = 3, \rho = 2000$) and LDGM($5, 0.5, 128000$) ensembles. Results are averaged over 50 instances. The minimum distortion occurs at $\beta^*_{(5,0.5,64,3)} \approx 1.03 \pm 0.01$ and $\beta^*_{(5,0.5)} \approx 0.71 \pm 0.01$.

of $N = 128000$). We see that the BPGD performance approaches the Shannon limit as the degrees increase. One obtains a good performance, for a range of rates, without any optimization on the degree sequence of the ensemble, and with simple BPGD schemes.

The simulations, suggest the following. Look at the regime $\rho \gg L \gg w \gg 1$. When these parameters go to infinity in the specified order *for the coupled ensemble,* $D_{\mathrm{BPGD}}(R)$ approaches $D_{\mathrm{opt}}(R)$. In words, the algorithmic distortion approaches the optimal distortion of the ensemble. When furthermore $p \to +\infty$ after the other parameters, $D_{\mathrm{BPGD}}(R)$ approaches $D_{\mathrm{sh}}(R)$. At this point it is not possible to assess from the simulations whether these limits are exactly attained.

7.3.2   The Choice of the Parameter β

We discuss the empirical observations for the dependence of the curves $D_{\mathrm{BPGD}}(\beta, R)$ on β at fixed rate. We illustrate our results for $R = 1/2$ and with the underlying LDGM($p = 5, R = 0.5, N = 128000$) and coupled SCLDGM($p = 5, R = 0.5, w = 3, L = 64, \rho = 2000$) ensembles.

In Figure 7.5 we plot the distortion $D_{\mathrm{BPGD-h}}(\beta, R = 1/2)$ of the hard decision rule. For *all* values of $0 < \beta < 3$, the algorithmic distortion $D_{\mathrm{BPGD-h}}(\beta, R)$ of the coupled ensemble is below the corresponding curve of the underlying ensemble. The most important feature is a clear minimum at a value β^* which is rate dependent. The rate-distortion curve for the hard decision rule in Figure 7.3 is computed at this β^* and is the result of the optimization

$$D_{\mathrm{BPGD-h}}(R) = \min_{\beta > 0} D_{\mathrm{BPGD-h}}(\beta, R). \tag{7.28}$$
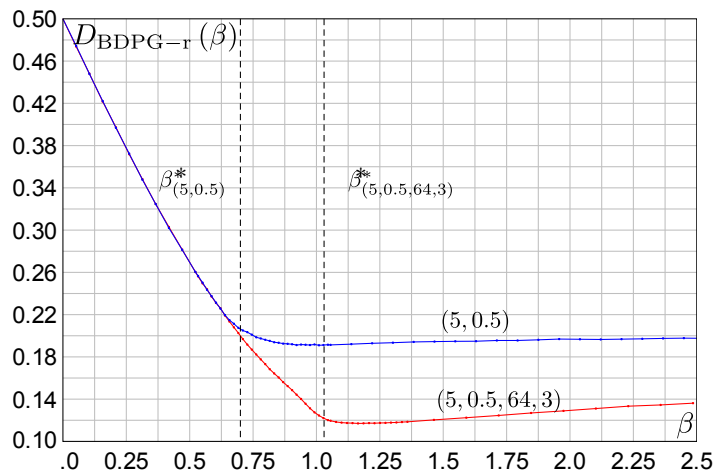
Figure 7.6: The BPGD-r algorithmic distortion versus β. Results are obtained for SCLDGM($p = 5, R = 0.5, L = 64, w = 3, \rho = 2000$) and LDGM($5, 0.5, 128000$) ensemble. Results are averaged over 50 instances. The values β^* of Figure 7.5 are reported for comparison.

We observe that the optimal value β^*_{cou} for the coupled ensemble is always larger than β^*_{un} for the underlying ensemble. Moreover we observe that as the check-degree p increases β^*_{un} tends to zero, whereas β^*_{cou} saturates to $\beta_{\mathrm{sh}}(R)$ where

$$\beta_{\mathrm{sh}}(R) \equiv \frac{1}{2} \ln \left(\frac{1 - D_{\mathrm{sh}}(R)}{D_{\mathrm{sh}}(R)} \right). \tag{7.29}$$

This is the information theoretic value corresponding to the amplitude of half log-likelihood ratios of a test-BSC with the noise tuned at capacity.

In Figure 7.6 we plot the curve $D_{\mathrm{BPGD-r}}(\beta, R = 1/2)$ for the randomized algorithm. The behavior of the underlying and coupled ensemble have the same flavor. The curves are first decreasing with respect to β and then flatten. The minimum is reached in the flattened region and as long as β is chosen in the flat region, the optimized distortion is not very sensitive to this choice. We take advantage of this feature, and compute the rate-distortion curve of the randomized decision rule at a predetermined value of β. This has the advantage of avoiding optimizing over β. For reasons that are discussed in Section 8.4 a good choice is to take $\beta_{\mathrm{sh}}(R)$ given by Equ. 7.29. With these considerations the distortion curve in Figure 7.4 is

$$D_{\mathrm{BPGD-r}}(R) = D_{\mathrm{BPGD-r}}(\beta_{\mathrm{sh}}, R). \tag{7.30}$$

117
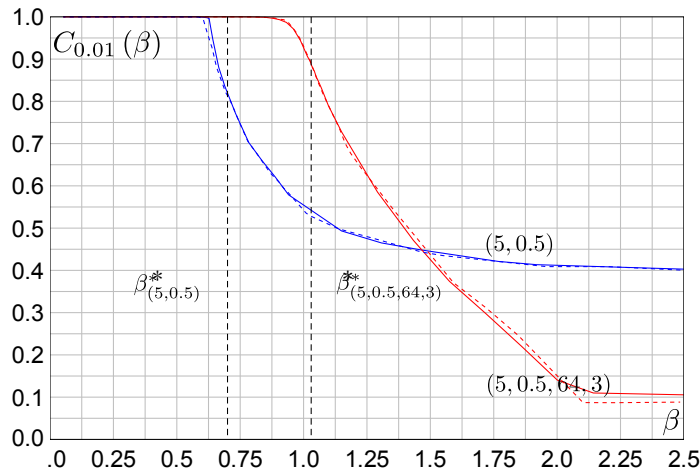
Figure 7.7: $C_{0.01}(\beta)$ versus β. Empirical convergence probability for SCLDGM($p = 5, R = 0.5, L = 64, w = 3, \rho = 2000$) and LDGM($5, 0.5, 128000$) ensembles. Solid (resp. dashed) lines are for the hard (resp. random) decision rule. Results are averaged over 50 instances.

7.3.3   Convergence

We have tested the convergence of the BPGD algorithms for both decision rules. We compute an *empirical probability of convergence* $C_{\epsilon,T}(\beta)$ defined as the fraction of decimation rounds that results from the convergence condition (7.23). In other words $C_{\epsilon,T}(\beta) = 1$ means that at every round of the decimation process the BP update rules converge in less than T iterations to a fixed point of the BP equations (7.20) up to a precision ϵ. Figure 7.7 shows $C_{\epsilon,T}(\beta)$ at $(\epsilon, T) = (0.01, 10)$ for the underlying and coupled ensembles. The hard decision rule is represented by solid lines and the random decision rule by dashed lines. The first observation is that both decision rules have nearly identical behaviors. This is not a priori obvious since the decimation rules are different, and as a result the graph evolves differently for each rule during the decimation process. This suggest that the convergence of the algorithms essentially depends on the convergence of the plain BP algorithm. The second observation is that the values of β where $C_{\epsilon,T}(\beta)$ drops below one are roughly comparable to the values where $D_{\mathrm{BPGD-r}}$ flattens and where $D_{\mathrm{BPGD-h}}$ attains its minimum.

8 A Perspective from the Cavity Method

In this chapter[1], we derive the phase diagram of the Gibbs distribution (7.8) using the cavity method. It is natural to expect that the behavior of belief propagation based algorithms should be related in one way or another to the phase diagram of the Gibbs distribution. As we vary β the nature of the Gibbs measure and the geometry of the space of its typical configurations changes at special *dynamical* and *condensation* thresholds β_d and β_c (see Section 1.5).

In Section 8.1, we give a primer on the cavity method.

In Section 8.2, we show the technical derivations for the present problem.

In Section 8.3, we display the set of fixed-point equations needed to compute the complexity of the coupled ensemble (Equations (8.51) on page 131 and (8.52) on page 131 and Equations (8.54) on page 131 and (8.55) on page 132). We will see that in the present problem for $\beta < \beta_c$ there is a further simplification of the cavity equations. With this extra simplification, the cavity equations reduce to standard density evolution equations associated with a coupled LDGM code over a BSC test-channel (Equations (8.63) on page 136 and (8.55) on page 132).

In Section 8.4, we discuss how the dynamical and condensation thresholds are affected by spatial coupling. We also discuss some heuristic insights that enable us to understand why the Shannon limit is approached with the BPGD algorithm for coupled ensembles with large check degrees.

In Section 8.5, we discuss some remaining open problems.

---

[1]The material and the presentation of this chapter is based on [89].

8.1   The Cavity Method

The treatment given here applies to *single instances*. Let $\Gamma = (V, C, E)$ be a factor graph which is assumed to be locally tree-like. We attach spins σ_j, $j \in V$, to variable nodes and constraint functions $\psi_a\left(\{\sigma_i, i \in \partial a\}\right)$, $a \in C$, to check nodes. We sometimes use the notation $\underline{\sigma}_{\partial a} = \{\sigma_i, i \in \partial a\}$ as a shorthand. The formalism developed in this appendix is valid for general spin variables belonging to a finite alphabet $\sigma_j \in \mathcal{X}$. The constraint functions depend only on the set of spins connected to a. We are interested in the thermodynamic limit where $|V| = n$ and $|C| = m$ tend to infinity and the ratio m/n is kept fixed. We consider the general class of Gibbs distributions of the form

$$\mu\left(\underline{\sigma}\right) = \frac{1}{Z} \prod_{a \in C} \psi_a\left(\underline{\sigma}_{\partial a}\right), \tag{8.1}$$

where Z is the partition function. The free energy of an instance is defined as usual

$$f\left(\beta\right) = -\frac{1}{n\beta} \ln Z\left(\beta\right). \tag{8.2}$$

One of the goals of the cavity method is to compute this free energy in the limit $n \to +\infty$.

Let us first outline the general strategy. For locally tree-like graphs, one can compute the marginals for a given node by restricting the measure to a tree. In the absence of long range correlations[2] the marginal does not depend on the boundary conditions at the leaf nodes, and the BP equations have one relevant solution. The BP marginals then constitute a good description of the measure (8.1). In particular, the true free energy is well approximated by replacing this solution in the Bethe free energy functional. As the control parameters vary, long range correlations could appear. In such a situation the marginals computed on a tree will depend on the boundary conditions at the leaf nodes, and the BP equations will have many relevant solutions yielding nearly the same Bethe free energy. The cavity method assumes that the measure (8.1) is then described by a convex superposition of "extremal measures". There could be a large number of extremal measures. A good proxy for the extremal measures is given by the BP marginals. The convex superposition of extremal measures yields a new statistical model on the same factor graph, the so-called *level-one model*. Assuming that the level one model does not display long range correlations, one can solve it using BP equations and the Bethe free energy. Otherwise, the cavity method iterates the previous considerations and constructs a level-two model. However, this usually becomes bewildering and one stops at the first level. In the following paragraphs we give a concrete implementation of these ideas.

We recall that the BP equations are the set of fixed-point equations satisfied by messages

---

[2]More precisely point-to-set correlations [44].

$$\{\nu_{i\to a}, \widehat{\nu}_{a\to i}\} = (\underline{\nu}, \underline{\widehat{\nu}}),$$

$$\widehat{\nu}_{a\to i} = \widehat{g}_{\mathrm{BP}}\left(\{\nu_{j\to a}\}_{j\in\partial a\backslash i}\right), \ \nu_{i\to a} = g_{\mathrm{BP}}\left(\{\widehat{\nu}_{b\to i}\}_{b\in\partial i\backslash a}\right), \tag{8.3}$$

where

$$\widehat{g}_{\mathrm{BP}}\left(\{\nu_{j\to a}\}_{j\in\partial a\backslash i}\right) = \frac{\sum_{\underline{\sigma}_{\partial a\backslash i}}\psi_a\left(\underline{\sigma}_{\partial a}\right)\prod_{j\in\partial a\backslash i}\nu_{j\to a}\left(\sigma_j\right)}{\sum_{\underline{\sigma}_{\partial a}}\psi_a\left(\underline{\sigma}_{\partial a}\right)\prod_{j\in\partial a\backslash i}\nu_{j\to a}\left(\sigma_j\right)}$$

$$g_{\mathrm{BP}}\left(\{\widehat{\nu}_{b\to i}\}_{b\in\partial i\backslash a}\right) = \frac{\prod_{b\in\partial i\backslash a}\widehat{\nu}_{b\to i}\left(\sigma_i\right)}{\sum_{\sigma_i}\prod_{b\in\partial i\backslash a}\widehat{\nu}_{b\to i}\left(\sigma_i\right)}. \tag{8.4}$$

When there is only one relevant solution, the BP marginal for σ_i is

$$\mu_i^{\mathrm{BP}}(\sigma_i) = \frac{\prod_{a\in\partial i}\nu_{a\to i}(\sigma_i)}{\sum_{\sigma_i}\prod_{a\in\partial i}\nu_{a\to i}(\sigma_i)}. \tag{8.5}$$

The set of messages is a proxy for the measure (8.1) in the sense that in principle one can "reconstruct" the measure from this set. The Bethe free energy functional which approximates $f(\beta)$ is given by

$$f^{\mathrm{Bethe}}\left(\underline{\nu}, \underline{\widehat{\nu}}\right) = \frac{1}{n}\left\{\sum_{i\in V}F_i + \sum_{a\in C}F_a - \sum_{(i,a)\in E}F_{ai}\right\} \tag{8.6}$$

where

$$F_i\left(\{\widehat{\nu}_{b\to i}\}_{b\in\partial i}\right) = -\frac{1}{\beta}\ln\sum_{\sigma_i}\prod_{b\in\partial i}\widehat{\nu}_{b\to i}\left(\sigma_i\right)$$

$$F_a\left(\{\nu_{j\to a}\}_{j\in\partial a}\right) = -\frac{1}{\beta}\ln\sum_{\underline{\sigma}_{\partial a}}\psi_a\left(\underline{\sigma}_{\partial a}\right)\prod_{j\in\partial a}\nu_{j\to a}\left(\sigma_j\right)$$

$$F_{ai}\left(\nu_{i\to a}, \widehat{\nu}_{a\to i}\right) = -\frac{1}{\beta}\ln\sum_{\sigma_i}\nu_{i\to a}\left(\sigma_i\right)\widehat{\nu}_{a\to i}\left(\sigma_i\right). \tag{8.7}$$

As explained in Section 1.5, in the presence of long range correlations this formalism is too simplistic. The cavity method assumes that:

1. The Gibbs distribution (8.1) is a convex sum of extremal measures.

2. To leading exponential order, the number of solutions of the BP equations is equal to the number of extremal measures.

3. The free energy of an extremal measure is well approximated by the Bethe free energy of the BP fixed-point.

These assumptions suggest that the Gibbs distribution (8.1) is well approximated by the

following convex superposition

$$\mu\left(\underline{\sigma}\right) \approx \frac{1}{Z} \sum_{\left(\underline{\nu},\widehat{\underline{\nu}}\right)\in\mathrm{BP}} e^{-\beta n f^{\mathrm{Bethe}}\left(\underline{\nu},\widehat{\underline{\nu}}\right)} \mu_{\left(\underline{\nu},\widehat{\underline{\nu}}\right)}\left(\underline{\sigma}\right) \tag{8.8}$$

The measures $\mu_{\left(\underline{\nu},\widehat{\underline{\nu}}\right)}$ are the ones whose marginals are given by the BP marginals computed from $(\underline{\nu},\widehat{\underline{\nu}})$. They play the role of the "extremal measures". The sum is over solutions of the BP equations. In principle one should sum only over stable solutions, i.e., local minima of the Bethe free energy. However at low temperatures these are expected to be exponentially more numerous than the other critical points and it is assumed to be a good approximation to sum over all BP solutions. The normalization factor yields the partition function

$$Z \approx \sum_{\left(\underline{\nu},\widehat{\underline{\nu}}\right)\in\mathrm{BP}} e^{-\beta n f^{\mathrm{Bethe}}\left(\underline{\nu},\widehat{\underline{\nu}}\right)}. \tag{8.9}$$

In order to compute this partition function and uncover the properties of the convex decomposition (8.8) we introduce the level-one statistical mechanical model. The dynamical variables of this model are the BP messages $(\underline{\nu},\widehat{\underline{\nu}})$. According to (8.8), (8.9) the probability distribution over $(\underline{\nu},\widehat{\underline{\nu}})$ is

$$\mu_{\mathrm{level-1}}\left(\underline{\nu},\widehat{\underline{\nu}}\right) = \frac{e^{-\beta n f^{\mathrm{Bethe}}\left(\underline{\nu},\widehat{\underline{\nu}}\right)}}{Z_{\mathrm{level-1}}} \mathbb{I}\left(\left(\underline{\nu},\widehat{\underline{\nu}}\right)\in\mathrm{BP}\right), \tag{8.10}$$

and

$$Z_{\mathrm{level-1}} = \sum_{\left(\underline{\nu},\widehat{\underline{\nu}}\right)\in\mathrm{BP}} e^{-\beta n f^{\mathrm{Bethe}}\left(\underline{\nu},\widehat{\underline{\nu}}\right)}, \tag{8.11}$$

The level-one free energy is defined as usual,

$$f_{\mathrm{level-1}}(\beta) = -\frac{1}{\beta n} \ln Z_{\mathrm{level-1}}. \tag{8.12}$$

From (8.9) it should be clear that $f(\beta) \approx f_{\mathrm{level-1}}(\beta)$. The average Bethe free energy, or level-one *internal energy*, is given by

$$\varphi_{\mathrm{int}}(\beta) = \frac{1}{n}\langle f^{\mathrm{Bethe}}[\underline{\nu},\widehat{\underline{\nu}}]\rangle_{\mathrm{level-1}} \tag{8.13}$$

Here the bracket denotes the average with respect to (8.10).

One also needs to compute the Shannon-Gibbs entropy $\Sigma(\beta)$ of $\mu_{\mathrm{level-1}}$. An important "trick" is to replace the *explicit β* dependence in (8.10), (8.11), (8.12) by βx where x is for the moment an arbitrary parameter[3]. This parameter turns out to play a crucial role and is called the Parisi parameter. This gives us an x-dependent level-one auxiliary

---

[3]Note that there is also an *implicit β* dependence in $f^{\mathrm{Bethe}}[\underline{\nu},\widehat{\underline{\nu}}]$.

model

$$\mu_{\text{level}-1}\left(\underline{\nu}, \widehat{\underline{\nu}}; x\right) = \frac{e^{-\beta x n f^{\text{Bethe}}\left(\underline{\nu}, \widehat{\underline{\nu}}\right)}}{Z_{\text{level}-1(x)}} \mathbb{I}\left(\left(\underline{\nu}, \widehat{\underline{\nu}}\right) \in \text{BP}\right), \tag{8.14}$$

and

$$Z_{\text{level}-1}(x) = \sum_{\left(\underline{\nu}, \widehat{\underline{\nu}}\right) \in \text{BP}} e^{-\beta x n f^{\text{Bethe}}\left(\underline{\nu}, \widehat{\underline{\nu}}\right)}, \tag{8.15}$$

and also

$$\varphi_{\text{level}-1}(\beta; x) = -\frac{1}{\beta x n} \ln Z_{\text{level}-1}(x). \tag{8.16}$$

It is then a matter of simple algebra to check that the Shannon-Gibbs entropy $\Sigma(\beta)$ is given by

$$\Sigma(\beta) = \Sigma(\beta; x) \equiv \beta x^2 \frac{\partial}{\partial x} \varphi_{\text{level}-1}(\beta; x) |_{x=1}, \tag{8.17}$$

and that

$$\Sigma(\beta) = \beta(\varphi_{\text{int}}(\beta) - f_{\text{level}-1}(\beta)). \tag{8.18}$$

Considering formulas (8.11), (8.13) and (8.18), it is not hard to argue that $e^{n\Sigma(\beta)}$ is (to leading exponential order) the number of BP solutions with free energy $\varphi_{\text{int}}(\beta)$ contributing to the sum (8.11). The quantity $\Sigma(\beta)$ (a kind of entropy) is called the complexity. It is the growth rate of the number of extremal measures dominating the convex decomposition (8.8) .

We explain later on how to concretely compute $f_{\text{level}-1}(\beta)$, $\varphi_{\text{int}}(\beta)$ and $\Sigma(\beta)$. Let us immediately describe how $\Sigma(\beta)$ informs us about the convex decomposition of the Gibbs distribution. We refer the reader to Figure 1.6 for a pictorial representation. For a large class of problems one finds that $\Sigma(\beta) = 0$ for $\beta < \beta_d$, which signals that only one extremal measure contributes to the Gibbs distribution. At β_d the complexity jumps to a non-zero value and then decreases as a function of β till β_c after which it takes on negative values. In the range $\beta_d < \beta < \beta_c$ where $\Sigma(\beta) > 0$ an exponentially large (with respect to n) number of extremal measures with the same internal free energy $\varphi_{\text{int}}(\beta)$ contribute to the Gibbs distribution. Beyond β_c one finds a negative complexity: this is inconsistent with the fact that it is an entropy. In order to enforce this constraint correctly one is forced to take the Parisi parameter $0 < x < 1$ in (8.17). More precisely, one sets x to the largest possible value (less than 1) such that $\Sigma(\beta) = 0$. With this prescription[4] for the correct value of x when $\beta > \beta_c$, one computes the internal free energy and the free energy and the complexity from the x-dependent level-one model. The complexity is zero by construction which means that there exist at most a sublinear (believed to be finite) number of extremal measures contributing to the Gibbs distribution. This phenomenon is called condensation.

The nature of the thresholds β_d and β_c has been discussed in the introduction (see Sect.

---

[4]One can argue that the Parisi parameter is a kind of "Lagrange multiplier" that enforces the non-negativity of the complexity in the level-one model.
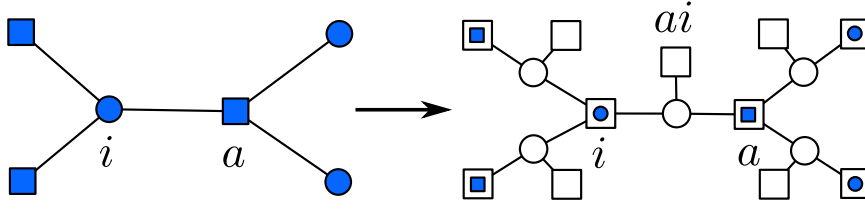
Figure 8.1:   On the left, an example of an original graph Γ. On the right its corresponding graph Γ_1 for the level-one model.

1.5) and we do not come back to this issue here.

We now show how the (x-dependent) level-one model is solved in practice. The main idea is to apply again the BP and Bethe equations for this model. The first step is to recognize that, if $\Gamma = (V, C, E)$ is the original factor graph, then the level-one model has the factor graph $\Gamma_1 = (V_1, C_1, E_1)$ described in Figure 8.1.

A variable node $i \in V$, becomes a function node $i \in C_1$, with the function

$$\psi_i^{(1)} = \prod_{a \in \partial i} \mathbb{I}\left(\nu_{i \to a} = g_{\mathrm{BP}}\right) e^{-x\beta F_i}. \tag{8.19}$$

A function node $a \in C$ remains a function node $a \in C_1$ with factor

$$\psi_a^{(1)} = \prod_{i \in \partial a} \mathbb{I}\left(\widehat{\nu}_{a \to i} = \widehat{g}_{\mathrm{BP}}\right) e^{-x\beta F_a}. \tag{8.20}$$

An edge $(a, i) \in E$, becomes a variable node $(a, i) \in V_1$. The dynamical variables are now couples of distributions $(\nu_{a \to i}, \widehat{\nu}_{a \to i})$. There is also an extra function node attached to each variable node of the new graph, or equivalently attached to each edge of the old graph. The corresponding function is

$$\psi_{ai}^{(1)} = e^{x\beta F_{ai}}. \tag{8.21}$$

With these definitions, Equ. (8.14) can be written as

$$\mu_{\mathrm{level}-1}(\underline{\nu}, \widehat{\underline{\nu}}; x) = \frac{1}{Z_{\mathrm{level}-1}(x)} \prod_{i \in V} \psi_i^{(1)} \prod_{a \in C} \psi_i^{(1)} \prod_{ai \in E} \psi_{ai}^{(1)}. \tag{8.22}$$

For the distributions $(\underline{\nu}, \widehat{\underline{\nu}})$ that satisfy the BP equations (8.3), some algebra leads to the useful formulas

$$\begin{cases} e^{-x\beta(F_a - F_{ai})} &= \widehat{z}_{a \to i}^x \\ e^{-x\beta(F_i - F_{ai})} &= z_{i \to a}^x \end{cases} \tag{8.23}$$
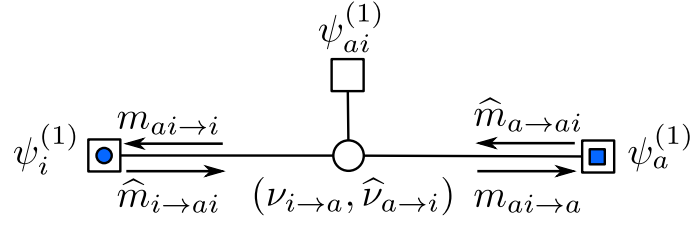
Figure 8.2: Messages are labeled by m if they are outgoing from a variable node in V_1 and by $\widehat{m}$ if they are outgoing from a function node in C_1.

where

$$
\begin{cases}
z_{i \to a} & = \sum_{\sigma_i} \prod_{b \in \partial i \backslash a} \widehat{\nu}_{b \to i} (\sigma_i) \\
\widehat{z}_{a \to i} & = \sum_{\sigma_{\partial a}} \psi_a (\sigma_{\partial a}) \prod_{\partial j \in a \backslash i} \nu_{j \to a} (\sigma_i)
\end{cases}
\tag{8.24}
$$

The BP equations for (8.22) involve four kind of messages as shown in Figure 8.2.

Messages from a (new) function node to a (new) variable node satisfy

$$
\begin{aligned}
\widehat{m}_{a \to ai} &\simeq \sum_{(\underline{\nu}, \underline{\widehat{\nu}}) \backslash (\nu_{i \to a}, \widehat{\nu}_{a \to i})} \psi_a^{(1)} \prod_{aj \in \partial a \backslash ai} m_{aj \to a} \\
&= e^{-x\beta F_{ai}} \sum_{\underline{\nu} \backslash \nu_{i \to a}} \mathbb{I} (\widehat{\nu}_{a \to i} = \widehat{g}_{\mathrm{BP}}) (\widehat{z}_{a \to i})^x \prod_{aj \in \partial a \backslash ai} m_{aj \to a}
\end{aligned}
\tag{8.25}
$$

and

$$
\begin{aligned}
\widehat{m}_{i \to ai} &\simeq \sum_{(\underline{\nu}, \underline{\widehat{\nu}}) \backslash (\nu_{i \to a}, \widehat{\nu}_{a \to i})} \psi_i^{(1)} \prod_{bi \in \partial i \backslash ai} m_{bi \to i} \\
&= e^{-x\beta F_{ai}} \sum_{\underline{\widehat{\nu}} \backslash \widehat{\nu}_{a \to i}} \mathbb{I} (\nu_{i \to a} = g_{\mathrm{BP}}) (z_{i \to a})^x \prod_{bi \in \partial i \backslash ai} m_{bi \to i}.
\end{aligned}
\tag{8.26}
$$

where the symbol $\simeq$ means equal up to a normalization factor. Messages from a (new) function node to a (new) variable node satisfy

$$
\begin{cases}
m_{ai \to i} \simeq e^{x\beta F_{ai}} \widehat{m}_{a \to ai} \\
m_{ai \to a} \simeq e^{x\beta F_{ai}} \widehat{m}_{i \to ai}.
\end{cases}
\tag{8.27}
$$

Notice that equations (8.25), (8.26) and (8.27) imply that $m_{ai \to a} (\nu_{i \to a}, \widehat{\nu}_{a \to i})$ (resp. $m_{ai \to i} (\nu_{i \to a}, \widehat{\nu}_{a \to i})$) is a constant function with respect to the argument $\widehat{\nu}_{a \to i}$ (resp. $\nu_{i \to a}$). This allows us to make a simplification by defining the following distributions

$$
\begin{cases}
Q_{i \to a} (\nu_{i \to a}) & = m_{ai \to a} (\nu_{i \to a}, \widehat{\nu}_{a \to i}) \\
\widehat{Q}_{a \to i} (\widehat{\nu}_{a \to i}) & = m_{ai \to i} (\nu_{i \to a}, \widehat{\nu}_{a \to i}).
\end{cases}
\tag{8.28}
$$

125

Distributions Q and $\widehat{Q}$ are called cavity messages, and live on the edges of the original factor graph $\Gamma = (V, C, E)$. From now on we can forget about the factor graph $\Gamma_1 = (V_1, C_1, E_1)$. The cavity messages satisfy

$$\widehat{Q}_{a \to i}(\widehat{\nu}_{a \to i}) \simeq \sum_{\underline{\nu}} \mathbb{I}(\widehat{\nu}_{a \to i} = \widehat{g}_{\mathrm{BP}}) \, \widehat{z}_{a \to i}^{x} \prod_{j \in \partial a \setminus i} Q_{j \to a}(\nu_{j \to a})$$

$$Q_{i \to a}(\nu_{i \to a}) \simeq \sum_{\underline{\widehat{\nu}}} \mathbb{I}(\nu_{i \to a} = g_{\mathrm{BP}}) \, z_{i \to a}^{x} \prod_{b \in \partial i \setminus a} \widehat{Q}_{b \to i}(\widehat{\nu}_{b \to i}). \tag{8.29}$$

The Bethe free energy functional of the level-one model can be expressed as a functional of the cavity messages (one way to determine this functional is to write down the functional whose critical points are given by Equ. (8.29)). This is an approximation for the true free energy (8.12) of the level-one model

$$f_{\mathrm{level}-1}^{\mathrm{Bethe}}(\underline{Q}, \underline{\widehat{Q}}; x) := \frac{1}{n} \left\{ \sum_{i \in V} \mathcal{F}_i + \sum_{a \in C} \mathcal{F}_a - \sum_{(i,a) \in E} \mathcal{F}_{ai} \right\}, \tag{8.30}$$

where

$$\mathcal{F}_i \left(\left\{ \widehat{Q}_{b \to i} \right\}_{b \in \partial i} \right) = -\frac{1}{x\beta} \ln \left(\sum_{\underline{\widehat{\nu}}} e^{-x\beta F_i} \prod_{b \in \partial i} \widehat{Q}_{b \to i} \right)$$

$$\mathcal{F}_a \left(\{ Q_{j \to a} \}_{j \in \partial a} \right) = -\frac{1}{x\beta} \ln \left(\sum_{\underline{\nu}} e^{-x\beta F_a} \prod_{j \in \partial a} Q_{j \to a} \right)$$

$$\mathcal{F}_{ai} \left(Q_{i \to a}, \widehat{Q}_{a \to i} \right) = -\frac{1}{x\beta} \ln \left(\sum_{\underline{\nu}, \underline{\widehat{\nu}}} e^{-x\beta F_{ai}} Q_{i \to a} \widehat{Q}_{a \to i} \right). \tag{8.31}$$

In principle one has to solve the cavity equations (8.29) for $0 < x \leq 1$, and compute the x-dependent free energy $f_{\mathrm{level}-1}^{\mathrm{Bethe}}$. From this free energy we obtain the complexity by computing the derivative in equation (8.17). This allows us to determine the thresholds β_d and β_c. For $\beta < \beta_c$ the free energy is given by $f_{\mathrm{level}-1}^{\mathrm{Bethe}}|_{x=1}$. This function has no singularities, which means that there are no static (thermodynamic) phase transitions for $\beta < \beta_c$ and in particular at $\beta = \beta_d$. In this phase one has $\Sigma(\beta; x = 1) = 0$ for $\beta < \beta_d \leq \beta_c$ and $\Sigma(\beta; x = 1) > 0$ for $\beta_d \leq \beta < \beta_c$. For $\beta > \beta_c$ one enforces a zero complexity by setting the Parisi parameter to a value $0 < x_* < 1$ s.t. $\Sigma(\beta; x_*) = 0$. The free energy is not analytic at β_c, due to the change of x parameter. This a static phase transition threshold.

In practice, as long as we are interested only in the range $\beta < \beta_c$ we can set $x = 1$. It is then possible to simplify the cavity equations (8.29) and the level-1 free energy (8.30).

8.2 Application to Spatially-Coupled LDGM codes

We apply the formalism of the cavity method to the Gibbs measure $\mu_\beta(\underline{u}|\underline{x})$ (7.8). Instead of working with the alphabet $\{0, 1\}$, we find it convenient to use the mapping $\sigma_i = (-1)^{u_i}$ and $J_a = (-1)^{x_a}$ to the alphabet $\{-1, +1\}$. With this change of variables, the Gibbs measure (7.8) takes the form

$$\mu_\beta(\underline{u}|\underline{x}) = \frac{1}{Z_\beta} \prod_{a \in C} \psi_a\left(\underline{\sigma}_{\partial a}\right) \tag{8.32}$$

with

$$\psi_a(\{\sigma_i, i \in \partial a\}) = e^{-\beta(1 - J_a \prod_{i \in \partial a} \sigma_i)}. \tag{8.33}$$

The probability distributions $\nu_{i \to a}(\sigma_i)$ and $\widehat{\nu}_{a \to i}(\sigma_i)$ are entirely characterized by their means, $\tanh \beta\eta_{i \to a}$ and $\tanh \beta\widehat{\eta}_{a \to i}$, as follows (we drop the subscripts)

$$\nu(\sigma) = \frac{1 + \sigma \tanh \beta\eta}{2}. \tag{8.34}$$

With this parametrization, the BP equations (8.3) for the model (8.32) become

$$\begin{cases} \widehat{\eta}_{a \to i} &= \widehat{g}_{\mathrm{BP}}(\{\eta_{j \to a}\}_{j \in \partial a \setminus i} \mid J_a) \\ \eta_{i \to a} &= g_{\mathrm{BP}}(\{\widehat{\eta}_{b \to i}\}_{b \in \partial i \setminus a}), \end{cases} \tag{8.35}$$

where

$$\begin{cases} g_{\mathrm{BP}}(\{\widehat{\eta}_{b \to i}\}_{b \in \partial i \setminus a}) = \sum_{b \in \partial i \setminus a} \widehat{\eta}_{bi}. \\ \widehat{g}_{\mathrm{BP}}(\{\eta_{j \to a}\}_{j \in \partial a \setminus i} \mid J_a) = \frac{J_a}{\beta} \operatorname{atanh}\left(\tanh \beta \prod_{j \in \partial a \setminus i} \tanh \beta\eta_{j \to a}\right) \end{cases} . \tag{8.36}$$

The Bethe free energy per check nodes (8.6) reads

$$\begin{aligned} f^{\mathrm{Bethe}}\left(\underline{\eta}, \underline{\widehat{\eta}}\right) = &- \frac{1}{\beta}\left(\ln\left(1 + e^{-2\beta}\right) + (R - 1)\ln 2\right) \\ &- \frac{1}{\beta N} \sum_{a \in C} (1 - |\partial a|) \ln Z_1(\{\eta_{j \to a}\}_{j \in \partial a} \mid J_a) \\ &- \frac{R}{\beta M} \sum_{i \in V} \ln Z_2(\{\widehat{\eta}_{b \to i}\}_{b \in \partial i}), \end{aligned} \tag{8.37}$$

where

$$\begin{cases} Z_1(\{\eta_{j \to a}\}_{j \in \partial a} \mid J_a) = 1 + J_a \tanh \beta \prod_{i \in \partial a} \tanh \beta\eta_{i \to a} \\ Z_2(\{\widehat{\eta}_{a \to i}\}_{a \in \partial i}) = \frac{1}{2} \sum_{s \in \{-1, 1\}} \prod_{a \in \partial i}(1 + s \tanh \beta\widehat{\eta}_{a \to i}). \end{cases} \tag{8.38}$$

Since we have parametrized the BP messages by real numbers, the cavity messages $Q_{i\to a}$, $\widehat{Q}_{a\to i}$ become distributions on $\eta_{i\to a}$, $\widehat{\eta}_{a\to i}$. The cavity equations (8.29) reduce to

$$
Q_{i\to a}(\eta_{i\to a}) \simeq \int \prod_{b\in\partial i\backslash a} d\widehat{\eta}_{b\to i}\widehat{Q}_{b\to i}(\widehat{\eta}_{b\to i})
$$

$$
\times Z_2^x(\{\widehat{\eta}_{b\to i}\}_{b\in\partial i\backslash a})\delta\left(\eta_{i\to a} - g_{\mathrm{BP}}(\{\widehat{\eta}_{b\to i}\}_{b\in\partial i\backslash a})\right) \qquad (8.39)
$$

and

$$
\widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i}) \simeq \int \prod_{j\in\partial a\backslash i} d\eta_{j\to a}Q_{j\to a}(\eta_{j\to a})
$$

$$
\times Z_1^x(\{\eta_{j\to a}\}_{b\in\partial i\backslash a})\delta\left(\widehat{\eta}_{a\to i} - \widehat{g}_{\mathrm{BP}}(\{\eta_{j\to a}\}_{j\in\partial a\backslash i} \mid J_a)\right). \qquad (8.40)
$$

For the Bethe free energy of the level-one model one finds

$$
\begin{aligned}
f_{\mathrm{level}-1}^{\mathrm{Bethe}}\left(\underline{\eta}, \underline{\widehat{\eta}}; x\right) &= -\frac{1}{\beta}\left(\ln(1 + e^{-2\beta}) + (R-1)\ln 2\right) \\
&\quad - \frac{1}{\beta x N}\sum_{a\in C}(1-|\partial a|)\ln\Big\{\int\prod_{i\in\partial a} d\eta_{i\to a}Q_{i\to a}(\eta_{i\to a}) \\
&\qquad \times Z_1^x(\{\eta_{i\to a}\}_{i\in\partial a} \mid J_a)\Big\} \\
&\quad - \frac{R}{\beta x M}\sum_{i\in V}\ln\Big\{\int\prod_{a\in\partial i} d\widehat{\eta}_{a\to i}\widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i}) \\
&\qquad \times Z_2^x(\{\widehat{\eta}_{a\to i}\}_{a\in\partial i})\Big\}.
\end{aligned} \qquad (8.41)
$$

We are interested in the range $\beta < \beta_c$ for which the Parisi parameter is set to $x = 1$. In this case the above equations greatly simplify. We first define average cavity messages

$$
\begin{cases} h_{i\to a} &= \mathrm{Av}[Q_{i\to a}] \\ \widehat{h}_{a\to i} &= \mathrm{Av}[\widehat{Q}_{a\to i}], \end{cases} \qquad (8.42)
$$

where the functional $\mathrm{Av}[P]$ is

$$
\mathrm{Av}[P] = \frac{1}{\beta}\mathrm{atanh}\Big\{\int d\eta P(\eta)\tanh\beta\eta\Big\}. \qquad (8.43)
$$

Thus $\tanh\beta h_{i\to a}$ and $\tanh\beta\widehat{h}_{a\to i}$ are real valued messages and are averages of $\tanh\beta\eta_{i\to a}$ and $\tanh\beta\widehat{\eta}_{a\to i}$ with respect to the cavity distributions $Q_{i\to a}(\eta_{i\to a})$ and $\widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i})$ respectively. The free energy of the level-one model for $x = 1$ can be expressed in terms

of these real valued messages, and one finds

$$f_{\text{level}-1}^{\text{Bethe}}\left(\underline{h}, \underline{\widehat{h}}\right) = -\frac{1}{\beta}\left(\ln(1 + e^{-2\beta}) + (R-1)\ln 2)\right)$$
$$-\frac{1}{\beta N}\sum_{a \in C}(1 - |\partial a|)\ln Z_1(\{h_{j \to a}\}_{j \in \partial a} \mid J_a)$$
$$-\frac{R}{\beta M}\sum_{i \in V}\ln Z_2(\{\widehat{h}_{b \to i}\}_{b \in \partial i}). \tag{8.44}$$

Remarkably, the level-one Bethe free energy (8.44) is equal to the original Bethe free energy functional $f^{\text{Bethe}}\left(\underline{\eta}, \underline{\widehat{\eta}}\right)$ defined in (8.37), but now evaluated for the average fields $h_{i \to a}$ and $\widehat{h}_{a \to i}$. From the cavity equations (8.39)-(8.40) for $x = 1$, one can deduce after some algebra that the average fields $h_{i \to a}$ and $\widehat{h}_{a \to i}$ satisfy

$$\begin{cases}\widehat{h}_{a \to i} &= \widehat{g}_{\text{BP}}(\{h_{j \to a}\}_{j \in \partial a \backslash i} \mid J_a) \\ h_{i \to a} &= g_{\text{BP}}(\{\widehat{h}_{b \to i}\}_{b \in \partial i \backslash a}).\end{cases} \tag{8.45}$$

Remarkably, the average fields $h_{i \to a}$ and $\widehat{h}_{a \to i}$ satisfy the BP equations (8.35).

To summarize, when $x = 1$, $f_{\text{level}-1}^{\text{Bethe}}$ equals f^{Bethe} computed at a certain appropriate BP fixed-point. This fixed point corresponds to messages $\tanh\beta h_{i \to a}$, $\tanh\beta\widehat{h}_{a \to i}$ which are an average of the BP solutions $\tanh\beta\eta_{i \to a}$, $\tanh\beta\widehat{\eta}_{a \to i}$ over the cavity distributions $Q_{i \to a}(\eta_{i \to a})$ and $\widehat{Q}_{a \to i}(\widehat{\eta}_{a \to i})$. The messages $\tanh\beta\eta_{i \to a}$, $\tanh\beta\widehat{\eta}_{a \to i}$ describe the "extremal states" whereas the messages $\tanh\beta h_{i \to a}$, $\tanh\beta\widehat{h}_{a \to i}$ describe their convex superposition.

8.3 Quenched Average and Density Evolution Equations

The discussion in Sections 8.1 and 8.2 is valid for a single instance. Unfortunately the cavity equations (8.39) and (8.40) are complicated to implement in practice as they involve distributions over real values. Furthermore, there is no known technique to solve the averaged equations (8.45) as it requires to select a particular BP fixed-point among an exponentially large set of solutions.

Instead of computing the free energy, internal free energy and complexity, we compute their quenched average

$$\overline{f} = \lim_{N \to \infty}\mathbb{E}_{\Gamma, \underline{J}}\left[f_{\text{level}-1}^{\text{Bethe}}\right], \quad \overline{\varphi}_{\text{int}} = \lim_{N \to \infty}\mathbb{E}_{\Gamma, \underline{J}}\left[\varphi_{\text{int}}\right], \quad \overline{\Sigma} = \lim_{N \to \infty}\mathbb{E}_{\Gamma, \underline{J}}\left[\Sigma\right]. \tag{8.46}$$

The quenched average is performed over the graph ensemble and the Bernoulli source. It is expected that the free energy, internal free energy and complexity concentrate on their quenched average.

To perform the quenched average we assume that the cavity messages $Q_{i \to a}(\eta_{i \to a})$ and $\widehat{Q}_{a \to i}(\widehat{\eta}_{a \to i})$ can be considered as i.i.d. realizations of random variables $Q_z(\eta)$ and $\widehat{Q}_z(\widehat{\eta})$. The random variables depend only on the *position* z along the spatial dimension and not on the *direction* of the edges $i \to a$ and $a \to i$. This key assumption is partially justified by the following reason: If one selects an edge (i, a) randomly, its neighborhood will be a tree with high probability. Therefore in the limit $N \to \infty$, the messages $Q_{i \to a}(\eta_{i \to a})$ and $\widehat{Q}_{a \to i}(\widehat{\eta}_{a \to i})$ will converge in distribution to the random variables $Q_z(\eta)$ and $\widehat{Q}_z(\widehat{\eta})$.

The distributions of the random variables $Q_z(\eta)$ and $\widehat{Q}_z(\widehat{\eta})$ are denoted $\mathcal{Q}_z$ and $\widehat{\mathcal{Q}}_z$. Note that the cavity messages are already distributions over real numbers, so that $\mathcal{Q}_z$ and $\widehat{\mathcal{Q}}_z$ are distributions of distributions. From the cavity equations (8.39), (8.40) it is easy to formally write down the set of integral equations that these distributions of distributions satisfy. These are the density evolution equations for the cavity message.

8.3.1   Fixed-Point Equations of the Cavity Method for $\beta \leq \beta_c$

Using the method of [16], we can simplify considerably the equation for the density evolution for $\beta < \beta_c$. The quenched free energy, internal free energy and complexity can be expressed with only six distribution over real values $q_z(h)$, $\widehat{q}_z(\widehat{h})$, $q_z^{\sigma = \pm 1}(\eta|h)$ and $\widehat{q}_z^{\sigma = \pm 1}(\widehat{\eta}|\widehat{h})$. Two of those densities are the probability distribution for the average fields $h_{i \to a}$ and $h_{a \to i}$,

$$
\begin{cases}
q_z(h) & = \int \mathcal{D}\mathcal{Q}_z[Q]\delta(h - Av[Q]) \\
\widehat{q}_z(\widehat{h}) & = \int \mathcal{D}\widehat{\mathcal{Q}}_z[\widehat{Q}]\delta(\widehat{h} - Av[\widehat{Q}]).
\end{cases}
\tag{8.47}
$$

The four remaining densities are transforms, for $\sigma = \pm 1$, of the conditional distributions $q_z(\eta|h)$ and $\widehat{q}_z(\widehat{\eta}|\widehat{h})$,

$$
\begin{cases}
q_z^\sigma(\eta|h) & = \frac{1 + \sigma \tanh \beta \eta}{1 + \sigma \tanh \beta h} q_z(\eta|h) \\
\widehat{q}_z^\sigma(\widehat{\eta}|\widehat{h}) & = \frac{1 + \sigma \tanh \beta \widehat{\eta}}{1 + \sigma \tanh \beta \widehat{h}} q_z(\widehat{\eta}|\widehat{h}),
\end{cases}
\tag{8.48}
$$

where the conditional distributions $q_z(\eta|h)$ and $\widehat{q}_z(\widehat{\eta}|\widehat{h})$ are defined with the respect to $q_z(h)$ and $\widehat{q}_z(\widehat{h})$,

$$
\begin{cases}
q_z(\eta|h)q_z(h) & = \int \mathcal{D}\mathcal{Q}_z[Q]Q(\eta)\delta(h - Av[Q]) \\
\widehat{q}_z(\widehat{\eta}|\widehat{h})\widehat{q}_z(\widehat{h}) & = \int \mathcal{D}\widehat{\mathcal{Q}}_z[\widehat{Q}]\widehat{Q}(\widehat{\eta})\delta(\widehat{h} - Av[\widehat{Q}]).
\end{cases}
\tag{8.49}
$$

It is convenient to define two functions g and $\widehat{g}$ which correspond to (8.36) in the case of

density evolution

$$
\begin{cases}
g(\widehat{h}_1, \ldots, \widehat{h}_{r-1}) = \sum_{i=1}^{r-1} \widehat{h}_i \\
\widehat{g}\,(h_1, \ldots, h_{p-1} \mid J) = J\beta^{-1} \tanh^{-1}\!\big(\tanh \beta \prod_{i=1}^{p-1} \tanh \beta h_i\big) \quad,
\end{cases}
\tag{8.50}
$$

where $J \equiv (-1)^x$ is the random variable representing the source bits. Furthermore we set $P(r) = \frac{(p/R)^r}{r!} e^{-p/R}$ for the Poisson-degree distribution of code-bit nodes.

With the independence assumption on the cavity messages, relations (8.45) imply that distributions $q_z\,(h)$, $\widehat{q}_z\left(\widehat{h}\right)$ satisfy a set of closed equations[5]

$$
q_z\,(h) = \sum_{r=0}^{\infty} \frac{P(r)}{w^r} \sum_{y_1,\ldots y_r=0}^{w-1} \int \prod_{a=1}^{r} d\widehat{h}_a \widehat{q}_{z-y_a}(\widehat{h}_a)\delta(h - g(\widehat{h}_1, \ldots, \widehat{h}_r))
\tag{8.51}
$$

and

$$
\widehat{q}_z(\widehat{h}) = \frac{1}{w^{p-1}} \sum_{y_1,\ldots,y_{p-1}=0}^{w-1} \int \prod_{i=1}^{p-1} dh_i q_{z+y_i}\,(h_i) \frac{1}{2} \sum_{J=\pm 1} \delta(\widehat{h} - \widehat{g}(h_1, \ldots, h_{p-1} \mid J)).
\tag{8.52}
$$

Let $\sigma_i = \pm 1$ denote auxiliary "spin" variables. We introduce the conditional measure over $\sigma_1, \ldots, \sigma_{l-1}$,

$$
\nu_1(\sigma_1, \ldots, \sigma_{l-1} \mid J, h_1, \ldots, h_{p-1}) = \frac{1 + J \tanh \beta \prod_{i=1}^{p-1} \sigma_j}{1 + J \tanh \beta \prod_{i=1}^{p-1} \tanh \beta h_i} \prod_{i=1}^{p-1} \frac{1 + \sigma_i \tanh \beta h_i}{2}.
\tag{8.53}
$$

The equations for distributions $q_z^{\sigma}(\eta \mid h)$ and $\widehat{q}_z^{\sigma}(\widehat{\eta} \mid \widehat{h})$ are

$$
q_z^{\sigma}(\eta \mid h) q_z(h) = \sum_{r=0}^{\infty} \frac{P(r)}{w^r} \sum_{y_1,\ldots,y_r=0}^{w-1} \int \prod_{a=1}^{r} d\widehat{h}_a d\widehat{\eta}_a \widehat{q}_{z-y_a}^{\sigma}(\widehat{\eta}_a \mid \widehat{h}_a) \widehat{q}_{z-y_a}(\widehat{h})
$$
$$
\times \, \delta(\eta - g(\widehat{\eta}_1, \ldots, \widehat{\eta}_r))\delta(h - g(\widehat{h}_1, \ldots, \widehat{h}_r))
\tag{8.54}
$$

---

[5]We use the convention that if z is out of range the corresponding distribution is a unit mass at zero.

and

$$\widehat{q}_z^{\sigma}(\widehat{\eta}|\widehat{h})\widehat{q}_z(\widehat{h}) = \frac{1}{w^{p-1}} \sum_{y_1,\ldots,y_{p-1}=0}^{w-1} \int \prod_{i=1}^{p-1} dh_i q_{z+y_i}(h_i)$$

$$\times \frac{1}{2} \sum_{J=\pm 1} \sum_{\sigma_1,\ldots,\sigma_{p-1}=\pm 1} \nu_1(\sigma_1,\ldots,\sigma_{p-1}|J\sigma,h_1,\ldots,h_{p-1})$$

$$\times \delta(\widehat{h} - \widehat{g}(h_1,\ldots,h_{p-1} \mid J))$$

$$\times \int \prod_{i=1}^{p-1} d\eta_i q_{z+y_i}^{\sigma_i}(\eta_i|h_i)\delta(\widehat{\eta} - \widehat{g}(\eta_1,\ldots,\eta_{p-1} \mid J)). \tag{8.55}$$

Equations (8.51), (8.52), (8.54), (8.55) constitutes a closed set of fixed-point equations for six probability distributions.

8.3.2   Complexity in Terms of Fixed-Point Densities

Let

$$\begin{cases} Z_1(h_1,\ldots,h_p \mid J) = 1 + J \tanh \beta \prod_{i=1}^{p} \tanh \beta h_i \\ Z_2(\widehat{h}_1,\ldots,\widehat{h}_r) = \frac{1}{2} \sum_{\sigma=\pm 1} \prod_{i=1}^{r}(1 + \sigma \tanh \beta \widehat{h}_i). \end{cases} \tag{8.56}$$

Weights Z_1 and Z_2 corresponds to (8.38) in density evolution. We are now ready to give the expression for the complexity in terms of the densities $q_z(h)$, $\widehat{q}_z(\widehat{h})$, $q_z^{\sigma}(\eta|h)$ and $\widehat{q}_z^{\sigma}(\widehat{\eta}|\widehat{h})$.

The expression of $\overline{f}$ is the simplest

$$-\beta\overline{f} = \ln(1 + e^{-2\beta}) + (R-1)\ln 2$$

$$- \frac{l-1}{L} \sum_{z=1}^{L} \frac{1}{w^p} \sum_{y_1,\ldots,y_p=0}^{w-1} \int \prod_{i=1}^{p} dh_i q_{z+y_i}(h_i)$$

$$\times \frac{1}{2} \sum_{J=\pm 1} \ln Z_1(h_1,\ldots,h_p \mid J)$$

$$+ \frac{R}{L+w-1} \sum_{z=1}^{L+w-1} \sum_{r=0}^{\infty} \frac{P(r)}{w^r}$$

$$\times \sum_{y_1,\ldots,y_r=0}^{w-1} \int \prod_{a=1}^{r} d\widehat{h}_a \widehat{q}_{z-y_a}(\widehat{h}_a) \ln Z_2(\widehat{h}_1,\ldots,\widehat{h}_r). \tag{8.57}$$

Recall formula (8.18) which expresses the complexity as

$$\Sigma(\beta) = \beta(\varphi_{\text{int}}(\beta) - f_{\text{level}-1}(\beta)). \tag{8.58}$$

Since we already know that $\overline{f}$ is given by (8.57), it remains to compute the internal free energy in the Bethe approximation. For this purpose we use

$$\varphi_{\text{int}}^{\text{Bethe}}(\beta) = \frac{\partial}{\partial x}(x f_{\text{level}-1}^{\text{Bethe}}(\beta; x))|_{x=1}. \tag{8.59}$$

We compute the x-derivative on (8.41), and average over the cavity distributions, the graph ensemble and the Bernoulli source. To express $\overline{\varphi}_{\text{int}}$ we first need to define the conditional measure over $\sigma = \pm 1$

$$\nu_2(\sigma|\widehat{h}_1, \ldots, \widehat{h}_k) = \frac{\prod_{a=1}^{k}(1 + \sigma \tanh \beta \widehat{h}_a)}{\prod_{a=1}^{k}(1 + \tanh \beta \widehat{h}_a) + \prod_{a=1}^{k}(1 - \tanh \beta \widehat{h}_a)}. \tag{8.60}$$

After some algebra one finds that $\overline{\varphi}_{\text{int}}(\beta)$ is given by

$$
\begin{aligned}
-\beta \overline{\varphi}_{\text{int}} =& \ln(1 + e^{-2\beta}) + (R - 1) \ln 2 \\
& - \frac{p-1}{L} \sum_{z=1}^{L} \frac{1}{w^p} \sum_{y_1, \ldots, y_p=0}^{w-1} \int \prod_{i=1}^{p} dh_i q_{z+y_i}(h_i) \\
& \times \frac{1}{2} \sum_{J=\pm 1} \sum_{\sigma_1, \ldots, \sigma_p=\pm 1} \nu_1(\sigma_1, \ldots, \sigma_p | J, h_1, \ldots, h_p) \\
& \times \int \prod_{i=1}^{p} d\eta_i q_{z+y_i}^{\sigma_i}(\eta_i | h_i) \ln Z_1(\eta_1, \ldots, \eta_p \mid J) \\
& + \frac{R}{L + w - 1} \sum_{z=1}^{L+w-1} \sum_{r=0}^{\infty} \frac{P(r)}{w^r} \\
& \times \sum_{y_1, \ldots, y_r=0}^{w-1} \int \prod_{a=1}^{r} d\widehat{h}_a \widehat{q}_{z-y_a}(\widehat{h}_a) \sum_{\sigma} \nu_2(\sigma|\widehat{h}_1, \ldots, \widehat{h}_r) \\
& \times \int \prod_{a=1}^{r} d\widehat{\eta}_a \widehat{q}_{z-y_a}^{\sigma}(\widehat{\eta}_a | \widehat{h}_a) \ln Z_2(\widehat{\eta}_1, \ldots, \widehat{\eta}_r).
\end{aligned} \tag{8.61}
$$

Thanks to (8.57), (8.61) the complexity $\overline{\Sigma}(\beta; L, w)$ of the coupled ensemble is computed, one reads off the dynamical and condensation thresholds $\beta_d(L, w)$ and $\beta_c(L, w)$. The corresponding quantities for the underlying ensemble are obtained by setting $L = w = 1$.

8.3.3 Population Dynamics

The fixed-point equations (8.51), (8.52), (8.54) and (8.55) cannot in general be solved in closed form. To solve those equations numerically we use a method called population dynamics.

The idea is to represent the densities $q_z(h)$, $q_z^{\pm}(\eta|h)$, $\widehat{q}_z(\widehat{h})$, and $\widehat{q}_z^{\pm}(\widehat{\eta}|\widehat{h})$ through a sample called a population. We have two populations: a code-bit population and a check

---

Algorithm 8.1 Population Dynamics for (8.51) and (8.52)

---

for $z = 1$ *to* $L + w - 1$ **do**
$\quad$ **for** $i = 1$ *to* n **do**
$\quad\quad$ Draw $\widehat{h}_{(z,i)}$ uniformly from $[-1, +1]$;

for $t \in \{1, \ldots, t_{max}\}$ **do**
$\quad$ **for** $z = 1$ *to* $L + w - 1$ **do**
$\quad\quad$ **for** $i = 1$ *to* n **do**
$\quad\quad\quad$ *Generate a new* $h_{(z,i)}$;
$\quad\quad\quad$ Choose $p - 1$ pair indices $a_1, \ldots, a_{p-1}$ uniformly from nw pairs (y, j),
$\quad\quad\quad$ $y \in [z - w + 1, z]$ and $j \in \{1, \ldots, n\}$;
$\quad\quad\quad$ **if** *for some index* k, $a_k = (y, j)$ *and* $y < 1$ **then**
$\quad\quad\quad\quad$ Set $\widehat{h}_{a_k} = 0$;
$\quad\quad\quad$ Set $h_{(z,i)} = \sum_{k=1}^{p-1} \widehat{h}_{a_k}$;

$\quad$ **for** $z = 1$ *to* L **do**
$\quad\quad$ **for** $a = 1$ *to* n **do**
$\quad\quad\quad$ *Generate J randomly and generate a new* $\widehat{h}_{(z,a)}$;
$\quad\quad\quad$ Choose $r - 1$ indices $i_1, \ldots, i_{r-1}$ uniformly from nw pairs (y, j),
$\quad\quad\quad$ $y \in [z, z + w - 1]$ and $j \in \{1, \ldots, n\}$;
$\quad\quad\quad$ Compute $\widehat{h}_{(z,a)}$ according to (8.52);

---

population. The code-bit population is constituted of $L + w - 1$ sets labeled by $z \in [1, L + w - 1]$. Each set, say z, has a population of size n, constituted of triples: $(h_{(z,i)}, \eta^+_{(z,i)}, \eta^-_{(z,i)})$, $1 \leq i \leq n$. The total size of the code-bit population is $(L + w - 1)n$. Similarly, we have a population of triples with size Ln for check nodes, i.e., $(\widehat{h}_{(z,a)}, \widehat{\eta}^+_{(z,a)}, \widehat{\eta}^-_{(z,a)})$, $z = 1, \ldots, L$, $a = 1, \ldots, n$. As inputs, they require the population size n, the maximum number of iterations $t_{\max}$, and the specifications of the SCLDGM (p, R, L, w) ensemble. First we solve the two equations (8.51) and (8.52) with Algorithm 8.1.

First we solve the two equations (8.51) and (8.52) with Algorithm 8.1. Then we solve (8.54) and (8.55) with the Algorithm[6] 8.2. From the final populations obtained after $t_{\max}$ iterations it is easy to compute the complexity and the thresholds β_d, β_c.

8.3.4   Further Simplifications of Fixed-Point Equations and Complexity

It is immediate to check that $q_z(h) = \delta(h)$ and $\widehat{q}_z(\widehat{h}) = \delta(\widehat{h})$ is a trivial fixed point of (8.51), (8.52). When we solve these equations by population dynamics with a uniform

---

[6]In the next to last line marked (*) the chosen index is not in a valid range. In an instance of a coupled ensemble, this happens at the boundary, in which the corresponding node has smaller degree. In the message-passing equation we discard these indices or equivalently assume that their triples are $(0, 0, 0)$.

Algorithm 8.2 Population Dynamics for (8.54) and (8.55)

---

for $z = 1$ *to* L **do**

$\quad$ **for** $i = 1$ *to* n **do**

$\quad\quad$ Set $\eta^{\pm}_{(z,i)} = \pm\infty$ and draw $h_{(z,i)}$ from $q_z(h)$;

for $t \in \{1, \ldots, t_{max}\}$ **do**

$\quad$ **for** $z = 1$ *to* L **do**

$\quad\quad$ **for** $a = 1$ *to* n **do**

$\quad\quad\quad$ *Generate J randomly and generate a new triple $(\widehat{h}_{(z,a)}, \widehat{\eta}^{+}_{(z,a)}, \widehat{\eta}^{-}_{(z,a)})$:*

$\quad\quad\quad$ Choose $r - 1$ indices $i_1, \ldots, i_{r-1}$ uniformly from nw pairs (y, j),

$\quad\quad\quad$ $y \in [z, z + w - 1]$ and $j \in \{1, \ldots, n\}$;

$\quad\quad\quad$ Compute $\widehat{h}_{(z,a)}$ according to (8.52);

$\quad\quad\quad$ Generate a configuration $\sigma_1, \ldots, \sigma_{r-1}$ from $\nu_1(\ldots | + J, h_{i_1}, \ldots, h_{i_{r-1}})$ in (8.53);

$\quad\quad\quad$ Compute $\widehat{\eta}^{+}_{(z,a)}$ by plugging $\eta^{\sigma_1}_{i_1}, \ldots, \eta^{\sigma_{r-1}}_{i_{r-1}}$ in (8.55);

$\quad\quad\quad$ Generate a configuration $\sigma_1, \ldots, \sigma_{r-1}$ from $\nu_1(\ldots | - J, h_{i_1}, \ldots, h_{i_{r-1}})$ in (8.53);

$\quad\quad\quad$ Compute $\widehat{\eta}^{-}_{(z,a)}$ by plugging $\eta^{\sigma_1}_{i_1}, \ldots, \eta^{\sigma_{r-1}}_{i_{r-1}}$ in (8.55);

$\quad$ **for** $z = 1$ *to* $L + w - 1$ **do**

$\quad\quad$ **for** $i = 1$ *to* n **do**

$\quad\quad\quad$ *Generate a new triple $(h_{(z,i)}, \eta^{+}_{(z,i)}, \eta^{-}_{(z,i)})$:* Choose $p - 1$ pair indices

$\quad\quad\quad$ $a_1, \ldots, a_{p-1}$ uniformly from nw pairs (y, j), $y \in [z - w + 1, z]$ and

$\quad\quad\quad$ $j \in \{1, \ldots, n\}$; **if** *for some index k, $a_k = (y, j)$ and $y < 1$* **then**

$\quad\quad\quad\quad$ Set $(\widehat{h}_{a_k}, \widehat{\eta}^{+}_{a_k}, \widehat{\eta}^{-}_{a_k}) = (0, 0, 0)$;(*)

$\quad\quad\quad$ Set $h_{(z,i)} = \sum_{k=1}^{p-1} \widehat{h}_{a_k}$ and $\eta^{\pm}_{(z,i)} = \sum_{k=1}^{p-1} \widehat{\eta}^{\pm}_{a_k}$;

---

initial condition over $[-1, +1]$ for $\widehat{h}$, we find that for fixed degrees and β fixed in a finite range depending on the degrees, the updates converge towards the trivial fixed point. Up to numerical precision, the values of h, $\widehat{h}$ are concentrated on 0. It turns out that the range of β for which this is valid is wider than the interval $[0, \beta_c]$. At first sight this could seem paradoxical, and one would have expected that this range of β is equal to $[0, \beta_c]$. In fact, one must recall that beyond β_c the equations of paragraph 8.3.1 are not valid (see Section 8.1), so there is *no* paradox. Theorem C.1 in Appendix C shows that, for a wide class of initial conditions and given β, for large enough degree p the iterative solution of (8.51), (8.52) tends to the trivial point. This theorem, together with the numerical evidence, provides strong support for the exactness of the following simplification.

We assume that for $\beta < \beta_c$, equations (8.51), (8.52) have a unique solution $q_z(h) = \delta(h)$ and $\widehat{q}_z(\widehat{h}) = \delta(\widehat{h})$. For $h = \hat{h} = 0$ the distributions in (8.54), (8.55) possess a symmetry $q_z^{\sigma=1}(\eta|0) = q_z^{\sigma=-1}(-\eta|0)$, $\widehat{q}_z^{\sigma=1}(\widehat{\eta}|0) = \widehat{q}_z^{\sigma=-1}(-\widehat{\eta}|0)$. It is therefore natural to look for symmetrical solutions, and set

$$q_z^+(\eta) = q_z^{\sigma=+1}(\eta|0), \qquad \text{and} \qquad \widehat{q}_z^+(\widehat{\eta}) = \widehat{q}_z^{\sigma=+1}(\widehat{\eta}|0). \tag{8.62}$$

Finally, equations (8.54), (8.55) simplify drastically,

$$q_z^+(\eta) = \sum_{r=0}^{\infty} \frac{P(r)}{w^r} \sum_{y_1,\ldots,y_r=0}^{w-1} \int \prod_{a=1}^{r} d\widehat{\eta}_a \widehat{q}_{z-y_a}^+(\widehat{\eta}_a)\delta(\eta - g(\widehat{\eta}_1, \ldots, \widehat{\eta}_r)), \tag{8.63}$$

and

$$\widehat{q}_z^+(\widehat{\eta}) = \frac{1}{w^{p-1}} \sum_{y_1,\ldots,y_{p-1}=0}^{w-1} \int \prod_{i=1}^{p-1} d\eta_i q_{z+y_i}^+(\eta_i)$$
$$\times \sum_{J=\pm 1} \frac{1 + J \tanh\beta}{2} \delta(\widehat{\eta} - \widehat{g}(\eta_1, \ldots, \eta_{p-1} \mid J)). \tag{8.64}$$

Remarkably, these are the standard density evolution equations for an LDGM code over a BSC test-channel with an amplitude of half log-likelihood ratios equal to β.

The quenched free energy (8.57) now takes a very simple form

$$-\beta\overline{f} = \ln(1 + e^{-2\beta}) + (R - 1)\ln 2. \tag{8.65}$$

At this point let us note that this simple formula has been proven by the *interpolation method* [110], for small enough β. Since it is expected that there is no (static) thermodynamic phase transition for $\beta < \beta_c$, the free energy is expected to be analytic for $\beta < \beta_c$. Thus by analytic continuation, formula (8.65) should hold for all $\beta < \beta_c$. This also provides a posteriori support for the triviality assumption made above for the fixed point. Indeed, a non-trivial fixed-point leading to the same free energy would entail miraculous cancellations.

When we compute the complexity, expression (8.65) cancels with the first line in $\overline{\varphi}_{\text{int}}$ (see Equ. (8.61)). We find

$$\overline{\Sigma}(\beta; L, w) = \frac{p-1}{L} \sum_{z=1}^{L} \frac{1}{w} \sum_{y=0}^{w-1} \overline{\Sigma}_e[q_{z+y}^+, \widehat{q}_z^+]$$
$$- \frac{l}{L} \sum_{z=1}^{L} \overline{\Sigma}_v\left[\widehat{q}_z^+\right] + \frac{R}{L+w-1} \sum_{z=1}^{L+w-1} \overline{\Sigma}_v\left[q_z^+\right], \quad (8.66)$$

where

$$\overline{\Sigma}_v[q^+] = \int d\eta\, q^+(\eta) \ln(1 + \tanh \beta \eta),$$
$$\overline{\Sigma}_e[q^+, \widehat{q}^+] = \int d\eta d\widehat{\eta}\, q^+(\eta)\widehat{q}^+(\widehat{\eta}) \ln(1 + \tanh \beta \eta \tanh \beta \widehat{\eta}). \quad (8.67)$$

For the underlying ensemble ($L = w = 1$) the complexity reduces to

$$\overline{\Sigma}(\beta) = (p-1)\overline{\Sigma}_e[q^+, \widehat{q}^+] - p\overline{\Sigma}_v[\widehat{q}^+] + R\overline{\Sigma}_v[q^+]. \quad (8.68)$$

The average distortion or internal energy (see (7.15), (7.16)) at temperature β is obtained by differentiating (8.65), which yields the simple formula

$$\frac{1}{2}u = \frac{1 - \tanh \beta}{2} \quad \text{for } \beta < \beta_c. \quad (8.69)$$

It has to be noted that this expression is only valid for $\beta < \beta_c$. To obtain the optimal distortion of the ensemble D_{opt} (see table 7.1) one needs to recourse to the full cavity formulas in order to take the limit $\beta \to +\infty$.

It is much simpler to solve the simplified fixed-point equations (8.63), (8.64). The population dynamics algorithm is almost the same as in Algorithm 8.1. The only difference is that J is generated according to the p.d.f. $(1 + J \tanh \beta)/2$ instead of $\text{Ber}(1/2)$. The big advantage is that there is no need to generate the 2^{r-1} configurations $\sigma_1, ..., \sigma_{r-1}$ which reduces the complexity of each iteration.

As expected the complexity obtained in either way is the same up to numerical precision.

8.3.5 Large Degree Limit

Inspection of the fixed-point equations (8.63) and (8.64) shows that the distributions[7]

$$q^+(\eta) = \delta_{+\infty}(\eta), \text{ and } \widehat{q}^+(\widehat{\eta}) = \sum_{J=\pm 1} \frac{1 + J \tanh \beta}{2} \delta(\widehat{\eta} - J). \quad (8.70)$$

---

[7]Here we adopt the notation $\delta_{+\infty}$ for a unit mass distribution at infinity.

are a fixed-point solution in the limit $p \to +\infty$, R fixed. This is (partially) justified by Theorem C.2 in Appendix C. The fixed point (8.70) leads to a complexity for the underlying model for $p \to +\infty$,

$$\lim_{p \to +\infty} \overline{\Sigma}(\beta) = (R-1)\ln 2 - \sum_{J=\pm 1} \frac{1 + J\tanh\beta}{2} \ln\left(\frac{1 + J\tanh\beta}{2}\right). \tag{8.71}$$

The condensation threshold $\lim_{p \to +\infty} \beta_c$ is obtained by setting this expression to zero

$$1 - R = \lim_{p \to +\infty} h_2\left(\frac{1 + \tanh\beta_c}{2}\right), \tag{8.72}$$

which is equivalent to

$$\lim_{p \to +\infty} \beta_c = \beta_{\mathrm{sh}} \equiv \frac{1}{2}\ln\left(\frac{1 - D_{\mathrm{sh}}(R)}{D_{\mathrm{sh}}(R)}\right). \tag{8.73}$$

In the large degree limit the condensation threshold is equal to the amplitude of half-log-likelihood ratios of a BSC test-channel with probability of error $D_{\mathrm{sh}}(R)$, i.e., tuned to capacity. Moreover the average distortion or internal energy is given by

$$\lim_{p \to +\infty} \frac{1}{2} u(\beta) = \begin{cases} \frac{1}{2}(1 - \tanh\beta) & \beta < \beta_{\mathrm{sh}}(R) \\ D_{\mathrm{sh}}(R) & \beta \geq \beta_{\mathrm{sh}}(R) \end{cases} \tag{8.74}$$

8.4 The Predictions of the Cavity Method

8.4.1 Complexity and Thresholds of the Underlying and Coupled Ensembles

We have computed the complexity and the thresholds from the cavity theory. These have been computed both from the full cavity equations of Section 8.3.1 and from the simplified ones of Section 8.3.4. Numerical values of the dynamical and condensation thresholds are presented in Tables 8.1 and 8.2. Results are obtained with population sizes $n = 30000$ (uncoupled), $n = 500 - 1000$ (coupled), and iteration number $t_{\max} = 3000$.

Since the free energies of the coupled and underlying ensembles are the same in the limit of infinite length (known from Theorem 7.1) and the condensation threshold is a singularity of the free energy (known from the cavity method), we can conclude on theoretical grounds that

$$\lim_{L \to +\infty} \beta_c(L, w) = \beta_c(w = 1). \tag{8.75}$$

The right-hand side of Equation (8.75) is the condensation threshold of the underlying system that is equivalent to a coupled system with $w = 1$.

Table 8.1: The numerical values of β_d and β_c for SCLDGM($p, R = 0.5, L, w = 3$) ensembles with $p = 3, 4$, and 5 and different values of L.

p	β	uncoupled ensemble	L		
			32	64	128
3	β_d	0.883	0.942	0.941	0.941
	β_c	0.940	0.958	0.948	0.946
4	β_d	0.875	1.010	1.010	1.009
	β_c	1.010	1.038	1.023	1.017
5	β_d	0.832	1.032	1.030	1.029
	β_c	1.032	1.067	1.048	1.039

Table 8.1 shows that the condensation threshold $\beta_c(L, w)$ of the coupled ensemble is higher than β_c for the uncoupled ensemble and decreases as L increases. The finite size effects are still clearly visible at lengths $L = 128$ and are more marked for larger w. This is not surprising since we expect the finite size corrections to be of order $O(w/L)$.

Let us now discuss the behavior of the dynamical threshold. Table 8.2 displays the results for the ensembles LDGM($p = 5, R = 0.5$) and SCLDGM($p = 5, R = 0.5, L, w$).

Table 8.2: The numerical values of β_d and β_c for SCLDGM($p = 5, R = 0.5, L, w$) ensembles with different values of L and w.

L	β	w			
		1	2	3	4
128	β_d	0.832	1.028	1.029	1.030
	β_c	1.032	1.038	1.039	1.043
256	β_d	0.832	1.023	1.027	1.029
	β_c	1.032	1.035	1.037	1.038

The column $w = 1$ gives the dynamical and condensation thresholds of the underlying ensemble, $\beta_d(w = 1)$ and $\beta_c(w = 1)$. We see that for each fixed L the dynamical threshold increases as a function of w. Closer inspection suggests that

$$\lim_{w \to +\infty} \lim_{L \to +\infty} \beta_d(L, w) = \beta_c(w = 1). \tag{8.76}$$

Equ. 8.76 indicates a threshold saturation phenomenon: for the coupled ensemble the phase of non-zero complexity shrinks to zero and the condensation point remains unchanged. This is analogous to the saturation of the BP threshold of LDPC codes towards the MAP threshold [78]. It is also analogous to the saturation of spinodal points in the Curie-Weiss chain [82]. Similar observations have been discussed for constraint

satisfaction problems in [83].

8.4.2   Comparison of β^* with β_d

We systematically observe that the optimal algorithmic value β^* of the BPGD-h algorithm is always lower, but somewhat close to β_d. For example for the uncoupled case $p = 5$ we have $(\beta^*, \beta_d) \approx (0.71, 0.832)$. For the coupled ensembles with $(L = 64, w = 3)$ we have $(\beta^*, \beta_d) \approx (1.03, 1.038)$. In fact, in the coupled case we observe $\beta^* \approx \beta_d \approx \beta_c$. Thus for the coupled ensemble BPGD-h operates well even close to the condensation threshold.

This is also the case for BPGD-r as we explain in the next paragraph. We use this fact in the next section to explain the good performance of the algorithm for coupled instances.

8.4.3   Sampling of the Gibbs Distribution with BPGD-r

Threshold saturation, equation (8.76), indicates that for L large, the phase of non-zero complexity, occupies a very small portion of the phase diagram close to β_c. This then suggests that for coupled ensembles Markov chain Monte Carlo dynamics, and BPGD-r algorithms are able to correctly sample the Gibbs measure for values of β up to $\approx \beta_c$. Let us discus in more detail this aspect of the BPGD-r algorithm.

By the Bayes rule:

$$\mu_\beta(\underline{u} \mid \underline{x}) = \prod_{i=1}^N \mu_\beta(u_i | \underline{x}, u_1, \ldots, u_{i-1}). \tag{8.77}$$

Thus we can sample $\underline{u}$ by first sampling u_1 from $\mu_\beta(u_1 | \underline{x})$, then u_2 from $\mu_\beta(u_2 | \underline{x}, u_1)$ and so on. Then, computing $x_a = \oplus_{i \in \partial a} u_i$ and the resulting average distortion, yields half the internal energy $u(\beta)/2$. With the BPGD-r algorithm the average distortion is computed in the same way except that the sampling is done with the BP marginals. So as long as the BP marginals are a good approximation of the true marginals, the average distortion $D_{\text{BPGD}-r}(\beta)$ should be close to $u(\beta)/2$. This can be conveniently tested because the cavity method predicts the simple formula (8.69) for the internal energy.

In Figure 8.3 we observe $D_{\text{BPGD}-r}(\beta) \approx (1 - \tanh \beta)/2$ for $\beta < \beta'$, with a value of β' lower but comparable to β_d. In particular for a coupled ensemble we observe $\beta' \approx \beta_d \approx \beta_c$. So Figure 8.3 strongly suggests that BPGD-r correctly samples the Gibbs distribution of coupled instances all the way up to $\approx \beta_c$, and that BP correctly computes marginals for the same range.
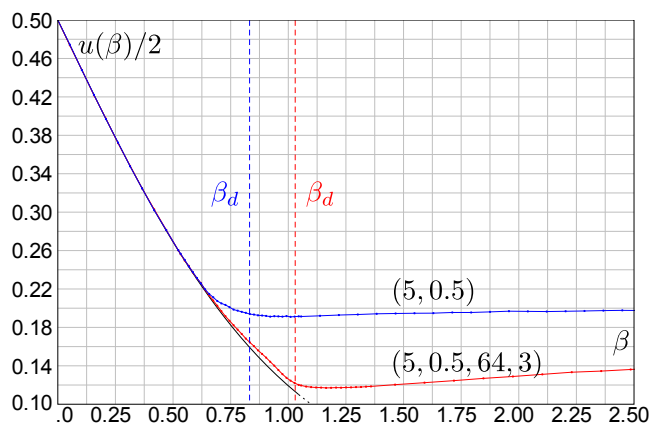
Figure 8.3: The performance of the BPGD-r algorithm. The plot shows that the algorithm can approximate average distortion quite precisely for $\beta < \beta' \approx \beta_d$. The black curve shows the average distortion $u(\beta)/2 = (1 - \tanh\beta)/2$ for $\beta < \beta_c$. The results are obtained for the underlying LDGM$(5, 0.5, 128000)$ and coupled SCLDGM$(5, 0.5, 64, 3, 2000)$ ensembles. The results are averaged over 50 instances. Numerical values of various thresholds are $\beta_{\mathrm{d,un}} = 0.832$, $\beta_{\mathrm{d,cou}} = 1.030$, $\beta_{\mathrm{c}} = 1.032$.

8.4.4 Large Degree Limit

According to the information theoretic approach to rate-distortion theory, we can view the encoding problem as a decoding problem for a random linear code on a BSC (q) test-channel with noise $q = D_{\mathrm{sh}}(R)$. Now, the Gibbs distribution (7.8) with $\beta = \frac{1}{2}\ln(1-q)/q$ is a MAP-decoder measure for a channel problem with the noise tuned to the Shannon limit. Moreover, for large degrees the LDGM ensemble is expected to be equivalent to the random linear code ensemble. These two remarks suggest that, since in the case of coupled ensembles with large degrees the BPGD-h encoder with optimal β^* approaches the rate-distortion limit, we should have

$$\beta^* \approx \frac{1}{2}\ln\frac{1-q}{q} \equiv \frac{1}{2}\ln\frac{1 - D_{\mathrm{sh}}(R)}{D_{\mathrm{sh}}(R)}. \tag{8.78}$$

In fact this is true. Indeed on the one hand, as explained above, for coupled codes we find $\beta^* \approx \beta_d \approx \beta_c$ (even for finite degrees). On the other hand the analytical *large degree* analysis of the cavity equations in Section 8.3.5 enables us to compute the complexity and to show the remarkable relation

$$\beta_c \approx \frac{1}{2}\ln\frac{1 - D_{\mathrm{sh}}(R)}{D_{\mathrm{sh}}(R)}, \text{ for } p \gg 1. \tag{8.79}$$

These remarks also show that the rate-distortion curve can be interpreted as a line of condensation thresholds for each R.

8.5   Discussion

8.5.1   Summary

We have investigated a simple spatially-coupled LDGM code ensemble for lossy source coding. No optimization on the degree distribution is required: the check degree is regular and the code-bit degree is Poisson. We have shown that the algorithmic rate-distortion curve of a low-complexity encoder based on BPGD enables us to approach the ultimate Shannon rate-distortion curve, for all compression rates, when the check degree grows large. The inverse temperature parameter (or equivalently, test-channel parameter) of the encoder could be optimized. However we have observed numerically and we have argued based on large degree calculations that a good universal choice is $\beta_{\mathrm{sh}}(R)$, given by tuning the test channel to capacity. We recall that for the underlying (uncoupled) ensemble the same encoder does not perform well; indeed, as the degree grows large, the difference between the algorithmic rate-distortion and Shannon rate-distortion curves grows. Insight into the excellent performance of the BPGD algorithm for spatially-coupled ensembles is gained by studying the phase diagram of the Gibbs measure on which the BPGD encoder is based. We have found, by applying the cavity method to the spatially-coupled ensemble, that the dynamical (inverse temperature) threshold β_d saturates towards the condensation (inverse temperature) threshold β_c. For this reason the BPGD encoder can operate close to the condensation threshold β_c, which itself tends in the large degree limit to $\beta_{\mathrm{sh}}(R)$, the test channel parameter tuned at capacity. For the underlying (uncoupled) ensemble the dynamical threshold moves in the opposite direction in the large degree limit so that the BPGD algorithm cannot operate close to the Shannon limit.

8.5.2   Open Questions

We mention some open questions that are left out by the present study and would deserve more investigations.

Tracking the Evolution of the Dynamical Threshold

For fixed degrees, the best value of the inverse temperature β_* of the BPGD algorithm is close to, but systematically lower, than the dynamical temperature β_d. Although the value of β_d can be calculated by the cavity theory, here we determine β_* by purely empirical means and it is not clear what the theoretical principles that enable us to determine its value are. As the graph is decimated the degree distribution changes and the effective dynamical temperature of the decimated graphs should evolve to slightly different values. It is tempting to conjecture that β_* is the limit of such a sequence of dynamical temperatures. A related phenomenon has been observed for the dynamical threshold with respect to clause density for random constraint satisfaction problems in

their SAT phase [111].

Analyzing the Decimation Process

The decimation process used in this chapter is hard to analyze rigorously because it is not clear how to keep track of the statistics of the decimated graph. However we would like to point out that a related process has been successfully analyzed in recent works [112] for the K-SAT problem in the large K limit up to the dynamical threshold (in the SAT phase). These methods could also be of use in the present case.

Proof of Threshold Saturation

Finally, whereas a rigorous control of the full cavity method is, in general, beyond present mathematical technology, there are sub-problems for which progress can presumably be made. For example, in the present case we have observed that the cavity equations reduce (in the dynamical phase $\beta_d < \beta < \beta_c$) to density evolution equations for an LDGM code on a BSC. The saturation of the dynamical temperature β_d to the condensation temperature β_c appears to be very similar to the threshold saturation phenomenon of channel coding theory. By now, we have a host of mathematical methods pertaining to this effect for LDPC on general binary memoryless channels [78], [80]. We think that these methods could be adapted to prove the saturation of β_d towards β_c. One extra difficulty faced in the present problem is that the "trivial" fixed-point of density-evolution equations of LDPC codes is not always present in the LDGM case.

A The All-zero Codeword Assumption for Symmetric Channels

The all-zero codeword assumption for symmetric channels is general for linear codes and symmetric decoder. We mention [59] for an extensive proof. We take the opportunity in this appendix to give an explanation of this assumption in the language of Gibbs measures and spins systems. The proof is given for LDPC codes but in fact does not require that the parity-check matrix is sparse.

We recall that a binary string $\underline{x}$ and its spin sequence representation $\underline{\sigma}$ are linked by the transformation

$$\sigma_i = (-1)^{x_i}. \tag{A.1}$$

We call $\underline{\sigma}$ a general codeword, $\underline{\sigma}^0$ the emitted codeword and $\underline{s}$ the received message after transmission over the channel. We say that a channel is symmetric if its transition probability satisfies

$$q\left(s_i \mid \sigma_i\right) = q\left(-s_i \mid -\sigma_i\right). \tag{A.2}$$

If we receive the message $\underline{s}$, the probability that the codeword $\underline{\sigma}$ was sent is given by the Gibbs measure

$$\mu_{\text{LDPC}}\left(\underline{\sigma} \mid \underline{h}\right) = \frac{1}{Z_{\text{LDPC}}} \prod_{a \in C} \left(1 + \prod_{i \in \partial a} \sigma_i\right) \prod_{i \in V} e^{h_i \sigma_i}, \tag{A.3}$$

where h_i are the half log-likelihood ratios

$$h_i\left(s_i\right) = \frac{1}{2} \ln \frac{q\left(s_i \mid 1\right)}{q\left(s_i \mid -1\right)}. \tag{A.4}$$

The probability distribution of the log-likelihood ratios is based on the distribution of the received message

$$\mathbb{P}_{\underline{S}}\left(\underline{s}\right) = \prod_{i \in V} q\left(s_i \mid \sigma_i^0\right). \tag{A.5}$$

Appendix A. The All-zero Codeword Assumption for Symmetric Channels

We will show now that up to a change of variable in the Gibbs measure, the codeword sent can be chosen as the all-zero codeword. We introduce the change of variable denoted by a tilde

$$\widetilde{\underline{\sigma}} = \underline{\sigma} \cdot \underline{\sigma}^0, \tag{A.6}$$

were the product is realized componentwise. Notice first that the symmetry of the channel (A.2) implies

$$h_i\left(s_i\right)\sigma_i^0 = h_i\left(s_i\sigma_i^0\right). \tag{A.7}$$

Therefore reparametrizing the codewords $\underline{\sigma}$ in the Gibbs distribution is equivalent as reparametrizing the channel observation $\underline{s}$

$$\mu_{\mathrm{LDPC}}\left(\widetilde{\underline{\sigma}} \mid \underline{h}\left(s\right)\right) = \mu_{\mathrm{LDPC}}\left(\underline{\sigma} \mid \underline{h}\left(\widetilde{\underline{s}}\right)\right). \tag{A.8}$$

The difference is that now the reparametrized channel observation is distributed as if the the all-zero codeword was sent

$$\begin{aligned}
\mathbb{P}_{\underline{S}}\left(\widetilde{\underline{s}}\right) &= \prod_{i \in V} q\left(s_i\sigma_i^0 \mid \sigma_i^0\right) \\
&= \prod_{i \in V} q\left(s_i \mid 1\right).
\end{aligned} \tag{A.9}$$

B On the Number of Rooted Polymers

The convergence criterion (4.32) in the polymer expansion requires an evaluation of the "entropy" of rooted polymers. The term "entropy" has to be understood as the number of polymers of a given size. The following lemma gives a bound on the entropy of polymers on a d-regular graph. Its generalization to irregular bipartite graphs with degrees of variable nodes and check nodes bounded by l_{max} and r_{max} is straightforward by setting $d = \max\left(l_{\max}, r_{\max}\right)$.

Lemma B.1 (Bound on the number of rooted polymers). *Let $\Gamma = (V, E)$ be a d-regular graph with vertex set V and edge set E. The number of polymers γ (connected subgraphs) of size $|\gamma \cap V| = t$ rooted to any vertex $x \in V$ is upper-bounded by*

$$\sum_{\gamma \ni x} \mathbb{I}\left(|\gamma \cap V| = t\right) \leq e^{dt}.$$

Proof. A polymer $\gamma \ni x$ is uniquely determined by one of its spanning tree T_γ plus the complementary set of edges $\gamma \backslash T_\gamma$. Figure (B.1) shows an example of this injective mapping. We ask the following question: If T is a spanning tree, how many different polymers have T as a spanning tree? In other words how many combinations of complementary edges can be made once T is given? Let g be the graph spanned by T which contains the most edges. As T is a tree, $|T \cap E| = |T \cap V| - 1$. Therefore the number of complementary edges unspecified by a spanning tree is at most

$$
\begin{aligned}
|g \backslash T| &= |g \cap E| - |T \cap E| \\
&= |g \cap E| - |T \cap V| + 1 \\
&= |g \cap E| - |g \cap V| + 1 \\
&\leq \left(\frac{d}{2} - 1\right)|g \cap V| + 1.
\end{aligned} \tag{B.1}
$$

Denote by $A_t(x)$ the number of polymers of size t rooted in $x \in V$ and call B_t the
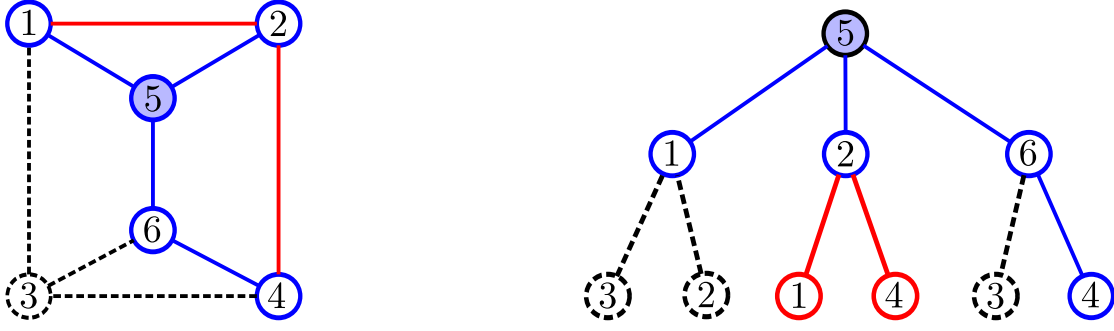
Figure B.1: On the left: a polymer is represented with colored solid lines. A spanning tree is shown in blue and the complementary edges in red. On the right: the spanning tree is shown on the computational tree in blue with a possible representation of the complementary edges in red.

number of rooted d-ary trees with size t. Based on the previous considerations

$$A_t\left(x\right) \leq 2^{\left(\frac{d}{2}-1\right)t+1}B_t. \tag{B.2}$$

To find a formula for B_t, we use a derivation based on generating functions similar to the one of Catalan numbers in [113]. Define the generating function

$$B\left(z\right) := \sum_{t=0}^{\infty} z^t B_t. \tag{B.3}$$

If one removes the root of a d-ary tree it splits the tree into d trees of smaller size. This yields the following equation for the generating function

$$B = 1 + zB^d. \tag{B.4}$$

By using the Lagrange-Bürmann formula on Equation (B.4) we find

$$B_t = \frac{1}{t\left(d-1\right)+1}\binom{td}{t}. \tag{B.5}$$

Finally we can relax the bound (B.2) to have a simpler expression by noticing that

$$2^{\left(\frac{d}{2}-1\right)t+1}B_t \leq e^{dt}. \tag{B.6}$$

$\square$

C Two Theorems for the Density Evolution Equations

Theorem C.1 provides theoretical support for the simplifications of the cavity equations discussed in Section 8.3.4.

Theorem C.1. *Consider the fixed-point equations (8.51) and (8.52) for the individual Poisson* LDGM(p, R) *ensemble with a fixed β. Take any initial continuous density $\hat{q}^{(0)}(\hat{h})$ and consider iterations $\hat{q}^{(t)}(\hat{h})$. There exists $p_0 \in \mathbb{N}$ such that for $p > p_0$, $\lim_{t \to \infty} \widehat{h}^{(t)} = 0$ almost surely.*

The proof[1] is presented in Section C.1. Note that p_0 depends on β and R. However we expect that as long as $\beta < \beta_c$ the result holds for all $p \geq 3$ and R. This is corroborated by the numerical observations. When we solve equations (8.51) and (8.52) by population dynamics with $\hat{q}^{(0)}(\hat{h})$ the uniform distribution, we observe that for a finite range of β depending on (p, R), the densities $q^{(t)}(h)$, $\widehat{q}^{(t)}(\widehat{h})$ tend to a Dirac distribution at the origin. The range of β for which this occurs always contains the interval $[0, \beta_c]$ irrespective of (p, R). These observations also hold for many other initial distributions. We note that these observations break down for β large enough.

Theorem C.2 partially justifies (8.70) which is the basis for the computation of the complexity in the large degree limit in Section 8.3.5.

Theorem C.2. *Consider the fixed-point equations (8.63) and (8.64) associated with the individual Poisson* LDGM(l, R) *ensemble for some p, R and β ($w = 1$ in the equations). Let $\widehat{\eta}^{(t)}$ be a random variable distributed according to $\widehat{q}^{+(t)}(\widehat{\eta})$ at iteration t. Assume that the initial density is*

$$\widehat{q}^{+(0)}(\hat{\eta}) = \sum_{J=\pm 1} \frac{1 + J \tanh(\beta)}{2} \delta(\hat{\eta} - J). \tag{C.1}$$

Then,

---

[1]It can be extended to other irregular degree distributions.

1. *For all t,*

$$\hat{q}^{+(t)}(-\hat{\eta}) = e^{-2\beta\hat{\eta}}\hat{q}^{+(t)}(\hat{\eta}),$$
$$q^{+(t)}(-\eta) = e^{-2\beta\eta}q^{+(t)}(\eta). \tag{C.2}$$

2. *For any $\delta > 0$, $\epsilon > 0$ and $B > 0$, there exits l_1 such that for $l > l_1$ and all t.*

$$\mathbb{P}\left\{1 - \epsilon \le \hat{\eta}^{(t)} \le 1\right\} > \frac{e^{2\beta}}{1 + e^{2\beta}}(1 - \delta), \tag{C.3}$$

$$\mathbb{P}\left\{-1 \le \hat{\eta}^{(t)} \le -1 + \epsilon\right\} > \frac{1}{1 + e^{2\beta}}(1 - \delta). \tag{C.4}$$

The proof is presented in Appendix C.2.

C.1 Proof of Theorem C.1

We first state two useful lemmas

Lemma C.1. *Let the random variable X be distributed according to a Poisson distribution with mean λ.*

$$\mathbb{P}(X < \frac{\lambda t}{2}) < \exp(-\frac{\lambda t}{10}), \quad t \le 1,$$
$$\mathbb{P}(X > \frac{3\lambda t}{2}) < \exp(-\frac{\lambda t}{10}), \quad t \ge 1. \tag{C.5}$$

Proof. Use the Chernoff bound. $\square$

Lemma C.2. *Let*

$$\epsilon_1 = \beta\frac{3p}{2R}(\tanh \epsilon_0)^{(pR^3)^{1/4}}$$
$$\delta_1 = \exp(-\frac{p}{10R}) + \frac{p}{R}\exp(-\frac{\epsilon_0\sqrt{Rp}}{\beta\sqrt{3\pi}}). \tag{C.6}$$

with $\epsilon_0 = \min(1/2, \beta/2)$. Consider the recursions for $t \ge 1$

$$\epsilon_{t+1} = (t + 1)\beta\frac{3p}{2R}(\tanh \epsilon_t)^{(pR)^{1/4}},$$
$$\delta_{t+1} = \exp(-\frac{p}{10R}(t + 1)) + \frac{p}{R}(2\sqrt{\delta_t})^{p-1}. \tag{C.7}$$

There exists an integer p_0 (depending only on R and β) such that for $p \ge p_0$,

1. *$\epsilon_t \le \frac{1}{2^{t+1}}$ for $t \ge 0$.*

2. $\delta_t < 2\exp(-\frac{p}{5R}t)$ *for* $t \geq 2$.

Proof. Consider the first property. At $t = 0$, $\epsilon_0 \leq 1/2$. Assume that $\epsilon_{t-1} \leq \frac{1}{2^t}$ for $t \geq 1$, then

$$\epsilon_t = t\beta\frac{3p}{2R}(\tanh\epsilon_{t-1})^{(pR^3)^{1/4}} \leq t\beta\frac{3p}{2R}(\epsilon_{t-1})^{(pR^3)^{1/4}}$$
$$\leq t\beta\frac{3p}{2R}(\frac{1}{2^t})^{(pR^3)^{1/4}} = \frac{t\beta\frac{3p}{R}}{2^{t((lR^3)^{1/4}-1)}} \times \frac{1}{2^{t+1}}. \tag{C.8}$$

The proof is complete if $t\beta\frac{3p}{R} < 2^{t(\sqrt[4]{pR^3}-1)}$ for $t \geq 1$. It is clear that this is true for p large enough.

Now consider the second property. Clearly for p large enough such that

$$\delta_2 = \exp\left(-\frac{p}{5R}\right) + \frac{p}{R}(2\sqrt{\delta_1})^{p-1} \leq 2\exp\left(-\frac{p}{5R}\right). \tag{C.9}$$

To complete the proof by induction, we remark that $\delta_t < 2\exp\left(-\frac{p}{5R}t)\right) < 1$ implies

$$\frac{p}{R}(2\sqrt{\delta_t})^{p-1} < \exp\left(-\frac{p}{5R}(t+1)\right) \tag{C.10}$$

for p large enough independent of t. $\qquad\square$

We now turn to the proof of Theorem C.1. It is organized in three steps:

1. We first show that for any small δ_1 and ϵ_1, one can find an integer p_1 such that for $p \geq p_1$

$$q_1 \equiv \mathbb{P}\left\{|h^{(1)}| \leq \frac{\epsilon_1}{\beta}\right\} \geq 1 - \delta_1. \tag{C.11}$$

2. We then show by induction on $t \geq 1$ that

$$q_t \equiv \mathbb{P}\left\{|h^{(t)}| < \frac{\epsilon_t}{\beta}\right\} \geq 1 - \delta_t. \tag{C.12}$$

3. Finally using Lemma C.2 we deduce that $h^{(t)} \to 0$ almost surely as $t \to +\infty$.

Proof. [Proof of Theorem C.1]

We begin by noting that regardless of the initial distribution, $\widehat{q}^{(t)}(\widehat{h})$ has a symmetric density due to the symmetric distribution of J. Moreover, $\left|\widehat{h}^{(t)}\right| \leq 1$ from (8.52). Thus, $\mathbb{E}_{\widehat{q}^{(t)}}(\widehat{h}^{(t)}) = 0$ and $\text{Var}(\widehat{h}^{(t)}) = \mathbb{E}_{\widehat{q}^{(t)}}(\widehat{h}^2) \leq 1$.

Appendix C. Two Theorems for the Density Evolution Equations

Step 1: We set $P(r) = e^{-\lambda} \frac{\lambda^r}{r!}$ and $\lambda = p/R$. Let $h^{(r,t)} = \sum_{a=1}^r \widehat{h}_a^{(t)}$ where $\widehat{h}_a^{(t)}$ are i.i.d. random variables with probability density $\widehat{q}^{(t)}(\widehat{h})$. Let $\sigma_0^2 = \mathbb{E}((\widehat{h}_a^{(0)})^2) \leq 1$. According to [114, Theorem 3.5.3] we have for any $\epsilon_0 > 0$. Thus, there exists $r'(\epsilon_0, \beta) \in \mathbb{N}$ such that for $r > r'$,

$$\mathbb{P}\left\{ |h^{(r,0)}| < \frac{\epsilon_0}{\beta} \right\} \geq \frac{\epsilon_0}{\beta\sqrt{2\pi r}}. \tag{C.13}$$

Take p such that $\lambda = p/R \geq p'/R = 2r'$, then

$$\begin{aligned}
q_0 &= \mathbb{P}\left\{ |h^{(0)}| < \frac{\epsilon_0}{\beta} \right\} \\
&= \sum_{r=0}^\infty P(r) \mathbb{P}\left\{ |h^{(r,0)}| < \frac{\epsilon_0}{\beta} \right\} \\
&\geq \sum_{r=\lambda/2}^{3\lambda/2} P(r) \mathbb{P}\left\{ |h^{(r,0)}| < \frac{\epsilon_0}{\beta} \right\} \\
&\geq \frac{\epsilon_0}{\beta\sqrt{3\pi\lambda}} \sum_{r=\lambda/2}^{3\lambda/2} P(r) \\
&> \frac{\epsilon_0}{\beta\sqrt{3\pi\lambda}}(1 - 2e^{-\frac{\lambda}{10}}). \tag{C.14}
\end{aligned}$$

The last inequality follows from lemma C.1. Thus for p large enough

$$q_0 = \mathbb{P}\left\{ |h^{(0)}| < \frac{\epsilon_0}{\beta} \right\} > \frac{\epsilon_0}{2\beta\sqrt{3\pi\lambda}} \equiv 1 - \delta_0. \tag{C.15}$$

Recall $\widehat{h}^{(t+1)} = \frac{1}{\beta} \tanh^{-1}(J \tanh\beta \prod_{i=1}^{p-1} \tanh\beta h_i^{(t)})$. From $\tanh^{-1}(a\tanh\beta) \leq a\beta$ for $0 < a < 1$, we have

$$\left| \widehat{h}^{(t+1)} \right| \leq \prod_{i=1}^{p-1} \tanh\left| \beta h_i^{(t)} \right|. \tag{C.16}$$

Define

$$Z_p^{(t)} \equiv \ln\left(\prod_{i=1}^{p-1} \tanh\left| \beta h_i^{(t)} \right| \right) = \sum_{i=1}^{p-1} \ln\left(\tanh\left| \beta h_i^{(t)} \right| \right). \tag{C.17}$$

Note that $Z_p^{(t)}$ is always negative and if one of $h_i^{(t)}$ tends to zero, it diverges to $-\infty$. Consider $t = 0$. We will show that $Z_p^{(0)}$ has a large negative value with high probability. Define

$$u_i \equiv \begin{cases} u_{i-1}, & \text{if } \left| h_{i-1}^{(0)} \right| > \frac{\epsilon_0}{\beta}, \\ u_{i-1} + \ln\tanh\epsilon_0, & \text{otherwise}, \end{cases} \tag{C.18}$$

with $u_0 = 0$. One can check for later use that $Z_p^{(0)} \leq u_p$. Moreover, because of (C.15)

one can consider u_p as a random walk (with negative jumps),

$$u_i = \begin{cases} u_{i-1}, & \text{with prob. } 1 - q_0 \\ u_{i-1} + \ln \tanh \epsilon_0, & \text{with prob. } q_0. \end{cases} \tag{C.19}$$

Let $s = \ln(\tanh(\epsilon_0))$. Using the Chernoff's theorem [115, Page 151],

$$\mathbb{P}\left\{\frac{1}{p-1}\frac{u_p}{s} < \lambda^{-3/4}\right\} < \exp\left(-(p-1)D(\lambda^{-3/4}||q_0)\right), \tag{C.20}$$

where $D(x||y) = x\ln(\frac{x}{y}) + (1-x)\ln(\frac{1-x}{1-y})$. Now, since

$$x\ln(\frac{x}{q_0}) > x\ln(x), \tag{C.21}$$

and

$$(1-x)\ln\left(\frac{1-x}{1-q_0}\right) > (1-x)\ln\left(\frac{1-x}{\delta_0}\right), \tag{C.22}$$

we have

$$D(\lambda^{-3/4}||q_0) > -H_2(\lambda^{-3/4})\ln(2) - (1 - \lambda^{-3/4})\ln(\delta_0), \tag{C.23}$$

for δ_0 defined in (C.15). By a large λ expansion of the right-hand side of (C.23):

$$-H_2(\lambda^{-3/4})\ln 2 - (1 - \lambda^{-3/4})\ln\delta_0 = \frac{\epsilon_0}{2\beta\sqrt{3\pi\lambda}} + o(\frac{1}{\sqrt{\lambda}}). \tag{C.24}$$

Thus, there exists $p'' \in \mathbb{N}$ depending on R, β and ϵ_0 such that for $p > p''$,

$$\mathbb{P}\left\{\frac{1}{p-1}\frac{u_p}{s} < \lambda^{-3/4}\right\} < \exp\left(-\frac{\epsilon_0(p-1)}{4\beta\sqrt{3\pi\lambda}}\right). \tag{C.25}$$

By replacing $s = \ln\tanh\epsilon_0$ and $\lambda = \frac{p}{R} \approx \frac{p-1}{R}$ for large degrees,

$$\mathbb{P}\left\{u_p > (lpR)^{1/4}\ln\tanh\epsilon_0\right\} < \exp\left(-\frac{\epsilon_0\sqrt{Rp}}{4\beta\sqrt{3\pi}}\right), \tag{C.26}$$

Note that the inequality in $\mathbb{P}(\dots)$ is reversed since $s < 0$. Now recall $Z_p^{(0)} \leq u_p$. Therefore,

$$\mathbb{P}\left\{Z_p^{(0)} \leq (pR)^{1/4}\ln\tanh\epsilon_0\right\} \geq \mathbb{P}\left\{u_p \leq (pR^3)^{1/4}\ln\tanh\epsilon_0\right\}$$

$$\geq 1 - \exp\left(-\frac{\epsilon_0\sqrt{Rp}}{4\beta\sqrt{3\pi}}\right). \tag{C.27}$$

Consequently,

$$\mathbb{P}\left\{\left|\widehat{h}^{(1)}\right| \le (\tanh \epsilon_0)^{(pR^3)^{1/4}}\right\} \ge \mathbb{P}\left\{Z_p^{(0)} \le (pR)^{1/4} \ln \tanh \epsilon_0\right\}$$

$$\ge 1 - \exp\left(-\frac{\epsilon_0 \sqrt{Rp}}{4\beta\sqrt{3\pi}}\right). \tag{C.28}$$

From $\left|h^{(r,1)}\right| = \left|\sum_{a=1}^{r} \widehat{h}_a^{(1)}\right| \le \sum_{a=1}^{r} \left|\widehat{h}_a^{(1)}\right|$, we deduce

$$\mathbb{P}\left\{\left|h^{(r,1)}\right| \le r(\tanh \epsilon_0)^{(pR^3)^{1/4}}\right\} \ge \mathbb{P}\left\{\left|\widehat{h}^{(1)}\right| \le (\tanh \epsilon_0)^{(pR^3)^{1/4}}\right\}^r$$

$$\ge \left\{1 - \exp\left(-\frac{\epsilon_0 \sqrt{Rp}}{4\beta\sqrt{3\pi}}\right)\right\}^r$$

$$\ge 1 - r\exp\left(-\frac{\epsilon_0 \sqrt{Rp}}{4\beta\sqrt{3\pi}}\right), \tag{C.29}$$

for p large enough. Therefore we have

$$\mathbb{P}\left\{\left|h^{(1)}\right| \le \frac{3}{2}\lambda(\tanh \epsilon_0)^{(pR^3)^{1/4}}\right\} = \sum_{r=0}^{\infty} P(r)\mathbb{P}\left\{\left|h^{(r,1)}\right| \le \frac{3}{2}\lambda(\tanh \epsilon_0)^{(pR^3)^{1/4}}\right\}$$

$$\ge \sum_{r=0}^{3\lambda/2} P(r)\mathbb{P}\left\{\left|h^{(r,1)}\right| \le \frac{3}{2}\lambda(\tanh \epsilon_0)^{(pR^3)^{1/4}}\right\}$$

$$\ge \sum_{r=0}^{3\lambda/2} P(r)\mathbb{P}\left\{\left|h^{(r,1)}\right| \le r(\tanh \epsilon_0)^{(pR^3)^{1/4}}\right\}$$

$$\ge \sum_{r=0}^{3\lambda/2} P(r)\left(1 - r\exp\left(-\frac{\epsilon_0 \sqrt{Rp}}{4\beta\sqrt{3\pi}}\right)\right)$$

$$\ge 1 - \exp(-0.1\lambda) - \lambda\exp\left(-\frac{\epsilon_0 \sqrt{Rp}}{4\beta\sqrt{3\pi}}\right). \tag{C.30}$$

To summarize, we have obtained

$$q_1 = \mathbb{P}\left\{\left|h^{(1)}\right| \le \frac{\epsilon_1}{\beta}\right\} \ge 1 - \delta_1. \tag{C.31}$$

This completes step 1.

Step 2: The proof is by induction. Assume that

$$q_t = \mathbb{P}\left\{|h^{(t)}| \le \frac{\epsilon_t}{\beta}\right\} \ge 1 - \delta_t. \tag{C.32}$$

We prove that this holds also for $t + 1$. This mainly consists in repeating the derivations

(C.15) to (C.31) for q_t, ϵ_t and δ_t. We briefly repeat them here:

$$\mathbb{P}\left\{\left|\widehat{h}^{(t+1)}\right| \leq (\tanh \epsilon_t)^{(pR^3)^{1/4}}\right\} \geq \mathbb{P}\left\{Z_l^{(t)} \leq (pR^3)^{1/4} \ln(\tanh \epsilon_t)\right\}$$
$$\geq 1 - \exp\left(-(p-1)D(\lambda^{-3/4}\|q_t)\right). \qquad \text{(C.33)}$$

Assume that $\delta_t \ll 1$. From (C.23),

$$D(\lambda^{-3/4}\|q_t) > -H_2(\lambda^{-3/4})\ln(2) - (1 - \lambda^{-3/4})\ln(\delta_t). \qquad \text{(C.34)}$$

If $\lambda^{-3/4} < \frac{1}{2}$ (equivalently, $p > 2^{4/3}R$),

$$D(\lambda^{-3/4}\|q_t) > -\ln 2 - \frac{1}{2}\ln \delta_t. \qquad \text{(C.35)}$$

Thus,

$$\mathbb{P}\left\{\left|\widehat{h}^{(t+1)}\right| \leq (\tanh \epsilon_t)^{(pR^3)^{1/4}}\right\} \geq 1 - (2\sqrt{\delta_t})^{p-1}, \qquad \text{(C.36)}$$

and finally,

$$\mathbb{P}\left\{\left|\widehat{h}^{(t+1)}\right| \leq (t+1)\frac{3}{2}\lambda(\tanh \epsilon_t)^{(lR^3)^{1/4}}\right\}$$
$$\geq \sum_{r=0}^{3(t+1)\lambda/2} P(r)\mathbb{P}\left\{\left|h^{(r,t+1)}\right| \leq (t+1)\frac{3}{2}\lambda(\tanh \epsilon_t)^{(pR^3)^{1/4}}\right\}$$
$$\geq \sum_{r=0}^{3(t+1)\lambda/2} P(r)\mathbb{P}\left\{\left|h^{(r,t+1)}\right| \leq r(\tanh \epsilon_t)^{(pR^3)^{1/4}}\right\}$$
$$\geq \sum_{r=0}^{3(t+1)\lambda/2} P(r)\left(1 - r(2\sqrt{\delta_t})^{p-1}\right)$$
$$\geq 1 - \exp(-(t+1)\frac{\lambda}{10}) - \lambda(2\sqrt{\delta_t})^{p-1}. \qquad \text{(C.37)}$$

Or equivalently,

$$p_{t+1} = \mathbb{P}\left\{|h^{(t+1)}| < \frac{\epsilon_{t+1}}{\beta}\right\} \geq 1 - \delta_{t+1}. \qquad \text{(C.38)}$$

This completes step 2.

Step 3: Using lemma C.2, for l large enough (depending on β and R, but independent of t)

$$\mathbb{P}\left\{\left|h^{(t)}\right| > \frac{1}{\beta 2^{(t+1)}}\right\} \leq \delta_t \leq 2\exp\left(-\frac{p}{5R}t\right). \qquad \text{(C.39)}$$

The Borel-Cantelli lemma [114, Theorem 2.3.1] states that, $h^{(t)} \to 0$ almost surely if for all $\alpha > 0$,

$$\sum_{t=1}^{\infty} \mathbb{P}\left\{\left|h^{(t)}\right| > \alpha\right\} < +\infty. \qquad \text{(C.40)}$$

Let us verify that $h^{(t)}$ has this property. For any α, there is τ such that $1/2^{\tau+1} < \beta\alpha$. Therefore, for $t \geq \tau$,

$$\mathbb{P}\left\{\left|h^{(t)}\right| > \epsilon\right\} \leq \mathbb{P}\left\{\left|h^{(t)}\right| > \frac{1}{2^{(t+1)\beta}}\right\} < \delta_t \tag{C.41}$$

and hence,

$$\sum_{t=1}^{\infty} \mathbb{P}\left\{\left|h^{(t)}\right| > \epsilon\right\} \leq \tau + \sum_{t=\tau}^{\infty} \mathbb{P}\left\{\left|h^{(t)}\right| > \epsilon\right\}$$

$$< \tau + \sum_{t=\tau}^{\infty} \delta_t$$

$$< \tau + \sum_{t=\tau}^{\infty} 2\exp\left(-\frac{p}{10R}t\right) < +\infty. \tag{C.42}$$

This completes step 3. $\qquad\square$

C.2 Proof of Theorem C.2

Proof. We first show the first property, which is called the Nishimori symmetry. Note that it is satisfied by $\widehat{q}^{+(0)}$ and $q^{+(0)}$. The equations (8.63) and (8.64) are the density evolution equations associated to an LDGM code over the BSC. It is known that the Nishimori symmetry is preserved under density evolution recursions (e.g., see [59] for similar properties in the case of LDPC codes).

Let us turn to the proof of the second property. First note that (C.3) implies (C.4). Indeed

$$\mathbb{P}\{\widehat{\eta}^{(t)} < -1+\epsilon\} = \int_{-1}^{-1+\epsilon} \widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta}$$

$$= \int_{1-\epsilon}^{1} e^{-2\beta\widehat{\eta}}\widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta}$$

$$\geq e^{-2\beta}\mathbb{P}\{\widehat{\eta}^{(t)} > 1-\epsilon\}$$

$$\geq \frac{1}{1+e^{2\beta}}(1-\delta). \tag{C.43}$$

So we only have to prove (C.3). We will use induction. The induction hypothesis is (C.3) for some $\delta > 0$ and $\epsilon > 0$ at iteration t. It is obviously true at $t = 0$. Let us first show that

$$\mathbb{E}(\eta^{(t)}) = \lambda\mathbb{E}(\widehat{\eta}^{(t)}) \geq 2\lambda s. \tag{C.44}$$

for $s = \frac{1}{2}(1 - \delta)(1 - \epsilon)(1 - e^{-2\beta(1-\epsilon)})/(1 + e^{-2\beta})$. We have

$$
\begin{aligned}
\mathbb{E}(\widehat{\eta}^{(t)}) &= \int_{-1}^{1} \widehat{\eta}\widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta} \\
&= \int_{-1}^{0} \widehat{\eta}\widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta} + \int_{0}^{1} \widehat{\eta}\widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta} \\
&= -\int_{0}^{1} \widehat{\eta}e^{-2\beta\widehat{\eta}}\widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta} + \int_{0}^{1} \widehat{\eta}\widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta} \\
&= \int_{0}^{1} \widehat{\eta}(1 - e^{-2\beta\widehat{\eta}})\widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta} \\
&\geq \int_{1-\epsilon}^{1} \widehat{\eta}(1 - e^{-2\beta\widehat{\eta}})\widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta} \\
&\geq (1 - e^{-2\beta(1-\epsilon)})(1 - \epsilon)\int_{1-\epsilon}^{1} \widehat{q}^{+(t)}(\widehat{\eta})\mathrm{d}\widehat{\eta} \\
&> (1 - \delta)(1 - \epsilon)\frac{1 - e^{-2\beta(1-\epsilon)}}{1 + e^{-2\beta}}.
\end{aligned}
\tag{C.45}
$$

This proves (C.44).

By applying Hoeffding's inequality [115] for $\lambda/2 < r < 3\lambda/2$,

$$
\begin{aligned}
\mathbb{P}\left\{\sum_{a=1}^{r} \widehat{\eta}_{a}^{(t)} < \lambda\frac{s}{2}\right\} &= \mathbb{P}\left\{\sum_{a=1}^{r}(\widehat{\eta}_{a}^{(t)} - \mathbb{E}(\widehat{\eta}^{(t)})) < \lambda\frac{s}{2} - r\mathbb{E}(\widehat{\eta}^{(t)})\right\} \\
&\leq \mathbb{P}\left\{\sum_{a=1}^{r}(\widehat{\eta}_{a}^{(t)} - \mathbb{E}(\widehat{\eta}^{(t)})) < \lambda\frac{s}{2} - 2rs\right\} \\
&\leq \mathbb{P}\left\{\sum_{a=1}^{r}(\widehat{\eta}_{a}^{(t)} - \mathbb{E}(\widehat{\eta}^{(t)})) < -\lambda\frac{s}{2}\right\} \\
&< \exp(-\frac{\lambda^{2}s^{2}}{8r}) \\
&< \exp(-\lambda\frac{s^{2}}{12}).
\end{aligned}
\tag{C.46}
$$

From

$$
\begin{aligned}
\mathbb{P}\left\{\eta^{(t)} < \lambda\frac{s}{2}\right\} &= \sum_{r=0}^{\infty} P(r)\mathbb{P}\left\{\sum_{a=1}^{r} \widehat{\eta}_{a}^{(t)} < \lambda\frac{s}{2}\right\} \\
&\leq \sum_{r=0}^{\lambda/2} P(r) + \sum_{r=\lambda/2}^{3\lambda/2} P(r)\mathbb{P}\left\{\sum_{a=1}^{r} \widehat{\eta}_{a}^{(t)} < \lambda\frac{s}{2}\right\} + \sum_{r=3\lambda/2}^{\infty} P(r).
\end{aligned}
\tag{C.47}
$$

and Lemma C.1, we get

$$
\mathbb{P}\left\{\eta^{(t)} > \lambda\frac{s}{2}\right\} > 1 - 2\exp\left(-\frac{\lambda}{10}\right) - \exp(-\lambda\frac{s^{2}}{12}).
\tag{C.48}
$$

Appendix C. Two Theorems for the Density Evolution Equations

Now consider the density evolution equation (8.64). We have

$$\mathbb{P}\left\{\widehat{\eta}^{(t+1)} > \frac{1}{\beta}\text{atanh}\left(\tanh(\beta)\left[\tanh(\beta\lambda\frac{s}{2})\right]^{p-1}\right)\right\}$$

$$\geq \mathbb{P}\left\{J = 1, \eta_1^{(t)} > \frac{\lambda s}{2}, \ldots, \eta_{p-1}^{(t)} > \frac{\lambda s}{2}\right\}$$

$$= \frac{1 + \tanh(\beta)}{2}\left(\mathbb{P}\left\{\eta^{(t)} > \frac{\lambda s}{2}\right\}\right)^{p-1}$$

$$\geq \frac{1 + \tanh(\beta)}{2}\left(1 - 2\exp\left(-\frac{\lambda}{10}\right) - \exp(-\lambda\frac{s^2}{12})\right)^{p-1}$$

$$\geq \frac{e^{2\beta}}{1 + e^{2\beta}}\left(1 - (p-1)\left(2\exp(-\frac{p}{10R}) + \exp(-\frac{ps^2}{12R})\right)\right) \tag{C.49}$$

Let

$$1 - \varepsilon(p, R, \beta) = \frac{1}{\beta}\text{atanh}\left(\tanh(\beta)\left[\tanh(\beta s\frac{p}{2R})\right]^{p-1}\right), \tag{C.50}$$

and

$$\Delta(p, R) = (p-1)\left(2\exp(-0.1\frac{p}{R}) + \exp(-\frac{ps^2}{12R})\right). \tag{C.51}$$

Inequality (C.3) holds at $t+1$, if $\varepsilon(p, R, \beta) \leq \epsilon$ and $\Delta(p, R) \leq \delta$. This is true for $p > p_1$ large enough since $\varepsilon(p, R, \beta)$ and $\Delta(p, R)$ are decreasing functions of p (for large values of p). $\square$

Bibliography

[1] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," *Phys. Rev. Lett.*, vol. 35, pp. 1792–1796, Dec 1975.

[2] G. Parisi, "Infinite number of order parameters for spin-glasses," *Physical Review Letters*, vol. 43, no. 23, p. 1754, 1979.

[3] G. Parisi, "A sequence of approximated solutions to the SK model for spin glasses," *Journal of Physics A: Mathematical and General*, vol. 13, no. 4, p. L115, 1980.

[4] M. Mézard, G. Parisi, and M. Virasoro, "SK model: The replica solution without replicas," *Europhys. Lett*, vol. 1, no. 2, pp. 77–82, 1986.

[5] M. Bailly-Bechet, A. Braunstein, A. Pagnani, M. Weigt, and R. Zecchina, "Inference of sparse combinatorial-control networks from gene-expression data: a message passing approach," *BMC bioinformatics*, vol. 11, no. 1, p. 355, 2010.

[6] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, "Identification of direct residue contacts in protein–protein interaction by message passing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 1, pp. 67–72, 2009.

[7] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, p. 021005, May 2012.

[8] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, 2012.

[9] M. Welling and Y. W. Teh, "Approximate inference in Boltzmann machines," *Artificial Intelligence*, vol. 143, no. 1, pp. 19–50, 2003.

[10] T. Tanaka, "Mean-field theory of Boltzmann machine learning," *Physical Review E*, vol. 58, no. 2, p. 2302, 1998.

[11] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–173, 2011.

Bibliography

[12] M. Mézard, "The space of interactions in neural networks: Gardner's computation with the cavity method," *Journal of Physics A: Mathematical and General*, vol. 22, no. 12, p. 2181, 1989.

[13] Y. Kabashima and D. Saad, "Statistical mechanics of low-density parity-check codes," *Journal of Physics A: Mathematical and General*, vol. 37, no. 6, p. R1, 2004.

[14] T. Murayama, "Thouless-Anderson-Palmer approach for lossy compression," *Phys. Rev. E*, vol. 69, p. 035105, Mar 2004.

[15] S. Mertens, M. Mézard, and R. Zecchina, "Threshold values of random K-SAT from the cavity method," *Random Structures & Algorithms*, vol. 28, no. 3, pp. 340–373, 2006.

[16] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, "Clusters of solutions and replica symmetry breaking in random K-satisfiability," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 04, p. P04004, 2008.

[17] F. Guerra, "Sum rules for the free energy in the mean field spin glass model," *Fields Institute Communications*, vol. 30, p. 161, 2001.

[18] M. Talagrand, "The Parisi formula," *Annals of Mathematics*, vol. 163, no. 1, pp. pp. 221–263, 2006.

[19] E. Ising, "Beitrag zur theorie des ferromagnetismus," *Zeitschrift für Physik*, vol. 31, no. 1, pp. 253–258, 1925.

[20] L. Onsager, "Crystal statistics. i. a two-dimensional model with an order-disorder transition," *Phys. Rev.*, vol. 65, pp. 117–149, Feb 1944.

[21] M. Aizenman, "Translation invariance and instability of phase coexistence in the two dimensional Ising system," *Communications in Mathematical Physics*, vol. 73, no. 1, pp. 83–94, 1980.

[22] C. N. Yang, "The spontaneous magnetization of a two-dimensional Ising model," *Phys. Rev.*, vol. 85, pp. 808–816, Mar 1952.

[23] T. T. Wu, B. M. McCoy, C. A. Tracy, and E. Barouch, "Spin-spin correlation functions for the two-dimensional Ising model: Exact theory in the scaling region," *Physical Review B*, vol. 13, no. 1, p. 316, 1976.

[24] P. G. Debenedetti and F. H. Stillinger, "Supercooled liquids and the glass transition," *Nature*, vol. 410, no. 6825, pp. 259–267, 2001.

[25] P. G. Debenedetti, T. M. Truskett, C. P. Lewis, and F. H. Stillinger, "Theory of supercooled liquids and glasses: energy landscape and statistical geometry perspectives," *Advances in Chemical Engineering*, vol. 28, pp. 21–79, 2001.

[26] G. Parisi, "Spin glasses and fragile glasses: Statics, dynamics, and complexity," *Proceedings of the National Academy of Sciences*, vol. 103, no. 21, pp. 7948–7955, 2006.

[27] J. Zarzycki, *Glasses and the vitreous state*, vol. 9. Cambridge University Press, 1991.

[28] G. W. White and S. H. Cakebread, "The glassy state in certain sugar-containing food products," *International Journal of Food Science & Technology*, vol. 1, no. 1, pp. 73–82, 1966.

[29] J. H. Crowe, J. F. Carpenter, and L. M. Crowe, "The role of vitrification in anhydrobiosis," *Annual Review of Physiology*, vol. 60, no. 1, pp. 73–103, 1998. PMID: 9558455.

[30] A. L. Greer and N. Mathur, "Materials science: Changing face of the chameleon," *Nature*, vol. 437, no. 7063, pp. 1246–1247, 2005.

[31] K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, and H. Hosono, "Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors," *Nature*, vol. 432, no. 7016, pp. 488–492, 2004.

[32] D. Uhlmann, "A kinetic treatment of glass formation," *Journal of Non-Crystalline Solids*, vol. 7, no. 4, pp. 337–348, 1972.

[33] R. Brüning and K. Samwer, "Glass transition on long time scales," *Physical Review B*, vol. 46, no. 18, p. 11318, 1992.

[34] C. T. Moynihan, A. J. Easteal, J. Wilder, and J. Tucker, "Dependence of the glass transition temperature on heating and cooling rate," *The Journal of Physical Chemistry*, vol. 78, no. 26, pp. 2673–2677, 1974.

[35] P. G. Debenedetti, *Metastable liquids: concepts and principles.* Princeton University Press, 1996.

[36] F. H. Stillinger, "A topographic view of supercooled liquids and glass formation," *Science*, vol. 267, no. 5206, pp. 1935–1939, 1995.

[37] M. Goldstein, "Viscous liquids and the glass transition: a potential energy barrier picture," *The Journal of Chemical Physics*, vol. 51, p. 3728, 1969.

[38] P. Goldbart, N. Goldenfeld, D. Sherrington, and S. Edwards, *Stealing the Gold: A Celebration of the Pioneering Physics of Sam Edwards.* Oxford University Press, 2007.

[39] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, p. 965, 1975.

Bibliography

[40] F. Guerra and F. L. Toninelli, "The thermodynamic limit in mean field spin glass models," *Communications in Mathematical Physics*, vol. 230, no. 1, pp. 71–79, 2002.

[41] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina, "Two solutions to diluted p-spin models and XORSAT problems," *Journal of Statistical Physics*, vol. 111, no. 3-4, pp. 505–533, 2003.

[42] H. Bethe, "Statistical theory of superlattices," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 150, no. 871, pp. 552–575, 1935.

[43] R. Peierls, "On Ising's model of ferromagnetism," in *Proc. Camb. Phil. Soc*, vol. 32, pp. 477–481, World Scientific, 1936.

[44] M. Mézard and A. Montanari, *Information, physics, and computation.* Oxford University Press, 2009.

[45] R. Monasson and R. Zecchina, "Entropy of the K-satisfiability problem," *Phys. Rev. Lett.*, vol. 76, pp. 3881–3885, May 1996.

[46] H.-O. Georgii, *Gibbs measures and phase transitions*, vol. 9. Walter de Gruyter, 2011.

[47] A. Montanari and G. Semerjian, "Rigorous inequalities between length and time scales in glassy systems," *Journal of statistical physics*, vol. 125, no. 1, pp. 23–54, 2006.

[48] J.-P. Bouchaud and G. Biroli, "On the Adam-Gibbs-Kirkpatrick-Thirumalai-Wolynes scenario for the viscosity increase in glasses," *The Journal of chemical physics*, vol. 121, p. 7347, 2004.

[49] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[50] R. W. Hamming, "Error detecting and error correcting codes," *Bell System technical journal*, vol. 29, no. 2, pp. 147–160, 1950.

[51] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Information Theory, IEEE Transactions on*, vol. 55, pp. 3051–3073, July 2009.

[52] R. Gallager, *Low-Density Parity-Check Codes.* PhD thesis, Massachusetts: M.I.T. Press, 1963.

[53] D. J. MacKay and R. M. Neal, "Near shannon limit performance of low density parity check codes," *Electronics letters*, vol. 32, no. 18, p. 1645, 1996.

[54] M. Sipser and D. A. Spielman, "Expander codes," *Information Theory, IEEE Transactions on*, vol. 42, no. 6, pp. 1710–1722, 1996.

[55] P. Oswald and A. Shokrollahi, "Capacity-achieving sequences for the erasure channel," *Information Theory, IEEE Transactions on*, vol. 48, no. 12, pp. 3017–3028, 2002.

[56] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *Information Theory, IEEE Transactions on*, vol. 51, no. 9, pp. 3221–3246, 2005.

[57] N. Macris, "Griffith–Kelly–Sherman correlation inequalities: A useful tool in the theory of error correcting codes," *Information Theory, IEEE Transactions on*, vol. 53, no. 2, pp. 664–683, 2007.

[58] S. Kudekar and N. Macris, "Sharp bounds for optimal decoding of low-density parity-check codes," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4635–4650, 2009.

[59] T. Richardson and R. L. Urbanke, *Modern coding theory*. Cambridge University Press, 2008.

[60] S. Korada, S. Kudekar, and N. Macris, "Exact solution for the conditional entropy of Poissonian LDPC codes over the binary erasure channel," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pp. 1016–1020, June 2007.

[61] S. Kudekar and N. Macris, "Decay of correlations for sparse graph error correcting codes," *SIAM Journal on Discrete Mathematics*, vol. 25, no. 2, pp. 956–988, 2011.

[62] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[63] R. G. Gallager, *Information theory and reliable communication*. Wiley, 1968.

[64] S. Kudekar, *Statistical physics methods for sparse graph codes*. PhD thesis, École polytechnique fédérale de Lausanne, 2009.

[65] T. J. Goblick, *Coding for discrete information source with a distortion measure*. PhD thesis, MIT, 1963.

[66] A. Viterbi and J. Omura, "Trellis encoding of memoryless discrete-time sources with a fidelity criterion," *Information Theory, IEEE Transactions on*, vol. 20, pp. 325–332, May 1974.

[67] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 58, no. 6, pp. 3309–3338, 2012.

[68] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *Information Theory, IEEE Transactions on*, vol. 56, pp. 1751–1768, April 2010.

[69] M. Wainwright, E. Maneva, and E. Martinian, "Lossy source compression using low-density generator matrix codes: Analysis and algorithms," *Information Theory, IEEE Transactions on*, vol. 56, pp. 1351–1368, March 2010.

[70] T. Filler and J. Fridrich, "Binary quantization using belief propagation with decimation over factor graphs of LDGM codes," in *Proc. Allerton Conf.*, 2007.

[71] D. Castanheira and A. Gameiro, "Lossy source coding using belief propagation and soft-decimation over LDGM codes," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, pp. 431–436, Sept 2010.

[72] A. Jimenez Felstrom and K. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *Information Theory, IEEE Transactions on*, vol. 45, pp. 2181–2191, Sep 1999.

[73] M. Lentmaier, A. Sridharan, D. Costello, and K. Zigangirov, "Iterative decoding threshold analysis for LDPC convolutional codes," *Information Theory, IEEE Transactions on*, vol. 56, pp. 5274–5289, Oct 2010.

[74] M. Lentmaier, A. Sridharan, K. Zigangirov, and D. Costello, "Terminated LDPC convolutional codes with thresholds close to capacity," in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pp. 1372–1376, Sept 2005.

[75] M. Lentmaier, D. G. M. Mitchell, G. P. Fettweis, and D. J. Costello, "Asymptotically regular LDPC codes with linear distance growth and thresholds close to capacity," in *Information Theory and Applications Workshop (ITA)*, pp. 1–8, January 2010.

[76] V. Aref and R. Urbanke, "Universal rateless codes from coupled LT codes," in *Information Theory Workshop (ITW), 2011 IEEE*, pp. 277–281, Oct 2011.

[77] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *Information Theory, IEEE Transactions on*, vol. 57, no. 2, pp. 803–834, 2011.

[78] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *Information Theory, IEEE Transactions on*, vol. 59, pp. 7761–7813, Dec 2013.

[79] A. Yedla, Y.-Y. Jian, P. Nguyen, and H. Pfister, "A simple proof of threshold saturation for coupled scalar recursions," in *Turbo Codes and Iterative Information Processing (ISTC), 2012 7th International Symposium on*, pp. 51–55, Aug 2012.

[80] S. Kumar, A. Young, N. Macris, and H. Pfister, "A proof of threshold saturation for irregular LDPC codes on BMS channels," in *Proc. of 50th Annual Allerton Conference on Communication, Control, and Computing (October 2012)*, 2012.

[81] S. Hassani, N. Macris, and R. Urbanke, "Coupled graphical models and their thresholds," in *Information Theory Workshop (ITW), 2010 IEEE*, pp. 1–5, Aug 2010.

[82] S. H. Hassani, N. Macris, and R. Urbanke, "Chains of mean-field models," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 02, p. P02011, 2012.

[83] S. Hamed Hassani, N. Macris, and R. Urbanke, "Threshold saturation in spatially coupled constraint satisfaction problems," *Journal of Statistical Physics*, vol. 150, no. 5, pp. 807–850, 2013.

[84] S. Kudekar and H. Pfister, "The effect of spatial coupling on compressive sensing," in *48th Annual Allerton Conference*, pp. 347 –353, Aug. 2010.

[85] D. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *Information Theory, IEEE Transactions on*, vol. 59, pp. 7434–7464, Nov 2013.

[86] S. Ciliberti and M. Mézard, "The theoretical capacity of the parity source coder," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 10, p. P10003, 2005.

[87] C. Moore and S. Mertens, *The Nature of Computation.* Oxford University Press, 2011.

[88] N. Macris and M. Vuffray, "The Bethe free energy allows to compute the conditional entropy of graphical code instances. a proof from the polymer expansion," *ArXiv e-prints*, Oct. 2013.

[89] V. Aref, N. Macris, and M. Vuffray, "Approaching the rate-distortion limit with spatial coupling, belief propagation and decimation," *ArXiv e-prints*, July 2013.

[90] M. Chertkov and V. Y. Chernyak, "Loop series for discrete statistical models on graphs," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, no. 06, p. P06009, 2006.

[91] D. C. Brydges, "A short course on cluster expansions," *Critical Phenomena, Random Systems, Gauge Theories*, vol. Les Houches, session XLIII, Part I, pp. 129–183, 1984.

[92] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *Information Theory, IEEE Transactions on*, vol. 51, no. 7, pp. 2282–2312, 2005.

Bibliography

[93] V. Chernyak and M. Chertkov, "Loop calculus and belief propagation for q-ary alphabet: Loop tower," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pp. 316–320, June 2007.

[94] J.-Q. Xiao and H. Zhou, "Partition function loop series for a general graphical model: free-energy corrections and message-passing equations," *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 42, p. 425001, 2011.

[95] G. D. Forney Jr. and P. Vontobel, "Partition functions of normal factor graphs," in *Proc. Information Theory and Applications Workshop, UC San Diego, La Jolla, CA, USA, Feb. 6-11*, 2011.

[96] R. Mori, "New understanding of the Bethe approximation and the replica method," *arXiv preprint arXiv:1303.2168*, 2013.

[97] J. Schiff, *Normal families.* Springer, 1993.

[98] J.-B. H. Urruty and C. Lemaréchal, *Fundamentals of convex analysis.* Springer, 2001.

[99] E. Sudderth, M. Wainwright, and A. Willsky, "Loop series and Bethe variational bounds in attractive graphical models," in *Advances in Neural Information Processing Systems 20* (J. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), pp. 1425–1432, Cambridge, MA: MIT Press, 2008.

[100] N. Ruozzi, "The Bethe partition function of log-supermodular graphical models," *ArXiv e-prints*, Feb. 2012.

[101] J. Mooij and H. Kappen, "Sufficient conditions for convergence of the sum–product algorithm," *Information Theory, IEEE Transactions on*, vol. 53, pp. 4422–4437, Dec 2007.

[102] B. D. McKay, "Subgraphs of random graphs with specified degrees," in *Proceedings of the International Congress of Mathematicians*, vol. 4, pp. 2489–2501, 2010.

[103] S. Ciliberti, M. Mézard, and R. Zecchina, "Message-passing algorithms for non-linear nodes and data compression," *Complexus*, vol. 3, no. 1-3, pp. 58–65, 2006.

[104] A. Giurgiu, N. Macris, and R. Urbanke, "And now to something completely different: Spatial coupling as a proof technique," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pp. 2443–2447, July 2013.

[105] M. Mézard and G. Parisi, "The Bethe lattice spin glass revisited," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 20, no. 2, pp. 217–233, 2001.

[106] S. Franz and M. Leone, "Replica bounds for optimization problems and diluted spin systems," *Journal of Statistical Physics*, vol. 111, no. 3-4, pp. 535–564, 2003.

[107] V. Aref, N. Macris, R. Urbanke, and M. Vuffray, "Lossy source coding via spatially coupled LDGM ensembles," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 373–377, July 2012.

[108] A. Iyengar, P. Siegel, R. Urbanke, and J. Wolf, "Windowed decoding of spatially coupled codes," *Information Theory, IEEE Transactions on*, vol. 59, pp. 2277–2292, April 2013.

[109] N. ul Hassan, A. Pusane, M. Lentmaier, G. Fettweis, and D. Costello, "Reduced complexity window decoding schedules for coupled LDPC codes," in *Information Theory Workshop (ITW), 2012 IEEE*, pp. 20–24, Sept 2012.

[110] F. Guerra and F. Toninelli, "The high temperature region of the Viana–Bray diluted spin glass model," *Journal of Statistical Physics*, vol. 115, no. 1-2, pp. 531–555, 2004.

[111] F. Ricci-Tersenghi and G. Semerjian, "On the cavity method for decimated random constraint satisfaction problems and the analysis of belief propagation guided decimation algorithms," *Journal of Statistical Mechanics: Theory and Experiment*, vol. P09001, 2009.

[112] A. Coja-Oghlan, "On belief propagation guided decimation for random K-SAT," in *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 957–966, SIAM, 2011.

[113] H. Wilf, *Generatingfunctionology*. Ak Peters Series, A K Peters, 2006.

[114] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2010.

[115] P. Billingsley, *Probability and measure*. John Wiley & Sons, 1995.

MARC VUFFRAY

Address: Rue du Centre 23, 1025 St-Sulpice, Switzerland
Tel: (+41) 21 691 88 31
Email: marc.vuffray@epfl.ch
Date of Birth: August 9th, 1984
Nationality: Swiss Citizen

<div style="border:1px solid">**Master of Science in Physics (EPFL) + Ph.D. in Computer Science (EPFL)**</div>

Education

2009 – 2013	**Doctoral studies at the school of computer and communication science of EPFL** *Study of diluted spin glasses and applications to communication problems* *Thesis directors: R. Ürbanke, N. Macris*
2003 – 2007	**Master of Science in Physics (EPFL), GPA 5.34/6** *Master Thesis: On the proof of existence and unicity of radially symmetric ground states* *of the nonlinear and nonlocal Schrödinger-Newton Equation*

Other Research and Academic Experience

May – July 2013	**Scientific collaboration with Dr. M. Chertkov (T-4, Los Alamos National Laboratory)** *Study of nucleation processes in spatially-coupled low density parity check codes*
2008 – 2009	**Scientific assistant for Dr. J. Stubbe (Mathematics, EPFL)** *Study of excited states of a nonlinear and nonlocal Schrödinger-Newton Equation*
2006 – 2013	**Scientific collaboration with Professor Emeritus Ph. Choquard (Physics, EPFL)** *Performing numerical simulations for a class of problems in fluid dynamics*
2005 – 2007	**Conference manager** *In the EPFL physics students association*

List of peer reviewed publications

- **Approaching the Rate-Distortion Limit with Spatial Coupling, Belief Propagation and Decimation**
  V. Aref, N. Macris, M. Vuffray, submitted *to Transaction on Information Theory*, 2014
  Not yet published, preprint available on arXiv:1307.5210

- **The Inviscid, Compressible and Rotational, 2D Isotropic Burgers and Pressureless Euler-Coriolis Fluids: Theory and Illustration**
  Ph. Choquard, M. Vuffray, submitted to *Physica D,* 2013
  Not yet published, preprint available on arXiv:1310.1337

- **The Bethe free energy allows to compute the conditional entropy of graphical codes instances: A proof from the polymer expansion**
  N. Macris, M. Vuffray, submitted to *Transaction on Information Theory*, 2013
  Not yet published, preprint available on arXiv:1310.1294

- **Approaching the Rate-Distortion Limit by Spatial Coupling with Belief Propagation and Decimation**
  Aref, V.; Macris, N.; Vuffray, M., *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pp.1177-1181, 7-12 July 2013

- **Lossy Source Coding via Spatially Coupled LDGM Ensembles**
  Aref, V.; Macris, N.; Urbanke, R.; Vuffray, M., *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp.373-377, 1-6 July 2012

- **Beyond the Bethe Free Energy of LDPC Codes via Polymer Expansions**
  Macris, N.; Vuffray, M, *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on* , vol., no., pp.2331,2335, 1-6 July 2012

- **Polymer Expansions for Cycle LDPC Codes**
  N. Macris, M. Vuffray, *Int. Zurich Seminar on Communications Proceedings (IZS)*, Pages 79-82, February 29 - March 2, 2012

- **Bound states of the Schrödinger-Newton model in low dimensions**
  J. Stubbe, M. Vuffray, *Nonlinear Analysis: Theory, Methods & Applications*, Volume 73, Issue 10, Pages 3171-3178, 2010

- **Stationary Solutions of the Schrödinger-Newton Model – An ODE Approach**
  Ph. Choquard, J. Stubbe, M. Vuffray, *Differential and Integral Equations*, vol. 21, no 7-8, p. 665-679, 2008

Contributed Talks

- **International Symposium on Information Theory**, 2013, Istanbul, Turkey

- **Joint Conference on Communications and Coding**, 2013, Holzgau, Austria

- **International Symposium on Information Theory**, 2012, Boston, USA

- **International Zurich Seminar on Communications**, 2012, Zürich, Switzerland

- **Bridging Statistical Physics and Optimization, Inference and Learning**, 2012, Les Houches, France

Other Conferences Attended

- **Information Theory Workshop**, 2012, Lausanne, Switzerland

- **Statistical Physics of Complexity, Optimization and Biological Information**, 2010, Orsay, France

- **Statistical Physics of Complexity, Optimization and Systems Biology**, 2010, Les Houches, France

Teaching Experience

Teaching assistant from 2007 to 20013 for undergraduate courses of Physics, Calculus, Probability and Quantum Information Theory

Languages

- English – Fluent

- French – Native Language

- German – Good

Computer Skills

Computer programming: C++, Basic, Fortran 90-95, Mathematica

Awards

- Mathematics prize, "Fond Jules Bezencenet", 2003

- Physics prize, "Département de la formation de la jeunesse", 2003

Other relevant experience

Creation and development of Visualis Physics (2003 - 2004)
A series of interactive simulation software for teaching undergraduate physics
Website: www.visualis-physics.com/en