Statistical Physics Methods for Sparse Graph Codes

THÈSE N^{O} 4442 (2009)

PRÉSENTÉE LE 15 JUILLET 2009

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

Shrinivas KUDEKAR

acceptée sur proposition du jury:

Prof. B. Rimoldi, président du jury
Prof. R. Urbanke, Dr N. Macris, directeurs de thèse
Prof. M. Chertkov, rapporteur
Prof. D. Shah, rapporteur
Prof. M. A. Shokrollahi, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2009

Abstract

This thesis deals with the asymptotic analysis of coding systems based on sparse graph codes. The goal of this work is to analyze the decoder performance when transmitting over a general binary-input memoryless symmetric-output (BMS) channel. We consider the two most fundamental decoders, the optimal maximum a posteriori (MAP) decoder and the sub-optimal belief propagation (BP) decoder. The BP decoder has low-complexity and its performance analysis is, hence, of great interest. The MAP decoder, on the other hand, is computationally expensive. However, the MAP decoder analysis provides fundamental limits on the code performance. As a result, the MAP-decoding analysis is important in designing codes which achieve the ultimate Shannon limit.

It would be fair to say that, over the binary erasure channel (BEC), the performance of the MAP and BP decoder has been thoroughly understood. However, much less is known in the case of transmission over general BMS channels. The combinatorial methods used for analyzing the case of BEC do not extend easily to the general case.

The main goal of this thesis is to advance the analysis in the case of transmission over general BMS channels. To do this, we use the recent convergence of statistical physics and coding theory. Sparse graph codes can be mapped into appropriate statistical physics spin-glass models. This allows us to use sophisticated methods from rigorous statistical mechanics like the correlation inequalities, interpolation method and cluster expansions for the purpose of our analysis. One of the main results of this thesis is that in some regimes of noise, the BP decoder is optimal for a typical code in an ensemble of codes. This result is a pleasing extension of the same result for the case of BEC. An important consequence of our results is that the heuristic predictions of the replica and cavity methods of spin-glass theory are correct in the realm of sparse graph codes.

Keywords: Low-Density Parity-Check Codes, Low-Density Generator-Matrix Codes, Maximum A Posteriori Decoder, Belief Propagation Decoder, Statistical Physics, Spin-Glass.

Résumé

Le sujet de cette thèse concerne l'analyse asymptotique de systèmes de codages basés sur les graphes dilués. Le but de ce travail est l'analyse de la performance du décodage lors de la transmission à travers un canal symétrique sans mémoire à entrées binaires (BMS). Nous considérons les décodeurs les plus fondamentaux, le décodeur optimal maximum a posteriori (MAP) et le décodeur sous-optimal de propagation des messages (BP). D'une part, le décodeur BP possède une faible complexité et son analyse est donc d'un grand intérêt. D'autre part, bien que le décodeur MAP possède une grande complexité, celui-ci donne les limites fondamentales de la performance du code considéré. Par conséquent son analyse est importante pour construire les codes qui atteignent la limite de Shannon.

Il est justifié de dire que sur le canal binaire à effacement (BEC) les performances des décodeurs MAP et BP sont bien comprises. Par contre, sur les canaux généraux BMS la situation est beaucoup moins bien comprise. En effet les méthodes combinatoires utilisées pour l'analyse du BEC ne se généralisent pas aisément au cas général.

Le but principal de cette thèse est de faire des progrès dans l'analyse du cas de la transmission à travers des canaux BMS généraux. Pour cela nous utilisons la connexion récente entre la physique statistique et le codage. Les codes correcteurs d'erreurs basés sur les graphes dilués peuvent être transformés en modèles équivalents de physique statistique des verres de spins. Ceci nous permet d'utiliser des méthodes sophistiquées provenant de la mécanique statistique rigoureuse telles que, les inégalités de corrélations, la méthode d'interpolation et le développement en clusters. L'un des résultats principaux de cette thèse est que, dans certains régimes de bruit, le décodeur BP est optimal pour un code typique dans l'ensemble considéré. Cela constitue une extension du même résultat obtenu précédemment sur le BEC. Une conséquence importante de nos résultats est aussi que les prédictions heuristiques de la théorie des verres de spins obtenues par la méthode des répliques et de la cavité, sont correctes en ce qui concerne les codes correcteurs sur des graphes dilués.

Mots clés: Low-Density Parity-Check Codes, Low-Density Generator-Matrix Codes, Décodeur Maximum A Posteriori, Décodeur Belief Propagation,

Physique Statistique, Verres de spins.

Acktionary

SNF: 1) Swiss National Foundation. Use in a sentence – I gratefully acknowledge the money provided by the Swiss National Foundation for funding my stay in EPFL.

Rüdiger Urbanke: 1) Person that eats, walks and thinks very, very fast. 2) Similar to the Road-Runner. 3) Supremely optimistic person. 4) Very creative and could have pursued an alternate career in film directing. Use in a sentence – I have learned so much about so many things from Rüdiger Urbanke, that I can thank him only for a very small subset. Rüdiger has truly been a friend, philosopher and guide for me during my stay at EPFL. Working with Rüdiger has been a pleasure. I am grateful to him for being patient and tolerating my statements starting with "But..." during our discussions. But ☺...for every hour I put into solving our problems, Rüdiger put two. His tremendous insights and persistence have been crucial to our understanding of the problems. Rüdiger's emphasis on good writing and presentation is legendary. Without his help, crucial parts of my thesis would have been unreadable. I hope that I have been able to grasp his research philosophy: work on important problems, think out of the box and never be afraid to extensively simulate your ideas. More importantly, I really hope I have been able to soak in his optimism, fun-loving nature, kindness and generosity of heart.

Nicolas Macris: 1) Person who looks like he is lost in space and time, but is *actually* thinking about space-time. 2) A person who believes ergodic theory also applies in life: "If any event has a non-zero probability, then it will happen – like sitting on a plane to an incorrect destination". 3) An unbridled enthusiasm for science. 4) Another name for a person who always believes the glass is half-full. Use in a sentence – I am deeply indebted to Nicolas for introducing and teaching me the subject of statistical physics. Working with Nicolas has been a delight. My thesis work would not have been possible without his initiation, insights, persistence and patience. Nicolas vast mathematical and physics background has helped me in many occasions to streamline my random thoughts in a precise manner. I sincerely hope I have been able to inculcate his meticulous style of research.

Muriel Bardet: 1) Superwoman. 2) Mom of the lab. Use in sentence – I seriously doubt if our lab could function without Muriel. I do not understand

how someone manages so many things yet remain always calm and smiling. Muriel made my stay in Lausanne and EPFL very very smooth. Without her diplomacy I do not think I would have got a nice apartment to stay in Lausanne. There are probably hundred of other things for which I am very grateful to Muriel. I will miss most those morning chatting sessions with her on Formula 1 and Roger Federer.

Damir: 1) Italian system administrator. 2) Person who can fix your computer and bake delicious cakes at the same time. 3) Loves to curse in Italian ☺. Use in a sentence – I am grateful to Damir for providing fantastic computing facilities in the lab. He was very patient when we tested his simulation system, `qsub`, by submitting about a million jobs. I am also thankful to him for maintaining the "/ipg/..." server. It made some gloomy days in Lausanne very enjoyable.

Dinkar and Satish: 1) Two stooges ☺. Use in a sentence – I have profited immensely from technical discussions, suggestions and professional advice provided to me by Dinkar and Satish. I have bothered them with many questions and they have always been patient in answering them. I must admit that a lot of things I have learned in my PhD were the result of our interactions. Some of the most enjoyable moments in my graduate life has been the weekend coffee conversations with Dinkar and Satish (and previously with Sanket and Vish). My stay in Lausanne – in and out of the office – was fun-filled and enjoyable mainly due to Dinkar and Satish. Dinkar's bollywood knowledge and wit has been a source of entertainment and will be sorely missed. Dinkar has probably cooked for me more than anybody else and because of this he is also my "Annadatta".

Potpourri: 1) Medley of people enriching life. Use in a sentence – Emre Telatar has given me advice on many occasions and topics. I am very grateful to him. He also answered all my doubts/questions in Information theory. I would also like to thank Yvonne and Francoise for their help and warmth.

Many thanks to my thesis committee, Prof. Devavrat Shah, Dr. Misha Chertkov and Prof. Amin Shokrollahi for carefully going through my thesis and providing me useful comments. I take this opportunity to thank Prof. Bixio Rimoldi for agreeing to be the president of my thesis committee.

Life in Lausanne was made very easy and enjoyable thanks to the company of the Indian group – Sibi & Reshmi, Sanket & Seema, Rajesh, Dinkar, Vish, Satish, Saket, Abhishek, Vasu & Anmol. I will always cherish the great outings I enjoyed with them and also the fun-filled evening dinners where I was allowed to vent. I am thankful to all the labmates past and present for providing a warm and friendly atmosphere in the lab.

A special thanks to Jérémie for many things. I highly enjoyed working with him for our TA duties: interesting conversations; the cursing – "Man, you are so stupid!"; the facebook surfing; coming up with questions; making fun of Rüdiger ☺; imitations. The annual "movie production" of our lab would not have been half as much fun without the craziness and energy of Jérémie, Marius and Rüdiger.

Last but not the least there is my family (parents + sister + grandparents). I can say without exaggeration: without them I would be nothing.

Contents

1 Introduction

The title "Statistical Physics Methods for Sparse Graph Codes" is in reference to the recent convergence of communication science and statistical physics. In this thesis we focus on the performance analysis of sparse graph codes when transmitting over binary-input memoryless symmetric-output (BMS) channels (see Section 1.2 for a definition) using tools from statistical physics.

Low-density parity-check (LDPC) codes based on sparse graphs (see Section 1.1) have emerged as a new chapter in the theory of error correcting codes. This is largely because they are amenable to low-complexity decoding and, at the same time, have good performance (measured as the gap to Shannon's capacity) [1], [2]. After their invention by Gallager in 1962 in his thesis [3], LDPC codes were largely forgotten due to a lack of computational resources to implement them. In 1993, Berrou, Glavieux and Thitimajshima provided an impetus to the search of low-complexity decoding by inventing Turbo codes [4]. They showed that Turbo codes achieved performance very close to the Shannon limit under low-complexity decoding.

Shortly afterwards, LDPC codes were re-invented by MacKay and Neal [5], [6] and independently by Spielman in his thesis on LDPC codes based on expander graphs [7]. One of the two most notable outcomes of this resurrection was the discovery of irregular LDPC codes (see Section 1.1) by Luby, Mitzenmacher, Shokrollahi and Spielman [8], which by optimizing the degrees, could approach the capacity of the binary erasure channel (BEC). The other notable outcome was the density evolution technique of Richardson and Urbanke [9]. This technique allowed to design codes which empirically approached the Shannon limit on the binary-input additive white-gaussian noise channel (BIAWGNC)[1] [1].

It was shown by Wiberg in his thesis [10], and independently by Tanner in

[1]See Section 1.2 for a definition of the BEC and BIAWGNC.

[11], that both Turbo codes and LDPC codes fall under the umbrella of codes based on sparse graphs. They also showed that their low-complexity decoding algorithm was an instance of a general *sum-product* rule, and is equivalently known as the *belief propagation* (BP) decoder. A summary of this development can be found in [12]. We would like to mention at this point that the BP rules have multiple and independent origins: they were suggested by Kim and Pearl in [13], [14] to approximately compute the a posteriori marginals in Bayesian inference networks in the field of artificial intelligence; they also appeared as the Bethe-Peierls equations for computing partition functions in the field of statistical physics [15].

The analysis of codes based on sparse graphs has evolved in two directions: (i) analysis of the *optimal decoding* (decoder minimizing the average bit-error rate) and (ii) analysis of the low-complexity (sub-optimal) BP decoder. There is an obvious interest in analyzing the BP decoder, as it is the choice for any practical decoding system. However, there is also an interest in analyzing the performance of the optimal decoder, since it makes it clear whether attempts to increase the performance should focus on finding better codes or better decoding algorithms.

It is probably fair to say that the analysis of the optimal decoder has lagged the BP decoder analysis. The basic techniques for analyzing the optimal decoder has not evolved much from its inception in Gallager's thesis. Although much effort has gone into refining the techniques and applying them to a large set of communication scenarios [16], [17].

The BP decoder analysis of sparse graph codes has been the main focus over the last decade. In the case of transmission over the BEC, the BP decoder has been thoroughly analyzed [18], [19], [20]. In fact there are code constructions which achieve the Shannon limit of the BEC under BP decoding [21], [22]. We refer to the recent book [23] for the state of the art of this general theory. However much less is known for the case of transmission over more general channels. One of the goals of this thesis is to advance the analysis of the BP decoder in the more general case.

Recently, Measson, Montanari and Urbanke in [24] have demonstrated the optimality of the BP decoder, with high probability, for some regimes of channel erasure probability, when transmitting over the BEC. The codes they use for transmission belong to a fairly general class of LDPC code ensembles. This is one of the few instances, where it is known that BP rules perform exact inference, with high probability, even in the presence of loops in the graphical representation! It would be desirable to have such a statement in the case of general channels. For this, the combinatorial techniques, employed in the analysis of the BEC, do not seem to be useful and radical new ideas seem to be necessary.

Sourlas in his work in [25], showed that the decoding problem of sparse graph codes can be mapped to the computation of *local magnetizations or magnetic fields* of a *spin-glass* system in statistical mechanics. Building on the work of Sourlas, Kanter and Saad in [26], Murayama, Kabashima, Saad

and Vicente in [27] and Franz, Leone, Montanari and Ricci-Tersenghi in [28] used methods from statistical physics to study both, optimal and BP decoding LDPC codes.

Recently, Montanari in [29] showed that predictions of statistical physics are tight bounds on the performance of the optimal decoder, when transmitting over general BMS channels. The methods used in [29] have their origins in the analysis of mean-field models of statistical physics of disordered systems. These methods were pioneered by Guerra, Toninelli and Talagrand in [30], [31], [32] for the analysis of the Sherrington-Kirkpatrick model of statistical physics.

In this thesis our main focus is to use tools from statistical physics for performance analysis of sparse graph codes when transmitting over general BMS channels. We study various quantities appearing in the optimal-decoder analysis. As a consequence of our analysis we show that the BP decoder does exact inference, with high probability, for appropriate regimes of channel noise value. We employ tools from statistical physics for the purpose of our analysis.

In the next section we define codes based on sparse graphs. We mention the two most popular codes based on sparse graphs, namely, the low-density parity-check codes and the low-density generator-matrix codes. This thesis focuses solely on the performance analysis of these two codes. We also describe the channels over which the transmission takes place. In Section 1.3, we describe the optimal MAP decoder. The sub-optimal BP decoder is outlined in Section 1.4. In Section 1.5 we mention the associated tools for performance analysis. A brief history of the performance analysis of sparse graph codes in given in Section 1.6. We show the connection to statistical physics in Section 1.7. Finally, we end the chapter by delineating the contribution and organization of this thesis.

1.1 Codes based on Sparse Graphs

It is our aim to study the performance of block codes defined on sparse graphs. We have k bits of information. We transform them via encoding into a codeword. Throughout this thesis, n will denote the block-length of the code. The two families of codes we consider are low-density parity-check codes and low-density generator-matrix codes. Most of our results deal with ensembles of codes rather than a fixed code. However, wherever applicable, we show that our results hold also for any fixed code. The codes we consider are defined over $\mathbb{F}_2 = \{0, 1\}$. Let us now describe these two code families in more detail.

1.1.1 Low-Density Parity-Check (LDPC) Codes

A parity-check code, C, is defined by a parity-check matrix H of dimension $m \times n$. Each entry of H belongs to $\{0, 1\}$. Here, m represents the number of parity-check constraints on the n code-bits. The codewords of C are all the

n length binary tuples $\underline{x} = (x_1, x_2, \ldots, x_n)$ which satisfy $\mathsf{H}\underline{x} = \underline{0}$. Thus the code is the null space of the parity-check matrix H. The rate of the code is defined to be the dimension of the null space normalized by the block-length. A related quantity is the design rate of the code, denoted by $\mathtt{R}$. It is given by $\mathtt{R} = 1 - \frac{m}{n}$. Note that the rate is always at least $\mathtt{R}$. The inequality is strict if some of the constraints are linearly dependent. Throughout this thesis we will call x_i the "code-bit" i. An example of a parity-check matrix code is shown in Example 1.1.

Example 1.1 (Parity-Check Code)**.** *Consider the parity-check matrix*

$$\mathsf{H} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}. \tag{1.1}$$

We have $n = 7$ and $m = 3$ (i.e., the block-length is 7 and there are 3 parity-check constraints). The design rate of the code is $\mathtt{R} = (7-3)/7 = 4/7$. The code consists of all $7-$tuples $\underline{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ satisfying the three equations

$$\begin{aligned} x_1 + x_2 + x_4 + x_5 &= 0, \\ x_1 + x_3 + x_4 + x_6 &= 0, \\ x_2 + x_3 + x_4 + x_7 &= 0. \end{aligned}$$

It is not hard to see that the above set of equations are linearly independent over the field $\mathbb{F}_2$. Thus the rate of the code is equal to the design rate.

We use the standard Tanner graph representation [23], [33] to visualize the parity-check code. In this representation the code-bits are represented by a circle and the parity-check constraints are represented by squares. The circles are usually called *variable* nodes and the squares are called *check nodes*. The Tanner graph is a bipartite graph with two sets of nodes; one set, $\mathfrak{V}$, consists of the variable nodes and the other set, $\mathfrak{C}$, consists of check nodes. There is an edge between a variable node $i \in \mathfrak{V}$ and a check node $a \in \mathfrak{C}$, if the corresponding code-bit i participates in the parity-check constraint a. Figure 1.1 shows the Tanner graph corresponding to the parity-check matrix of (1.1).

Low-Density parity-check codes are parity-check codes which have a *sparse* Tanner graph representation [23]. By sparse we mean that the number of edges in the Tanner graph is proportional to n. From the Tanner graph one can deduce a degree distribution of the variable nodes and a degree distribution of the check nodes. The variable node degree distribution is given by the function $\Lambda(x) \triangleq \sum_i \Lambda_i x^i$, where Λ_i denotes the fraction of nodes with degree i. Similarly, the check node degree distribution is given by $P(x) \triangleq \sum P_i x^i$. It is standard practice to represent the code by its node degree distribution pair $(\Lambda(x), P(x))$. We will also be interested in the degree distribution from the

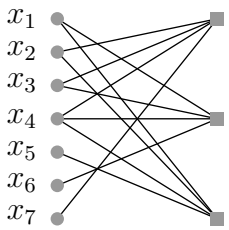

Figure 1.1: The Tanner graph corresponding to the code in Example 1.1. Here $n = 7$ and the (design) rate is equal to $\frac{4}{7}$. The variable node degree distribution is given by $\Lambda(x) = \frac{3}{7}x + \frac{3}{7}x^2 + \frac{1}{7}x^3$. The check node degree distribution is given by $P(x) = x^4$. The average variable node degree is equal to $\mathtt{l}_{\text{avg}} = \Lambda'(1) = \frac{12}{7}$ and average check node degree is equal to $\mathtt{r}_{\text{avg}} = P'(1) = 4$. The edge perspective degree distributions are given by $\lambda(x) = \frac{1}{4} + \frac{1}{2}x + \frac{1}{4}x^2$ and $\rho(x) = x^3$.

edge perspective. We denote by $\lambda(x) \triangleq \sum_i \lambda_i x^{i-1}$ the edge degree distribution of variable nodes. Here λ_i denotes the fraction of edges attached to variable nodes of degree i. Similarly, we denote by $\rho(x) \triangleq \sum_i \rho_i x^{i-1}$ the edge degree distribution of the check nodes. Here ρ_i gives the fraction of edges connected to check nodes of degree i. Clearly, $\mathtt{l}_{\text{avg}} \triangleq \Lambda'(1)$ and $\mathtt{r}_{\text{avg}} \triangleq P'(1)$ gives the average variable node degree and the average check node degree respectively. Since $n\mathtt{l}_{\text{avg}} = m\mathtt{r}_{\text{avg}} =$ number of edges, the design rate is also given by $\mathtt{R} = 1 - \frac{m}{n} = 1 - \frac{\mathtt{l}_{\text{avg}}}{\mathtt{r}_{\text{avg}}}$. From the definitions above we deduce $\lambda_i = \frac{i\Lambda_i}{\Lambda'(1)}$ and $\rho_i = \frac{iP_i}{P'(1)}$. We will use, whenever convenient, either $(\Lambda(x), P(x))$ or $(\lambda(x), \rho(x))$ to denote the code. We let $\mathtt{l}_{\max}$ and $\mathtt{r}_{\max}$ denote the maximum variable node and check node degrees. In most of our results, we consider codes with bounded $\mathtt{l}_{\max}$ and $\mathtt{r}_{\max}$.

We now define an ensemble of codes, $\mathcal{C}$, using the variable node and check node degree distributions. Thus we define an ensemble of codes of degree distribution $(\Lambda(x), P(x))$ as follows.

Definition 1.1 (Standard Ensemble)**.** *A code from the standard ensemble $(n, \Lambda(x), P(x))$ is constructed as follows. Let the maximum variable node degree be given by $\mathtt{l}_{max}$ and the maximum check node degree be given by $\mathtt{r}_{max}$. Associate i sockets to each of $n\Lambda_i$ variable nodes of degree i (choose n such that $n\Lambda_i$ is an integer for all i), for all $1 \leq i \leq \mathtt{l}_{max}$. Similarly, associate k sockets for mP_k check nodes (choose m such that mP_k is an integer for all k) of degree k. Number the sockets on both sides. Consider a random permutation, π of $n\,\mathtt{l}_{\text{avg}}$ objects and connect socket number i on the variable node side to socket number $\pi(i)$ on the check node side. For each permutation we get a code in the ensemble and we associate a uniform probability over all the permutations.*

The most popular example of LDPC code ensembles, is the regular LDPC code ensemble, for which $\Lambda(x) = x^{\mathtt{l}}$ and $P(x) = x^{\mathtt{r}}$. In words, all the variable nodes have degree $\mathtt{l}$ and all the check nodes have degree $\mathtt{r}$. These codes are

also known as Gallager codes, as they first apppeared in his thesis [3]. Irregular LDPC code ensembles refer to codes which do not have the same variable node degree and the same check node degree.

1.1.2 Low-Density Generator-Matrix (LDGM) Codes

In the previous section we defined a code via the null space of a matrix. Let us now take a dual approach and define a code by the image space of a matrix. More precisely, suppose that we have m information bits. Consider a matrix G of dimension $m \times n$. Let $\underline{u} \in \{0,1\}^m$ denote the vector consisting of the m information bits, $u_1, u_2, \ldots, u_m$. We say that $\underline{u}$ is the information word. Then the codeword, generated by the matrix G, for the information word $\underline{u}$ is given by $\underline{x}^\top = \underline{u}\mathsf{G}$. Thus the code, C, is the image set of the matrix G. We say that C is a generator-matrix code. As above, we represent the code via a Tanner graph. Consult Figure 1.2 for an illustration. In this case, the variable nodes (circles) on the left, denote the information bits. The variable nodes on the right (attached to check nodes (squares)) denote the code-bits which are transmitted over the channel. An edge is present between a variable node, representing an information bit, and a check node if the corresponding information bit is used in the generation of the code-bit. The design rate is given by $\mathsf{R} = \frac{m}{n}$. Contrary to parity-check codes, the design rate of a generator-matrix code is an upper bound to the actual rate. Similar to LDPC codes, *low-density generator-matrix* codes are generator-matrix codes having a Tanner graph representation which is sparse. Again we can specify $\Lambda(x), P(x), \lambda(x), \rho(x)$ in the same manner as for LDPC codes. The standard ensemble of LDGM codes is also defined in an equivalent way as Definition 1.1.

Example 1.2 (Generator-Matrix Code)**.** *Consider the generator-matrix*

$$\mathsf{G} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}. \tag{1.2}$$

The block-length is 7*, the dimension of the code is* 4 *and the design rate is equal to* $4/7$*. The Tanner graph representation is shown in Figure 1.2.*

1.1.3 Notation

We fix the following notation for both LDPC and LDGM codes. Variable nodes (resp. check nodes) will be denoted by i, j (resp. a, b). The set of variable nodes (resp. check nodes) is denoted by $\mathfrak{V}$ (resp. $\mathfrak{C}$). The *neighborhood* of a node v (variable node or a check node), defined to be the set of nodes which have an edge with the node v in the Tanner graph, is denoted by ∂v. We define the *graph distance* between any two nodes (variable or check) i, j, denoted by $\mathrm{dist}(i,j)$, to be the minimum number of edges between the nodes i, j. For

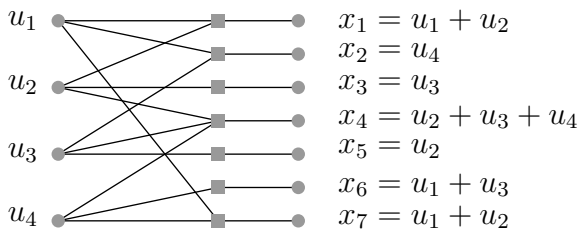

Figure 1.2: The variable nodes on the left represent the information bits and the check nodes on the right represent the code-bits. The degree distributions from the node perspective are given by $\Lambda(x) = x^3, P(x) = \frac{3}{7}x + \frac{3}{7}x^2 + \frac{1}{7}x^3$. The average variable node degree is equal to $\Lambda'(1) = 3$ and the average check node degree is equal to $P'(1) = 12/7$. The edge perspective degree distributions are given by $\lambda(x) = x^2, \rho(x) = \frac{1}{4} + \frac{1}{2}x + \frac{1}{4}x^2$.

any variable node i, for d even, denote by $N_d(i)$ the neighborhood of depth d around the variable node i. More precisely, $N_d(i)$ contains all variable nodes j such that $\text{dist}(i, j) \leq d$. Similarly for any check node a, for d odd, $N_d(a)$ denotes the neighborhood of depth d around the check node a and contains all variable nodes j such that $\text{dist}(a, j) \leq d$. We use $\mathring{N}_d(i)$ to denote the set of check nodes at a distance $d + 1$ from the root node i. We call $\mathring{N}_d(i)$ as the boundary of the node i. Figure 1.3 shows a neighborhood of depth 3 around a check node a and its boundary in an LDGM code. It also shows a neighborhood of depth 2 around a variable node i and its boundary of check nodes in a LDPC code.

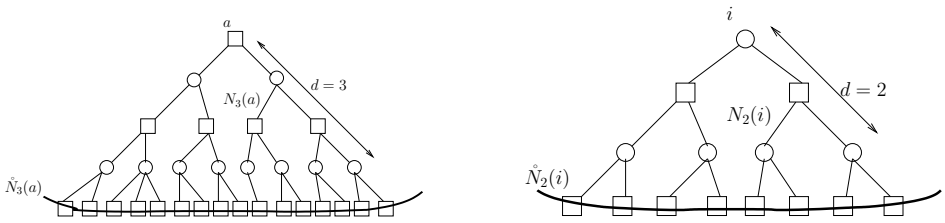

Figure 1.3: The figure on the left shows a neighborhood of depth 3, denoted by $N_3(a)$, around the check node a in an LDGM code. The boundary of check nodes is also shown and denoted by $\mathring{N}_3(a)$. The figure on the right shows a neighborhood of depth 2, denoted by $N_2(a)$, around the variable node i in an LDPC code. The boundary of check nodes is also shown and denoted by $\mathring{N}_2(i)$.

1.2 Channel Models

We consider noisy communication, where we transmit information over a channel by encoding it into a codeword of block-length n and the channel introduces random errors according to a certain model. We consider binary-input channels with transition probability between input and output given by $p_{Y|X}(y \mid x)$. Although the code is defined over $\mathbb{F}_2$, we map $0 \to +1$ and $1 \to -1$ before transmitting the symbol on the channel. Thus $x \in \{-1, +1\}$. The channels we consider are memoryless and without feedback, which implies

$$p_{\underline{Y}|\underline{X}}(\underline{y} \mid \underline{x}) = \prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid x_i) \quad \text{(memoryless and no feedback).} \tag{1.3}$$

Let us now review the three most fundamental channels which we will consider throughout this thesis:

Example 1.3 (Binary Erasure Channel – BEC(ϵ))**.** *The input alphabet is $x \in \{-1, +1\}$ and the output alphabet is $y \in \{-1, *, +1\}$, where $*$ denotes the erasure symbol. The channel transition probability is given by $p_{Y|X}(+1 \mid -1) = p_{Y|X}(-1 \mid +1) = 0$, $p_{Y|X}(+1 \mid +1) = p_{Y|X}(-1 \mid -1) = 1 - \epsilon$ and $p_{Y|X}(* \mid -1) = p_{Y|X}(* \mid +1) = \epsilon$, where ϵ is the erasure probability, $\epsilon \in [0, 1]$.*

Example 1.4 (Binary Symmetric Channel – BSC(ϵ))**.** *The input alphabet is $x \in \{-1, +1\}$ and the output alphabet is $y \in \{-1, +1\}$. The channel transition probability is given by $p_{Y|X}(+1 \mid -1) = p_{Y|X}(-1 \mid +1) = \epsilon$ and $p_{Y|X}(+1 \mid +1) = p_{Y|X}(-1 \mid -1) = 1 - \epsilon$, where ϵ is the flip probability, $\epsilon \in [0, \frac{1}{2}]$.*

Example 1.5 (Binary-Input Additive White Gaussian Noise Channel – BIAWGNC(ϵ))**.** *The input alphabet is $x \in \{-1, +1\}$ and the output alphabet is $y \in \mathbb{R}$. The channel transition probability is given by $p_{Y|X}(y \mid x) = \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{(y-x)^2}{2\epsilon^2}}$ where ϵ^2 is the noise variance, $\epsilon \in [0, +\infty)$.*

All three channels above are parametrized by a single scalar value ϵ. In general we will denote a channel by BMS(ϵ) where the scalar parameter is in one-to-one correspondence with the channel entropy (equals the single letter entropy, $H(X \mid Y)$).

Definition 1.2 (Channel Symmetry)**.** *A binary-input channel with transition probability $p_{Y|X}(y \mid x)$ is* symmetric *if*

$$p_{Y|X}(y \mid x) = p_{Y|X}(-y \mid -x).$$

All three channels presented above satisfy the channel symmetry. An important consequence of channel symmetry is that the uniform distribution over the input alphabet achieves the capacity. Another consequence of channel symmetry (coupled with the fact that the code is a vector space) is that for the purpose of performance analysis of the decoders we consider (see next section), one can assume without loss of generality that the all-one codeword is transmitted.

Definition 1.3 (Channel Loglikelihood Ratio)**.** *For any binary-input channel, the loglikelihood ratio is given by*

$$l(y) = \ln \frac{p_{Y|X}(y \mid +1)}{p_{Y|X}(y \mid -1)}.$$

Since the output of the channel y is random, the loglikelihood ratio, called henceforth LLR, is also a random variable. We denote the distribution of l, under the assumption that $+1$ is transmitted, by $c_L(l)$. In the literature [23], this is known as the $L-$density. Another representation of the densities is the $G-$domain representation. The $G-$domain representation of $c_L(l)$ is given by $g(l) = (\mathfrak{H}(l), \ln\coth|l/2|)$, where $\mathfrak{H}(l) = +1$ if $l > 0$ and $\mathfrak{H}(l) = -1$ if $l < 0$ and is equal to $+1$ or -1 with equal probability when $l = 0$ (see [23] for more details). For the three channels above we have

$$\begin{aligned}
c_L(l) &= \epsilon\delta_0(l) + (1-\epsilon)\delta_\infty(l), && \text{BEC}(\epsilon)\\
c_L(l) &= \epsilon\delta_{\ln\frac{\epsilon}{1-\epsilon}}(l) + (1-\epsilon)\delta_{\ln\frac{1-\epsilon}{\epsilon}}(l), && \text{BSC}(\epsilon)\\
c_L(l) &= \sqrt{\frac{1}{8\pi\epsilon^{-2}}}e^{-\frac{(l-2\epsilon^{-2})^2}{8\epsilon^{-2}}}. && \text{BIAWGNC}(\epsilon)
\end{aligned}$$

It is not difficult to show the following fact about the channel LLR [23].

Fact 1.1. *Consider any binary-input symmetric channel with transition probability given by $p_{Y|X}(y \mid x)$. Let $c_L(l)$ denote the distribution of the output* LLR *given that $+1$ is transmitted. Then $c_L(-l) = e^{-l}c_L(l)$.*

In this thesis our focus will be on a fixed LDPC or LDGM code or a fixed ensemble of LDPC or LDGM codes and its performance under various decoders when the underlying scalar parameter characterizing the channel noise is varied. Inherent in the above statement is the notion of a *family* of channels, characterized by a single scalar parameter, which is *complete*, *ordered* and *smooth*. All of the above notions are standard and details can be found in [34], [23], but for completeness we state them here. We will consider a *family* of binary-input memoryless symmetric channels denoted by BMS(ϵ), characterized by the scalar ϵ which is in one-to-one correspondence with the channel entropy. For any general BMS(ϵ), we define the range of the noise parameter to be $[0, \epsilon_{\max}]$, where $\epsilon_{\max} \in \overline{\mathbb{R}}^+$. For all BMS($\epsilon$) channels, $\epsilon = 0$ denotes a noiseless channel and $\epsilon = \epsilon_{\max}$ denotes a maximally noisy channel. For example, for all the three channels mentioned above, the corresponding channel families are given by BEC(ϵ) with $\epsilon \in [0, 1]$ ($\epsilon_{\max} = 1$), BSC(ϵ) with $\epsilon \in [0, 1/2]$ ($\epsilon_{\max} = 1/2$) and BIAWGNC(ϵ) with $\epsilon \in [0, \infty)$ ($\epsilon_{\max} = \infty$). *Completeness* requires that the channel entropy varies over the entire range from 0 ($\epsilon = 0$) to the entropy of the input, $H(X)$ ($\epsilon = \epsilon_{\max}$), when ϵ is varied over its entire range.

We say that the channel $p_{Z|X}(z \mid x)$ is *physically degraded* with respect to $p_{Y|X}(y \mid x)$, if there exists a joint distribution $p_{Y,Z|X}(y, z \mid x) = p_{Y|X}(y \mid$

$x)p_{Z|Y}(z \mid y)$. A simple way to understand this is that, physical degradation implies that the symbol error probability of the channel $p_{Z|X}(z \mid x)$ is greater than the symbol error probability of the channel $p_{Y|X}(y \mid x)$. The channels are said to be *ordered* by physical degradation if and only if $\epsilon_1 < \epsilon_2$ implies that the channel $p^{\epsilon_2}_{Y|X}$ is physically degraded with respect to the channel $p^{\epsilon_1}_{Y|X}$. The notion of ordered family allows us to define decoding thresholds (depending on the decoder) beyond which reliable communication is not possible using that decoder.

A natural assumption on the channel family that we impose is that of *smoothness*, which implies that the channel entropy varies smoothly with respect to the channel parameter ϵ.

Definition 1.4 (Channel Smoothness, [23])**.** *The channel family BMS(ϵ) is said to be smooth if for all $x \in \{-1, +1\}$ and all bounded continuously differentiable functions $f(y)$, the integral $\int f(y)p^{\epsilon}_{Y|X}(y \mid x)dy$ exists and is a continuously differentiable function with respect to ϵ.*

Above, the notation $p^{\epsilon}_{Y|X}(y \mid x)$ denotes the transition probability of a channel with noise value ϵ. If the channel family is smooth, one can formally set $\frac{d}{d\epsilon}\int f(y)p^{\epsilon}_{Y|X}(y \mid x)dy \triangleq \int f(y)\frac{dp^{\epsilon}_{Y|X}(y|x)}{d\epsilon}dy$.

A quick computation tells us that the channel families: BEC(ϵ), BSC(ϵ) and BIAWGNC(ϵ) are complete, ordered and smooth.

1.3 Maximum a Posteriori Decoder

Consider transmission over a BMS(ϵ) channel using a code of block-length n. Denote the transmitted codeword by $\underline{x} = (x_1, x_2, \ldots, x_n)$. One of the most fundamental decoders is the bit maximum a posteriori decoder (MAP). The decoder is defined as follows. The estimate provided by the bit MAP decoder for the i^{th} bit is denoted by $\hat{x}_{i,\mathrm{MAP}}$ and is given by,

$$\hat{x}_{i,\mathrm{MAP}} \triangleq \mathrm{argmax}_{x_i \in \{-1,+1\}} p_{X_i|\underline{Y}}(x_i \mid \underline{y}). \tag{1.4}$$

An important fact is that the bit MAP decoder minimizes the bit probability of error under random channel errors. Thus it is *optimal* with respect to bit error rate.

We denote by $\underline{y}^{\sim i}$, the vector of all observations except of the code-bit at position i. Crucial to our analysis are the notions of *intrinsic* and *extrinsic* MAP estimate given in the LLR form by,

$$\phi_{i,\mathrm{MAP}}(\underline{y}) \triangleq \log \frac{p_{X_i|\underline{Y}}(+1 \mid \underline{y})}{p_{X_i|\underline{Y}}(-1 \mid \underline{y})}, \quad \phi_{i,\mathrm{MAP}}(\underline{y}^{\sim i}) \triangleq \log \frac{p_{X_i|\underline{Y}^{\sim i}}(+1 \mid \underline{y}^{\sim i})}{p_{X_i|\underline{Y}^{\sim i}}(-1 \mid \underline{y}^{\sim i})}. \tag{1.5}$$

In words, $\phi_{i,\mathrm{MAP}}(\underline{y}^{\sim i})$ is the estimate of the MAP decoder of the code-bit i, when all the observations except y_i are available to the decoder. The intrinsic

and extrinsic MAP estimate are a random variables since the outputs $\underline{y}$ and $\underline{y}^{\sim i}$ are random. We denote the random intrinsic and extrinsic MAP estimates by $\Phi_{i,\mathrm{MAP}}(\underline{Y})$ and $\Phi_{i,\mathrm{MAP}}(\underline{Y}^{\sim i})$ respectively. An important fact is that the extrinsic (intrinsic) MAP estimate is a sufficient statistic for estimating the code-bit i given $\underline{Y}^{\sim i}$ ($\underline{Y}$) (see [23]). In the language of information theory, this means that $H(X_i \mid \underline{Y}^{\sim i}) = H(X_i \mid \Phi_{i,\mathrm{MAP}}(\underline{Y}^{\sim i}))$ and $H(X_i \mid \underline{Y}) = H(X_i \mid \Phi_{i,\mathrm{MAP}}(\underline{Y}))$.

Since the family of channels that we consider here are physically degraded, one defines the notion of the *MAP threshold*, which is denoted by ϵ_{MAP}, as the minimum channel noise for which reliable communication is not possible on an average. In other words, above the MAP threshold, the average (over the ensemble of codes) error probability, denoted by $\mathbb{P}_{\mathrm{e,\,MAP}}$, is strictly greater than zero.

Although the MAP decoder is optimal with respect to the bit error rate, it is computational expensive when we consider coded transmission. More precisely, to perform MAP estimation we are required to know the *a posteriori* marginal distribution $p_{X_i|\underline{Y}}(x_i \mid \underline{y})$. Computing this marginal involves summing out all the code-bits except code-bit i in the joint distribution $p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y})$. This sum involves an exponential number of terms and optimal decoding comes at a price of expensive computation time. More precisely, consider the case of LDPC codes as an example. The a posteriori distribution can be written as

$$\begin{aligned} p_{X_i|\underline{Y}}(x_i \mid \underline{y}) &= \sum_{\sim x_i} p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y}) = \frac{1}{p_{\underline{Y}}(\underline{y})} \sum_{\sim x_i} p_{\underline{Y}|\underline{X}}(\underline{y} \mid \underline{x}) p_{\underline{X}}(\underline{x}) \\ &= \frac{1}{|C| p_{\underline{Y}}(\underline{y})} \sum_{\sim x_i} \prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid x_i) \mathbb{1}(\underline{x} \in C) \\ &= \frac{1}{|C| p_{\underline{Y}}(\underline{y})} \sum_{\sim x_i} \prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid x_i) \prod_{a=1}^{m} \mathbb{1}(\oplus_{k \in \partial a} x_k = 0), \end{aligned} \tag{1.6}$$

where $p_{\underline{X}}(\underline{x}) = \frac{\prod_{a=1}^{m} \mathbb{1}(\oplus_{k \in \partial a} x_k = 0)}{|C|}$, is the uniform distribution over all the codewords and x_k denotes the code-bit k present in the check constraint a. Above, $\sum_{\sim x_i}$ denotes sum over all variables except x_i. From above we can conclude that to evaluate the exact marginal (MAP decoder), one has to consider an exponential in n number of terms.

1.4 Belief Propagation Decoder

LDPC/LDGM codes are described by constraints which act locally. This has resulted in the application of the *sum-product* rule [23], [33], [18], [35] to estimate the a posteriori marginal distributions. The resulting decoder is called the belief propagation decoder (BP).

The first step in developing the sum-product rules is to recognize that the function, whose marginal is required, is a product of a large number of

local functions (often called as *kernels*). Indeed in the case of transmission using LDPC codes, the function in (1.6) is a product of n kernels of the form $p_{Y_i|X_i}(y_i \mid x_i)$ (depends only on x_i) for $1 \leq i \leq n$ and m kernels of the form $\mathbb{1}(\oplus_{k\in\partial a} x_k = 0)$ (depends only on ∂a) for $1 \leq a \leq m$. Both these kernels act locally and the application of the distributive law of the underlying field (the field with which we work is $\mathbb{R}$) yields the sum-product rules (see [23], [33], [35]) to *approximately* evaluate the marginal. A very important theorem regarding the sum-product rule is that it is *exact* if the Tanner graph is a tree. On general Tanner graphs with loops, the rules do not evaluate the marginals exactly, and the estimation is truly an approximation. The complexity of decoding all code-bits, for the BP decoder with fixed iterations for a bounded degree code, is proportional to n.

Let us now state the rules more precisely for the case of transmission using LDPC codes. It is convenient to state the sum-product rules in terms of the channel LLR , defined in Definition 1.3, which gets rid of the common factor, $|C|p_{\underline{Y}}(\underline{y})$ in (1.6). The sum-product rule is stated in terms of *messages*, which are exchanged between the variable nodes and check nodes of the Tanner graph, depending on a schedule (see [23], [9]). Each edge carries a message in both the directions.

The message from the variable node i to the check node a is denoted by $\zeta_{i\to a}(\in \mathbb{R})$ and intuitively corresponds to the *belief* of the variable node i of its LLR value depending only on the information provided by the check nodes other than a and the channel LLR . The rule assumes that the information provided are conditionally independent given the value of the variable node.

The message from the check node a to the variable node i is denoted by $\zeta_{a\to i}(\in \mathbb{R})$ and intuitively corresponds to the belief provided by the check node a of the LLR value of the variable node i, given only the information provided by the remaining variable nodes in its neighborhood. The schedule that we consider throughout this thesis is the parallel schedule, where $\zeta_{i\to a}$ are computed first for all variable nodes i and check nodes a, followed by the computation of $\zeta_{a\to i}$. The rules are

- Variable node rule (message from variable node i to check node a):

$$\zeta_{i\to a}^{(d)} = l_i + \sum_{b\in\partial i\setminus a} \zeta_{b\to i}^{(d-1)}. \tag{1.7}$$

- Check node rule (message from variable node a to check node i):

$$\zeta_{a\to i}^{(d)} = 2\tanh^{-1}\left(\prod_{j\in\partial a\setminus i} \tanh\left(\frac{\zeta_{j\to a}^{(d-1)}}{2}\right)\right). \tag{1.8}$$

- Initial value (initial message from variable node i to check node a):

$$\zeta_{i\to a}^{(0)} = l_i. \tag{1.9}$$

- Final value (BP estimate of the code-bit i after d iterations):

$$\phi_{i,BP}^{(d)}(\underline{y}) = l_i + \sum_{a \in \partial i} \zeta_{a \to i}^{(d)}, \tag{1.10}$$

where d denotes the iteration number. The final estimate of the BP decoder, $\phi_{i,BP}^{(d)}(\underline{y})$ is a random variable with respect to randomness in the channel LLR values and also with respect to the randomness in the code picked from the ensemble. A related quantity is the extrinsic BP estimate denoted by $\phi_{i,BP}^{(d)}(\underline{y}^{\sim i})$ which is equal to $\sum_{a \in \partial i} \zeta_{a \to i}^{(d)}$. The initial values are given by the channel LLR present at the variable nodes.

For the case of transmission using LDGM codes, the channel observation is present at the check nodes and not at the variable nodes, hence the variable node message $\zeta_{a \to i}^{(d)}$, at iteration d, is given by

$$\zeta_{a \to i}^{(d)} = \sum_{j \in \partial a \setminus i} \zeta_{j \to a}^{(d-1)},$$

the check node rule modifies to

$$\zeta_{i \to a}^{(d)} = 2 \tanh^{-1} \left(\tanh\left(\frac{l_i}{2}\right) \prod_{b \in \partial i \setminus a} \tanh\left(\frac{\zeta_{b \to i}^{(d-1)}}{2}\right) \right),$$

and the final estimate of the information bit a becomes

$$\phi_{a,BP}^{(d)}(\underline{y}) = \sum_{i \in \partial a} \zeta_{i \to a}^{(d)}.$$

A related estimate which we will need is the BP estimate of the check node bit i, denoted by

$$\phi_{i,BP}^{(d)}(\underline{y}) = 2 \tanh^{-1} \Big(\tanh\left(\frac{l_i}{2}\right) \prod_{a \in \partial i} \tanh\left(\frac{\zeta_{a \to i}^{(d-1)}}{2}\right) \Big).$$

The initial values are given by complete erasure messages from the variable nodes, i.e. $\zeta^(0)_{a \to i} = 0$. Similar to LDPC codes, we define $\phi_{i,BP}^{(d)}(\underline{y}^{\sim i})$ as the extrinsic BP estimate for the code-bit i. As a consequence of the rules, the BP decoder for LDGM codes never starts. However, with the presence of an arbitrarily small fraction of degree one check nodes, this problem can be overcome.

Again, since the family of channels that we consider here are physically degraded, one defines the notion of the BP threshold, which is denoted by ϵ_{BP}, as the minimum channel noise for which reliable communication using BP decoder (average (over the code ensemble) error probability, $\mathbb{P}_{e,\text{BP}} > 0$) is not possible.

An important simplification in the analysis of the BP decoder presented above is that the error probability under BP decoding is independent of the codeword transmitted and one can thus restrict to the all-one codeword (see Lemma 4.90 in [23]).

Example 1.6 (BP decoder for the BEC(ϵ))**.** *Assume the all-one codeword transmission over the BEC(ϵ) with erasure fraction given by $\epsilon \in [0,1]$. Since initial* LLR *values are either* 0 *or* $+\infty$*, the messages along any edge at any time in the schedule is always* 0 *or* $+\infty$*. We call* 0 *message (resp.* $+\infty$*) an erasure (resp. known) message. The BP decoder rules in this case simplify to: (i) $\zeta_{i\to a}$ is erasure if all the incoming messages (this includes the channel* LLR *in the case of LDPC codes) $\zeta_{b\to i}, b \in \partial i \setminus a$ are erasures, else $\zeta_{i\to a}$ is known; (ii) $\zeta_{a\to i}$ is erasure if any incoming message (this includes the channel* LLR *in the case of transmission using LDGM codes) is an erasure and is equal to ∞ if all incoming are known.*

As mentioned previously, the sum-product rule defined above, exactly computes the a posteriori distribution $p_{X_i|\underline{Y}}(x_i \mid \underline{y})$ for a code whose Tanner graph is a tree. An intuitive reason for this is that, the messages on a tree are independent, which is what the BP decoding rules assume. The sum-product rule on a general graph with cycles does not give the exact estimate. In fact, the rules need not even converge in general.

Nevertheless, for the case of transmission over the BEC(ϵ) using regular LDPC code ensemble, it was shown recently by Measson et al. [24], that above ϵ_{MAP}, the bit MAP error rate equals the BP error rate, with high probability[2]. I.e., BP decoder, in this case, is computing the exact marginals, with high probability, even in the presence of loops! It was possible to show this result, in the case of BEC, mainly due the simplicity of the message space ($\{0,\infty\}$). One of the main contribution of this thesis is to establish that even in the presence of cycles, there are non-trivial regimes of the noise parameter ϵ, for a fairly general class of BMS(ϵ) channels, where the BP decoder estimate is exact with high probability.

1.5 Tools for Performance Analysis

In this section we briefly review the important tools and quantities which appear in the asymptotic analysis of the bit MAP and BP decoders. In the literature, as we will see in the next section, the analysis of the MAP and BP decoders have largely been developed in an orthogonal fashion. On one hand, the MAP decoding analysis heavily relied on the weight spectrum of the code [16], and on the other hand the BP decoder (when transmitting over general BMS channels) was analyzed using the *density evolution* technique of Richardson and Urbanke (see [23], [9]).

The convergence of statistical physics and communication has brought out a very deep connection exists between the density evolution analysis and the MAP decoding analysis in the limit of infinite block-length [26], [28], [27]. We will talk more about this connection in Section 1.6. The tools important to this thesis are the conditional entropy (described in Section 1.5.2) and the density

[2]The probability here is with respect to the ensemble of codes.

evolution equations (described in Section 1.5.1), for the MAP decoding and BP decoding analysis respectively.

Recently, Measson et al. [24], [36] introduced and used the *generalized extrinsic information transfer function* (GEXIT) to demonstrate an intimate relationship between the BP decoder and the MAP decoder. We describe the GEXIT functions in Section 1.5.3, which will be central to this thesis.

Some of the results in this thesis hold for *any* fixed code, but some results we show only on an average with respect to the code ensemble. However most of the quantities of interest in the analysis are very tightly concentrated around the ensemble average and we suffer minimal loss in showing average results [23], [29].

1.5.1 Density Evolution: Asymptotic Analysis of the BP Decoder

Consider transmission over a general BMS channel using an LDPC or LDGM code, of block-length n, picked u.a.r. from an ensemble of codes. As usual we can restrict the analysis to the all-one codeword transmission [23]. Let $\phi_{i,BP}^{(d)}(\underline{y})$ denote the estimate of the code-bit i (in terms of LLR) after d iterations of the BP decoder. The *density evolution* technique gives a formula to evaluate the average (over the code ensemble) of the distribution of the estimate $\phi_{i,BP}^{(d)}(\underline{y})$, when n goes to ∞. I.e., for a fixed number of iterations, it provides a formula for the density of the BP estimate for large block-lengths.

Often one is interested in a different limit, i.e., the BP estimate when one fixes the block-length n and the number the iterations d go to infinity. It is conjectured that the probability of error of the BP decoder under the limit $\lim_{d\to+\infty}\lim_{n\to+\infty}$ is the same as that under the limit $\lim_{n\to+\infty}\lim_{d\to+\infty}$ and this was recently shown for some cases to be true in [37]. For the BP-GEXIT function, which we will introduce shortly, we will show in some cases also that under both order of limits (first $n\to+\infty$ and then $d\to+\infty$ and vice-versa), the computation of the average BP-GEXIT function is the same.

An important object which occurs in the density evolution analysis is the *computation tree*. The computation tree of depth d, as defined in [9], is the unraveling of the graph, starting from the code-bit to be estimated, to include all variable nodes which are at most a distance d from the root node. Note that in the computation tree one can have labels of variable nodes and check nodes repeating. Thus it is not a tree in the typical graph-theoretical sense.

The advantage of taking the block-length to infinity first is that, as shown in [9], the sparsity of the code ensures that with high probability the computation tree of depth d around the root node becomes a tree in the graph-theoretic sense and the BP estimate, after a fixed d number of iterations, can be easily evaluated. This gives rise to the density evolution technique. Here we only state the final result of density evolution. For details see [23], [18], [9], [19], [8]. An important result of [9] is that for a fixed d and block-length n going to in-

finity, the BP estimate is exactly given by the density evolution estimate, with high probability. We present here a slightly weaker version of their theorem.

Theorem 1.1 (Richardson and Urbanke)**.** *Consider an LDPC code picked u.a.r. from an ensemble of codes with edge perspective (resp. node perspective) degree distributions given by $\lambda(x), \rho(x)$ (resp. $\Lambda(x)$ and $P(x)$). Consider transmission over a general BMS(ϵ) channel with the distribution of the* LLR *l under the all-one codeword transmission given by $c_L(l)$. Consider d iterations of the BP decoder and let $\phi^{(d)}_{i,BP}(\underline{y})$ denote the BP estimate of code-bit i. Then the random variable $\frac{1}{n}\sum_{i=1}^{n}\phi^{(d)}_{i,BP}(\underline{y})$, for any $d \geq 1$, converges in distribution to the random variable with probability density function given by*

$$c_L(l) \circledast \Lambda(\rho(a_{d-1})),$$

where

$$\begin{aligned} a_{d-1} &= c_L(l) \circledast \lambda(\rho(a_{d-1})), \\ a_0 &= c_L(l), \end{aligned} \tag{1.11}$$

and where $\Lambda(a) = \sum_i \Lambda_i a^{\circledast(i-1)}$, $\lambda(a) = \sum_i \lambda_i a^{\circledast(i-1)}$ and $\rho(a) = \sum_k \rho_k a^{\boxtimes(k-1)}$. Here $\circledast$ denotes the normal convolution of two functions in $\mathbb{R}$ and $\boxtimes$ is the convolution of densities represented in the $G-$domain (see Section 1.2). Also a_0 is fixed to $c_L(l)$.

Remark 1.1. *In fact one can show that [23] the probability of error of the BP decoder with d iterations converges (as $n \to \infty$) almost surely to $\mathfrak{E}(a_d)$ where $\mathfrak{E}$ is the "error functional" (see [23] for a precise definition).*

Remark 1.2. *Consider the case of transmission using an LDGM code picked u.a.r. from an ensemble of codes with edge perspective degree distributions given by $\lambda(x), \rho(x)$ ($\Lambda(x)$ denotes the variable node degree distribution and $P(x)$ denotes the check node degree distribution). Then $\frac{1}{n}\sum_{i=1}^{n}\phi^{(d)}_{i,BP}(\underline{y})$, for any $d \geq 1$, converges in distribution to the random variable with probability density function given by $c_L(l) \circledast P(a_d)$, where*

$$\begin{aligned} a_d &= \lambda(c_L(l) \circledast \rho(a_{d-1})), \\ a_0 &= \delta_0(l), \end{aligned} \tag{1.12}$$

and where $P(a) = \sum_k P_k a^{\boxtimes(k-1)}$, $\rho(a) = \sum_k \rho_k a^{\boxtimes(k-1)}$ and $\lambda(a) = \sum_i \lambda_i a^{\circledast(i-1)}$.

Remark 1.3. *In this thesis the above equations will appear in integral form. For the case of LDPC code ensembles we have the following. The average BP estimate $\frac{1}{n}\sum_i \phi^{(d)}_{i,BP}$ converges in distribution to the random variable $\Delta^{(d)} = l + \sum_{a=1}^{\ell} u_a^{(d)}$. Here ℓ is random and distributed as $\Lambda(x)$ and l is the* LLR

distributed as $c_L(l)$. The $u_a^{(d)}$ are i.i.d. random variables with distribution obtained from the iterative system equations

$$\widehat{\eta}^{(d)}(u) = \sum_{\imath} \lambda_{\imath} \int \prod_{j=1}^{\imath-1} dv_j \, \eta^{(d)}(v_j) \, \delta\left(u - 2\tanh^{-1}\left(\prod_{j=1}^{\imath-1} \tanh \frac{v_j}{2} \right) \right),$$

$$\eta^{(d)}(v) = \sum_{k} \rho_k \int dl \, c_L(l) \prod_{a=1}^{k-1} du_a \, \widehat{\eta}^{(d-1)}(u_a) \, \delta\left(v - l - \sum_{a=1}^{k-1} u_a^{(d)} \right).$$

The initial condition is given by $\eta^{(0)}(v) = c_L(v)$. One can write similar formulas for the case of LDGM code ensembles.

The density evolution equations gives us a nice recursive procedure to evaluate the average BP decoder estimate after d iterations for a code-bit.

Example 1.7 (Density Evolution for BEC(ϵ))**.** *Consider transmission over the BEC(ϵ) using an LDPC code ensemble given by the edge perspective degree distributions $(\lambda(x), \rho(x))$. Then Theorem 1.1 tells us that the probability of a code-bit being in erasure after fixed d iterations of the BP decoder, is given by $\epsilon\Lambda(1-\rho(1-a_d))$, with high probability, where*

$$a_d = \epsilon\lambda(1-\rho(1-a_{d-1})),$$

$a_0 = \epsilon$ and $\Lambda(x)$ is the node perspective degree distribution of the variable nodes. Note that for the BEC(ϵ), $\lambda(a_d) = \sum_i \lambda_i (a_d)^{i-1}$ and $\rho(a_d) = \sum_i \rho_i (a_d)^{i-1}$.

For transmission using LDGM code ensemble given by edge degree distributions $(\lambda(x), \rho(x))$, the probability of a code-bit being in erasure after d iterations of the BP decoder, is given by $1-(1-\epsilon)P(1-a_d))$, with high probability, where

$$a_d = \lambda(1-(1-\epsilon)\rho(1-a_{d-1})), \tag{1.13}$$

$a_0 = 1$ and $\Lambda(x)$ is the node perspective degree distribution of the variable nodes representing the information bits. As is seen from the equation (1.13), if the fraction of edges attached to degree one check nodes $\rho_1 = 0$, then $a_d = 1$ for all d. For any ensemble with $\rho_1 > 0$, the iterations start and will give a non-trivial fixed-point.

1.5.2 Conditional Entropy

The conditional entropy is the entropy of the transmitted codeword (information word) given the received message, $H(\underline{X} \mid \underline{Y})$ ($H(\underline{U} \mid \underline{Y})$) for transmission using an LDPC code (LDGM code). From Fano's inequality, see [29], we know that for a given code C of block-length n and rate r,

$$\mathbb{P}_{\text{block,MAP}} \geq \frac{H(\underline{Z} \mid \underline{Y}) - 1}{nr}$$

$$h(\mathbb{P}_{\text{bit,MAP}}) \geq \frac{H(\underline{Z} \mid \underline{Y})}{n},$$

where $\underline{Z} = \underline{X}(\underline{U})$ is the codeword (information word) in case of LDPC codes (LDGM codes) and $h(\cdot) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function. From the above, it is clear that knowing the conditional entropy exactly would provide an upper bound on the channel parameter beyond which reliable transmission using the code C with given parameters n, r is not possible. Indeed, channel noise for which the per-bit conditional entropy is strictly positive, i.e., $H(\underline{Z} \mid \underline{Y})/n > 0$, implies a non-zero bit error probability for the MAP decoder. Thus the conditional entropy is an important quantity in the analysis of the MAP decoder. As it turns out, in this thesis we will show that the conditional entropy has a very intimate relation to the density evolution analysis of the BP decoder.

Let us look at the conditional entropy of a bit more closely. Assume transmission using LDPC codes. The conditional entropy is by definition

$$H(\underline{X} \mid \underline{Y}) = -\mathbb{E}_{\underline{Y}}\Big[\sum_{\underline{x}} p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y}) \ln p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y})\Big],$$

where $\mathbb{E}_{\underline{Y}}$ denotes the expectation with respect to $p_{\underline{Y}}(\underline{y})$. Since the channel is output symmetric, i.e., $p_{Y|X}(y \mid x) = p_{Y|X}(-y \mid -x)$, and the code used for transmission is a vector space, it is not difficult to show that the conditional entropy can equivalently be written as

$$H(\underline{X} \mid \underline{Y}) = -\mathbb{E}_{\underline{Y}|\underline{1}}\Big[\sum_{\underline{x}} p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y}) \ln p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y})\Big]. \tag{1.14}$$

Above we use $\mathbb{E}_{\underline{Y}|\underline{1}}$ to denote the expectation with respect to the probability measure $\prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid +1)$. As seen from the above equation, to evaluate the conditional entropy, one must perform a sum which has exponentially many terms. Thus again we see that computation of the conditional entropy is hard. One of the main contributions of this thesis is an analytical formula for the conditional entropy in terms of the code and the noise parameters only, as the block-length goes to infinity.

1.5.3 GEXIT Functions

It is observed that under MAP decoding there is a *phase transition* phenomenon. More precisely, there is a critical noise value at which the error probability of the MAP decoder changes abruptly. We stress that there can be many discontinuities in the error probability of the MAP decoder. At one such critical noise value, the average conditional entropy becomes non-zero (from zero). This is known as the MAP decoding threshold denoted by ϵ_{MAP}. At ϵ_{MAP}, the derivative of the conditional entropy is discontinuous. Thus it seems natural to study the derivative of the conditional entropy to estimate the MAP threshold.

Consider transmission of a fixed code C (LDPC or LDGM) over a BMS(ϵ) channel. Let us denote the per-bit conditional entropy by $h_n = \frac{1}{n}H(\underline{Z} \mid \underline{Y})$,

where $\underline{Z} = \underline{X}$ in the case of LDPC codes and $\underline{Z} = \underline{U}$ in the case of LDGM codes. The dependence on the code C has been suppressed for notational convenience. Under the assumption of a smooth channel family BMS(ϵ), the MAP generalized extrinsic information transfer function or in short, the MAP-GEXIT function is defined as in [36]

Definition 1.5 (MAP-GEXIT)**.** *Assume transmission over a smooth channel family BMS(ϵ) using a fixed LDPC or LDGM code of block-length n. Then the MAP-GEXIT function, denoted by $g_n(\epsilon)$ is given by the derivative, with respect to the noise parameter, of the conditional entropy. More precisely,*

$$g_n(\epsilon) = \frac{dh_n}{d\epsilon}. \tag{1.15}$$

Under the assumption of a smooth channel family it can be shown, for both the LDPC and LDGM case, [36] that

$$g_n(\epsilon) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial H(X_i \mid \underline{Y})}{\partial \epsilon_i}\Big|_{\epsilon_i=\epsilon}. \tag{1.16}$$

For the MAP decoder there is an integral formula of $\frac{\partial H(X_i|\underline{Y})}{\partial \epsilon_i}\Big|_{\epsilon_i=\epsilon}$. We have

Theorem 1.2 (Integral Formula – Measson et al. [36])**.** *Assume that transmission takes place over a smooth BMS(ϵ) channel family using a LDPC or LDGM code C*[3]*.*

$$\frac{\partial H(X_i \mid \underline{Y})}{\partial \epsilon_i}\Big|_{\epsilon_i=\epsilon} = \int dl_i \frac{dc_L(l_i)}{d\epsilon} \mathbb{E}_{\underline{Y}^{\sim i}|\underline{1}} \ln\left[\frac{1+\tanh\frac{l_i}{2}\tanh\left(\frac{\phi_{i,MAP}(\underline{Y}^{\sim i})}{2}\right)}{1+\tanh\frac{l_i}{2}}\right], \tag{1.17}$$

where $\mathbb{E}_{\underline{Y}^{\sim i}|\underline{1}}$ is the expectation with respect to all the $\underline{Y}$ except Y_i, under the assumption of the all-one codeword transmission. Also, $c_L(l_i)$ denotes the distribution of the LLR *l_i under the assumption of the all-one codeword transmission.*

Proof. We will provide a proof of this in Section 2.3 after we introduce a convenient language to state and prove all our results. □

Note that in the equation (1.17), the extrinsic MAP estimate $\phi_{i,\mathrm{MAP}}(\underline{Y}^{\sim i})$ requires an exponential in the block-length n operations to compute its value. As a result the MAP-GEXIT is a hard quantity to determine. Also the definition and the integral formula of the MAP-GEXIT function is not restricted to sparse codes and can be easily extended to any linear code.

[3]We require that the code is *proper*, meaning that the apriori symbol probability is 1/2. But this is the case if the dual of the code C has a generator-matrix representation with no zero column, see [34].

Example 1.8 (MAP-GEXIT for the BEC(ϵ))**.** *Consider transmission over a BEC(ϵ) with erasure probability ϵ using an LDPC code. It is convenient to imagine that each code-bit i is transmitted through an independent BEC with erasure probability given by ϵ_i. It can be shown that the MAP-GEXIT function is equal to $\frac{1}{n}\sum_{i=1}^{n} H(X_i \mid \underline{Y}^{\sim i})$ (see [23], [34]). Indeed, for the erasure channel we have,*

$$H(X_i \mid \underline{Y}) = H(X_i \mid \underline{Y}^{\sim i}, Y_i) = \sum_{y_i} H(X_i \mid \underline{Y}^{\sim i}, Y_i = y_i)\mathbb{P}(Y_i = y_i)$$
$$= \mathbb{P}(y_i = *)H(X_i \mid \underline{Y}^{\sim i}) = \epsilon_i H(X_i \mid \underline{Y}^{\sim i}),$$

Thus taking the derivative w.r.t. ϵ_i we find,

$$\frac{\partial H(X_i \mid \underline{Y})}{\partial \epsilon_i} = H(X_i \mid \underline{Y}^{\sim i}).$$

Above, $*$ denotes the erasure symbol. In obtaining the above equations we used the fact that (i) if y_i is either $+1$ or -1, then X_i is perfectly known, implying that $H(X_i \mid \underline{Y}^{\sim i}, y_i \in \{+1, -1\}) = 0$ and (ii) $H(X_i \mid \underline{Y}^{\sim i})$ is independent of the noise parameter ϵ_i. Thus we have $g_n(\epsilon) = \frac{1}{n}\sum_i H(X_i \mid \underline{Y}^{\sim i})$. The function $H(X_i \mid \underline{Y}^{\sim i})$ is called the average extrinsic information transfer function (EXIT) and was introduced by Ashikhmin et al. in [38]. Intuitively the EXIT function determines the amount of information that all other noisy code-bits provide about the code-bit i. More precisely, one can show that $H(X_i \mid \underline{Y}^{\sim i})$ is equal to the erasure probability of the MAP decoder given the observations $\underline{Y}^{\sim i}$. Although for the BEC(ϵ), this turns out to be a pivotal quantity in design and analysis of LDPC codes [23], it seems to have lesser use in the analysis of LDPC codes over general BMS(ϵ) channels, most notably because of its failure to satisfy the area theorem [38], [23], [36]. The area theorem for the BEC(ϵ) channel states that the area under the EXIT function, when ϵ is varied from $0 \to 1$, is equal to the rate of the code. Unfortunately, if we consider the EXIT function for general BMS(ϵ) channels, the area theorem is not satisfied. But the MAP-GEXIT function, from its definition, trivially satisfies the area theorem. Also, importantly, as shown in this example, it matches the definition of the EXIT function for the BEC(ϵ).

Definition 1.6 (Average MAP-GEXIT)**.** *Assume transmission over a smooth channel family BMS(ϵ) using a fixed LDPC or LDGM code of block-length n picked u.a.r. from an ensemble of codes given by edge degree distributions $(\lambda(x), \rho(x))$. Then the average MAP-GEXIT function, with abuse of notation, is denoted by $g_n(\epsilon)$ and is given by its average over the code ensemble, $\mathbb{E}_{\mathcal{C}}[g_n(\epsilon)]$.*

Example 1.9. *In the Figure 1.4 we illustrate the asymptotic average MAP-GEXIT function for two classes of codes when transmitting over the BEC. The figure on the left is the asymptotic average MAP-GEXIT curve for the LDPC($\Lambda(x) = x^3, P(x) = x^6$) code ensemble. The figure on the right is the asymptotic average MAP-GEXIT function for the LDGM($\Lambda(x) = x^3, P(x) =$*

$0.3x + 0.7x^3$) *code ensemble. Both the curves have one discontinuity. From the Example 1.8, we have that the error probability of the MAP decoder is proportional to the MAP-GEXIT function for the case of transmission over the BEC. Thus for the LDPC*($\Lambda(x) = x^3, P(x) = x^6$) *code ensemble, the jump in asymptotic average MAP-GEXIT signifies the MAP-decoding threshold. For the LDGM codes we know that the error probability is always non-zero. Thus we stress that the discontinuity in the asymptotic average MAP-GEXIT curve for LDGM*($\Lambda(x) = x^3, P(x) = 0.3x + 0.7x^3$) *does not signify the MAP threshold. Nevertheless, it signifies a phase transition in the average (over the code ensemble) error probability of the MAP decoder.*

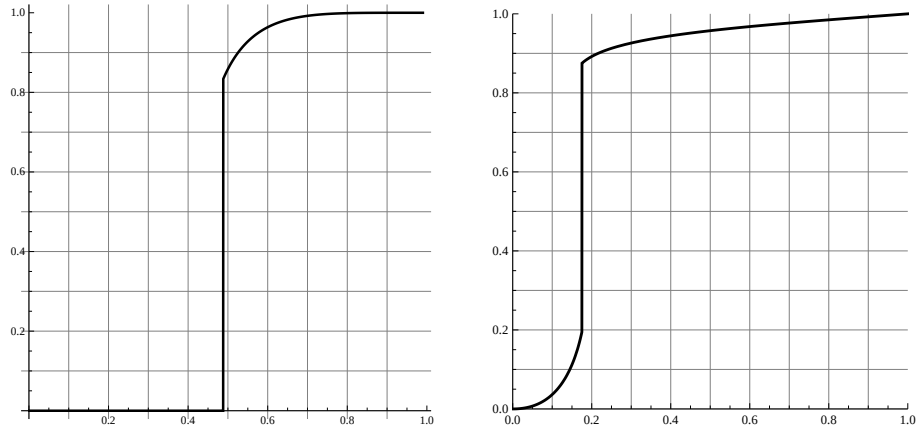

Figure 1.4: The figure on the left plots the asymptotic average MAP-GEXIT for transmission over the BEC using a code from the LDPC($\Lambda(x) = x^3, P(x) = x^6$) ensemble. The figure on the right shows the asymptotic average MAP-GEXIT for transmission over the BEC using a code from the LDGM($\Lambda(x) = x^3, P(x) = 0.3x + 0.7x^3$) ensemble.

Similar to the MAP-GEXIT, we define the GEXIT function associated to the BP decoder.

Definition 1.7 (BP-GEXIT)**.** *Assume transmission over a smooth channel family BMS*(ϵ) *using a fixed LDPC or LDGM code of block-length* n*. Consider the BP decoder with* d *iterations. Then the BP-GEXIT function, denoted by* $g_{n,d}^{BP}(\epsilon)$ *is given by*

$$g_{n,d}^{BP}(\epsilon) = \frac{1}{n}\sum_{i=1}^{n} g_{n,i,d}^{BP}(\epsilon).$$

$$g_{n,i,d}^{BP}(\epsilon) = \int dl_i \frac{dc_L(l_i)}{d\epsilon} \mathbb{E}_{\underline{Y}^{\sim i}|\underline{1}} \ln \left[\frac{1 + \tanh \frac{l_i}{2} \tanh\left(\frac{\phi_{i,BP}(\underline{Y}^{\sim i})}{2}\right)}{1 + \tanh \frac{l_i}{2}} \right]. \qquad (1.18)$$

Similar to the Definition 1.6, we define the average (w.r.t the code ensemble) BP-GEXIT function, and with abuse of notation, denote it by $g_{n,d}^{BP}$.

Remark 1.4. *This definition differs from the one made in [23]. The above definition, as we will see later, is natural when we want to find an analytical formula for the average asymptotic MAP-GEXIT function. The definition in [23] is more natural from an information theoretic point of view and closely follows the equation* (1.16)*. In the asymptotic limit* $\lim_{d\to\infty}\lim_{n\to\infty}$ *the two definitions of the average BP-GEXIT function match.*

The average BP-GEXIT function is given by $\mathbb{E}_{\mathcal{C}}[g_{n,d}^{BP}(\epsilon)]$.

Example 1.10 (Asymptotic Average BP-GEXIT for the BEC(ϵ))**.** *Assume transmission over a BEC*(ϵ) *using LDPC codes from an ensemble given by edge degree distributions* $(\lambda(x), \rho(x))$*. In this example we will compute the asymptotic average BP-GEXIT function. As we will see in Section 4.4, the limiting object* $\lim_{d\to\infty}\lim_{n\to\infty} g_{n,d}^{BP}(\epsilon)$ *exists and is equal to*

$$\lim_{d\to\infty}\int dl\frac{dc_L(l)}{d\epsilon}\mathbb{E}_{\Delta_d}\ln\Big[\frac{1+\tanh\frac{l}{2}\tanh\Delta_d}{1+\tanh\frac{l}{2}}\Big],$$

where Δ_d *is the density evolution estimate. For the BEC*(ϵ)*,* $\lim_{d\to\infty}\Delta_d$ *is distributed as* $\Lambda(1-\rho(1-x_\epsilon))\delta_0(\cdot)+(1-\Lambda(1-\rho(1-x_\epsilon)))\delta_\infty(\cdot))$ *where* x_ϵ *is the largest solution of the fixed point equation* $x=\epsilon\lambda(1-\rho(1-x))$*. Since* $c_L(l)=\epsilon\delta_0(l)+(1-\epsilon)\delta_\infty(l)$ *we have* $\frac{dc_L(l)}{d\epsilon}=\delta_0(l)-\delta_\infty(l)$*. Thus we get*

$$\begin{aligned}\lim_{d\to\infty}\lim_{n\to\infty} g_{n,d}^{BP}(\epsilon) &= \lim_{d\to\infty}\int dl(\delta_0(l)-\delta_\infty(l))\mathbb{E}_{\Delta_d}\ln\left[\frac{1+\tanh\frac{l}{2}\tanh\Delta_d}{1+\tanh\frac{l}{2}}\right]\\ &= -\lim_{d\to\infty}\mathbb{E}_{\Delta_d}\ln\left[\frac{1+\tanh\Delta_d}{2}\right]\\ &= (\ln 2)\Lambda(1-\rho(1-x_\epsilon)).\end{aligned}$$

In the Figure 1.5 we plot the asymptotic average BP-GEXIT function which is proportional to the bit error probability of the BP decoder (in the large block-length and iterations limits). The figure also illustrates the asymptotic average BP-GEXIT function using LDGM($\Lambda(x)=x^3, P(x)=0.3x+0.7x^3$) *code ensemble when transmitting over the BEC.*

1.6 Brief History

The literature on the performance analysis of LDPC codes can be split into their MAP-decoder and BP-decoder analysis. Most of the work on MAP-decoder analysis of LDPC codes can be traced back to the seminal work of Gallager in his thesis [3]. For upper bounding the block probability of error of

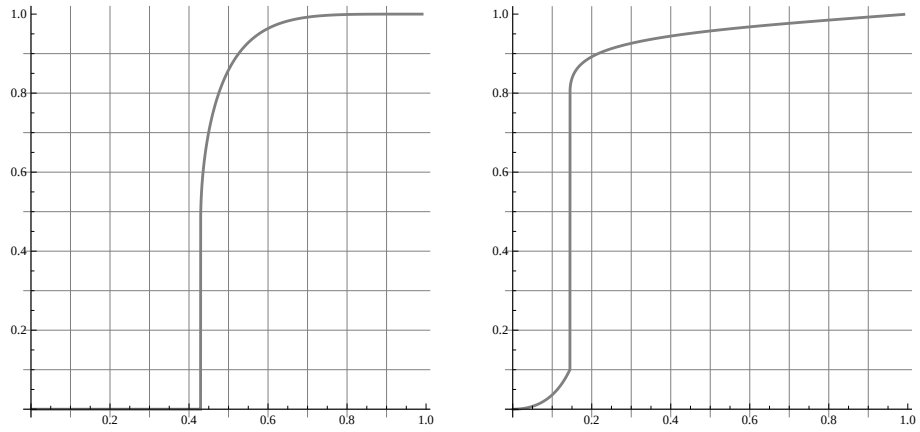

Figure 1.5: The figure on the left shows the asymptotic average BP-GEXIT curve, normalized by $\ln 2$, for transmission over a BEC(ϵ) using a LDPC code from an ensemble with edge degree distributions given by $(\lambda(x) = x^2, \rho(x) = x^5)$. The figure on the right illustrates the asymptotic average BP-GEXIT function for transmission over the BEC using LDGM($\Lambda(x) = x^3, P(x) = 0.3x + 0.7x^3$) code ensemble.

MAP decoding of LDPC codes, Gallager uses the average weight distribution function of the given LDPC code. The idea of the bound is used as a starting point to get better bounds in [16], [39]. A summary of bounds on the MAP-decoder error probability, using the idea of Gallager, can be found in [17]. Let us describe one such bound in more details. The upper bound, as described by Burshtein and Miller in [16], involves estimating two terms. One term involves a union bound over all possible codewords, at a given maximum distance from the transmitted codeword. The Bhattacharya parameter is used to bound the probability of error of confusing a codeword, of a particular weight, with the transmitted codeword ("Bhattacharya distance"). This term usually corresponds to the low-weight codewords, and gives rise to the dominant term in the error probability. The second term, involving the codewords of remaining weights, utilizes the error-exponent technique of Gallager in the context of random coding [40], to provide an exponential in the block-length estimate for the error probability. In [16], the authors also provide a lower bound on the error probability for regular LDPC codes, by finding the probability of finding at least one column of all-zero vector in the parity-check matrix. In general, the upper and lower bounds in [16] are not tight.

Gallager also provided a lower bound on the conditional entropy when transmitting over the BSC in terms of the degree of the check nodes. This was extended to the case of general BMS by Burshtein et al. [41]. This bound, although not tight in general, is valid for any fixed LDPC code. The basic idea behind the bound is to compute the conditional entropy of any syndrome bit given the reliability values of the channel observations. A nice discussion

on this bound can be found in Section 4.11 of [23].

1.6.1 BP in Communication

Over the last decade there has been a plethora of work on the analysis of BP decoder. Most of the work has focused on the BEC and this has culminated in the construction of capacity-achieving sequences for the BEC [42], [21]. Let us mention here the breakthrough works of Luby, Mitzenmacher, Shokrollahi, Spielman and Stemann in [19] and [18] and Richardson and Urbanke in [9]. Luby et al. study the BP decoder by analyzing the equivalent peeling decoder using the differential equation method of Wormald [43]. This was later generalized to the case of transmission over general BMS channels in the notable work of Richardson and Urbanke. In [9], the authors developed the density evolution technique for analyzing the behavior of the BP decoder in the asymptotic block-length limit. The density evolution technique takes advantage of the locally tree-like behavior of sparse graphs. More precisely, in the limit of infinite block-length, any neighborhood of finite depth d around a code-bit (which corresponds to the subgraph "observed" by the BP decoder with d iterations) is a tree with high probability. As a result, the BP decoder messages become independent, yielding the density evolution equation for the BP estimate of the root code-bit. The analysis relies on two basic simplifications: restriction to the all-one codeword when transmitting over BMS channels and considering the average over the ensemble performance (tight concentration) [23].

It is fair to say that for the case of transmission over the BEC, the performance analysis of the BP decoder both in the asymptotic block-length limit and for finite-length have been understood in quite some details. Since the messages in the BP decoding over the BEC are either erasures or provide complete information, it is not hard to show the convergence of the BP decoder [23]. The density evolution technique for asymptotic average analysis of the BP decoder, gives a pleasing one-dimensional fixed-point equation. This allows easy determination of the BP decoding threshold in the limit of infinite block-lengths. The EXIT functions introduced by Ashikhmin et al. [38], serves multiple purposes: it is used as a graphical tool to visualize the BP decoder, it provides the *matching condition* (see [44]) for capacity-achieving codes and also provides the remarkable *area theorem*. The content of the area theorem is that, the area under the EXIT curve, for the MAP decoder (also called as the MAP-EXIT) of any LDPC code equals the rate of the code. In the sequel when we say average MAP-GEXIT or average BP-GEXIT, we mean it in the limit of infinite block-length. In the case of BEC it is also not hard to show that the asymptotic limits of large block-length and number of iterations can be exchanged when computing the bit-error rate of the BP decoder [23].

Exact finite-length analysis for regular LDPC code ensemble when transmitting over the BEC was done by Di et al. [20]. In their analysis, they used the *stopping set* characterization of the failure of the BP decoder, provided by Richardson and Urbanke [45]. Recently, Amraoui et al. [46] have provided an

alternative approach to finite-length analysis using scaling laws of statistical physics.

As we have seen above, the analysis techniques for the MAP and BP decoder have evolved quite independently. It is well known that for a code whose graph is a tree, the BP algorithm has the same performance as the MAP decoder. This essentially comes from the fact that on a tree the computational graph of a node matches the original graph itself [23]. However, codes based on tree graphs have poor performance and one needs to consider graphs with loops or cycles. *With cycles in the original graph, the messages (of the BP decoder) on the computational tree are no longer independent, and it is not a priori clear, if and why, the BP decoder should retain any close relationship to the MAP decoder.*

Recently, Measson, Montanari, Richardson and Urbanke demonstrated in [24], [34], [47] that the BP decoder and the MAP decoder are intimately related when transmitting over the BEC using LDPC codes with loops present in their Tanner graphs. For transmission over the BEC, the MAP-EXIT function simply corresponds to the bit erasure probability of the MAP decoder. They also define the BP-EXIT function which is the bit erasure probability under BP decoding. In [24], Measson et al. show that for regular LDPC codes, above the MAP threshold, the average (over the ensemble of codes) BP-EXIT and the average MAP-EXIT curves match, in the limit of infinite block-length. *This means that the BP decoder above the MAP threshold, for the special case of regular LDPC codes, is optimal, with high probability! In other words, for some noise regime, the asymptotic average MAP-EXIT function can be evaluated by the density evolution equations.*

As an immediate consequence, they obtain the exact value of the MAP decoding threshold for the regular LDPC code ensemble. The proof relies crucially on the fact that for the regular codes the design rate is equal to the actual rate, which was first proved by Miller and Cohen in [48]. In fact the result of Measson et al. holds for a large class of LDPC code ensemble for which the residual graph (the graph remaining after BP decoding) has design rate equal to the actual rate.

Measson et al. also provide an operational interpretation of this connection via the Maxwell decoder and the extended-BP EXIT curve (EBP-EXIT). The EBP-EXIT curve is a plot depicting all the fixed-points of the density evolution equation, for all values of the channel erasure probability in the case of BEC. It is the C-shaped curve in Figure 1.6 for the case of transmission using LDPC($\Lambda(x) = x^3, P(x) = x^6$) ensemble. They show how it is possible to construct the average MAP-EXIT curve from the EBP-EXIT curve by "matching areas" – consider a vertical line and adjust its position till the area to the left of this line and bounded on the left by the EBP-EXIT equals the area to the right of this vertical line and bounded above by the EBP-EXIT curve. The main result of [24] is that the channel erasure fraction at which the two areas become equal is the MAP threshold. This immediately provides the construction of the average MAP-EXIT curve, which is zero till the MAP

threshold and then jumps and clings onto the EBP-EXIT curve. *Thus away from the transition thresholds, (ϵ_{BP} and ϵ_{MAP} in this case[4]) the average MAP-EXIT equals the average BP-EXIT.* This construction is illustrated in Figure 1.6. The dark area bounded on the left by the EBP-EXIT and bounded on the right by the vertical line through the MAP threshold is proportional to the amount of "guessing work" done by the Maxwell decoder and the other dark area is proportional to the "confirmation work" of the Maxwell decoder.

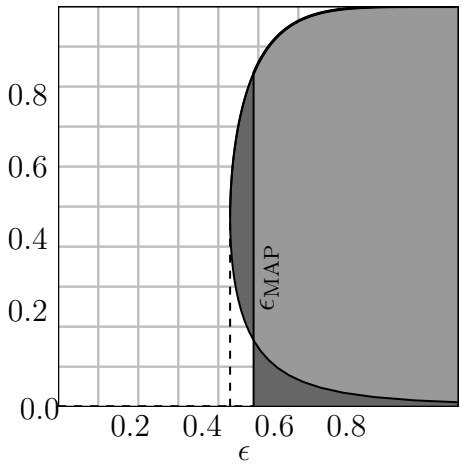

Figure 1.6: In the figure, the C-shaped curve is the EBP-EXIT curve for the case of transmission over the BEC using regular LDPC($\Lambda(x) = x^3, P(x) = x^6$) code ensemble. As shown in [24], the average MAP-EXIT curve is constructed by "sweeping" the EBP-EXIT curve from the right, till the two regions, shown in dark, have equal areas. This figure is borrowed from [23] and is gratefully acknowledged.

This construction is proved rigorously for transmission over the BEC for the regular LDPC code ensembles. To extend their results to general BMS channels, Measson et al. introduced the GEXIT functions. The main motivation for this was that the MAP-EXIT function did not satisfy the area theorem for transmission over general BMS channels. Recall that the MAP-GEXIT function is given by the derivative of the conditional entropy with respect to the noise. Hence by definition, it satisfies the area theorem. Remarkably, the MAP-GEXIT function equals the EXIT function for the BEC as shown in [36], [23]. The BP-GEXIT functions and EBP-GEXIT functions for the general BMS channels were also introduced in [36].

For the trivial case of LDPC code ensembles with a smooth average BP-GEXIT function (no discontinuities), using the area theorem for the MAP-

[4]For irregular LDPC code ensembles there are many discontinuities (or thresholds) in both the average MAP-EXIT and the average BP-EXIT curves. The Maxwell construction, in this case, states that the average BP-EXIT and MAP-EXIT are equal everywhere except between discontinuities.

GEXIT, it was shown in [36], that the average MAP-GEXIT equals the average BP-GEXIT. In general when discontinuities (phase transitions) are present, not much is known: whether the average MAP-GEXIT and BP-GEXIT are equal, away from transition thresholds and whether there is a Maxwell construction which yields the average MAP-GEXIT from the EBP-GEXIT curve. The techniques used for BEC are combinatorial in nature and cannot be easily extended to the case of transmission over general BMS channels and radically new methods have to be used. It is believed that in terms of the GEXIT functions, the results obtained for the BEC still hold. *In particular, the average GEXIT curves for the BP and the MAP decoder should match for high and low noise regimes away from the thresholds.* Recently the convergence of statistical physics and communication science has provided an alternative interpretation of the problem of decoding of sparse graph codes. The work of Montanari in [29] is one of the breakthrough papers in this direction. We will investigate this connection and mention some of the other most notable works in the next section.

1.6.2 BP in Other Problems

Let us mention few other instances of rigorous results concerning the BP algorithm. We do not attempt to cover all examples, but mention very briefly a few notable ones.

The *max-weight independent set (MWIS)* problem is to find the heaviest set of mutually non-adjacent nodes (independent set) in an arbitrary graph with positive weights on nodes. One can formulate an integer program for this problem and provide approximate solutions by solving the corresponding linear programming (LP) relaxation. For this problem also one can construct a probability distribution whose MAP estimate is the MWIS of the given graph. In [49], Sanghavi, Shah and Willsky show that there is a one-to-one correspondence between the fixed-points of the BP (also called as max-product in [49]) algorithm and the extreme points of the LP polytope. As a result they can conclude that the optimality of the BP estimate implies the tightness of the LP relaxation. They also show how any MAP estimation problem can be cast as a problem of finding a MWIS on some graph. Similar results were also shown by Bayati, Shah and Sharma [50] for the *maximum weight matching* problem.

Other models where BP does exact inference is the *ferromagnetic random field Ising model* of statistical physics, as shown by Chertkov in [51]. Similar to the previous work, the MAP solution for this problem involves solving an integer program. It turns out for this problem that the LP relaxation is tight. As shown in [51], the optimal solution of the LP relaxation minimizes the Bethe-Free energy approximation for this problem, which is also got via the BP rules.

Various works [52], [53] have also shown how belief propagation is related to the Bethe approximation of the free energy, which has led to the inter-

pretation of the BP rules as finding the minimum of the Bethe-Free energy approximation. As a result of this observation, various other generalized BP algorithms have been developed based on a similar *variational approach* [54], [52].

In [55], [56], Chertkov and Chernyak develop an expansion of the partition function corresponding to a problem on a fixed graph. The first term in the series corresponds to the Bethe-Free energy approximation. The rest of the terms in the expansion involves average (under BP estimations) of functions of the spins, over generalized loops in the graph. In this thesis we will follow a similar philosophy of perturbing the optimal MAP estimate to bring out the BP estimate plus some correction term, for the problem of decoding of sparse graph codes. We then show that in appropriate range of the noise value, one can control the correction term.

Salez and Shah have shown in [57] the optimality of the BP algorithm, with high probability, in the *random assignment* problem in which the goal is to find the minimum cost matching in a complete bipartite-graph with edge weights being i.i.d with an exponential distribution, in the limit large number of nodes.

Montanari and Shah [58] show the optimality of the BP algorithm in the *random k-satisfiability* problem. Let us discuss this in some more details. In the random k-satisfiability problem can be stated on a bipartite-graph as follows. There are n variable nodes representing n Boolean variables, denoted by $x_1, x_2, \ldots, x_n$. There are m check nodes in the graph. Each of them imposes a logical-OR constraint on k randomly chosen variable nodes. The parameter of the problem (analogous to the channel noise in the coding problem), denoted by α, is the ratio of the number of constraints to the number of variable nodes. For this problem also one can formulate a probability distribution and use the BP algorithm to estimate the marginal distribution for the variable nodes. Montanari and Shah show that if α is small enough, then one can exactly count the normalized (by n) logarithm of the number of satisfying assignments, with high probability, by using the estimates provided by the BP algorithm and an appropriate interpolation method. To show that the BP estimate is optimal, with high probability, they use the fact that α is small enough, which allows them to show a "worst-case" decay of correlation.

Similar results were obtained by Bandyopadhyay and Gamarnik in [59] and by Weitz in [60], for the counting the number of *independent sets* in graphs with large girth. In graph theory, an independent set is a set of nodes of the graph which are mutually disconnected. Bandyopadhyay and Gamarnik show the "worst case" decay of correlations for low degree graphs with large girth, to give an approximate value of the normalized (by the number of nodes in the graph) logarithm of the number of independent sets.

In this thesis we also study decay of correlations in the context of decoding of sparse graph codes. An important difference we would like to point out at this moment is that, we study the "typical-case" correlation decay, which is a stronger form of decay of correlation, as opposed to the "worst case"

correlation decay shown in the previous work. Also the techniques we use are completely different and arise in the spin-glass theory of statistical mechanics. As a consequence we will show the correctness of the BP estimate, with high probability, whenever correlations decay fast enough.

In [61], an arbitrary graph was considered, wherein nodes represented jointly Gaussian random variables $\underline{x}$ (the distribution $P(\underline{x})$ factorizes on the cliques of the graph and variable nodes in a clique are jointly gaussian). The observations $\underline{y}$ and $\underline{x}$ are also jointly gaussian and as usual, the question is to determine when BP gives the correct estimate of the conditional mean and covariance, given the observations. Weiss et al. show that if BP converges then, under certain conditions on the covariance matrix of $\underline{x}$, it gives exact values for the marginal posterior probabilities. Here also the basic idea is to relate true means to those obtained by BP via the correlations. Then under assumed conditions, the correlations decay rapid enough to justify the exactness of the BP estimates.

In [62], Tatikonda et al. used the Dobrushin's criteria of Gibbs measure theory of statistical physics to show the convergence and exactness of BP on loopy graph models. To show this they define the Gibbs measure on the computational tree of the graph. And if the Dobrushin's criteria is satisfied, then there exists a unique Gibbs measure for the computation tree. This then implies the convergence and exactness of BP. Unfortunately, the decoding problem of sparse graph codes over general BMS channels most often does not satisfy the Dobrushin's criteria and we resort to more sophisticated techniques in this thesis to show the exactness of BP, with high probability, in appropriate regimes of the channel noise.

1.7 Connections to Statistical Physics

In this section we show how statistical physics and decoding of sparse graph codes are related. This connection was first established in the pioneering work of Sourlas [25], [63]. There has been a considerable amount of recent work on developing this connection and understanding the coding problem from the statistical physics point of view [64], [65], [66], [27], [26], [67], [68], [28], [69], [70]. The main tool used is the *replica method*, equivalently known as the cavity method [71], [72]. This method is one of the main tools in statistical physics of *spin-glasses*. In communication, the replica method has provided a recipe to determine various quantities, like MAP/BP decoding thresholds, per-bit conditional entropy, error exponents and weight enumerator function for different codes which include LDPC code ensembles and the Turbo codes. Vicente, Saad and Kabashima used the replica method to provide a formula for the average per-bit conditional entropy in [67]. This is known as the *replica solution*. As a consequence they obtain the decoding threshold of the MAP decoder. In [28], Franz, Leone, Montanari and Ricci-Tersenghi also perform *replica symmetric* calculations to determine the MAP threshold.

They further show that, using the *replica symmetric breaking* ansatz, the BP threshold is an "ultimate" threshold for local decoding algorithms. I.e., at the BP threshold there are exponentially many codewords, with energy very close to the transmitted codeword, which would cause all local algorithms, including BP, to fail. In [65], Montanari analyzes the random codeword model, using replica analysis, by exploiting its similarity to the *random energy model* of Derrida [73].

Let us say a few more words on *spin-glasses* and the replica solution.

1.7.1 Random Spin-Glass

Consider a spin system. By a spin system, we mean here a system consisting of n degrees of freedom that we call "spins". We label each spin as x_i for $1 \leq i \leq n$. Each spin takes a value in the set $\{+1, -1\}$. Such spins are also known as Ising spins. The state of the system is denoted by $\underline{x} = (x_1, x_2, \ldots, x_n)$. The state of the system is also known as the configuration of the system. The energy of the state is given by the function or the Hamiltonian, $\mathcal{H}(\underline{x})$. Then the probability that the system is present in a particular state $\underline{x}$ is given by the Boltzmann distribution[5]

$$\mathbb{P}(\underline{x}) = \frac{1}{Z} e^{-\beta \mathcal{H}(\underline{x})}, \tag{1.19}$$

where β is the inverse temperature and

$$Z = \sum_{\underline{x}} e^{-\beta \mathcal{H}(\underline{x})}. \tag{1.20}$$

Z is known as the normalization factor or the partition function. Note that as the inverse temperature varies, the Boltzmann distribution varies from a uniform distribution ($\beta \to 0$), to concentrating on a few configurations which have the minimum energy, also known as ground states ($\beta \to +\infty$). The averages of observables w.r.t. the Boltzmann measure is usually denoted by the Gibbs bracket $\langle - \rangle$. Later in this section, we will consider systems with random "external" conditions. The average w.r.t. this randomness will be denoted as usual by $\mathbb{E}$. It is important to differentiate the two randomness (randomness w.r.t. Gibbs measure and randomness in the "external" conditions). It is clear that one cannot interchange the order of the two expectations $\mathbb{E}$ and $\langle - \rangle$. For taking the average of any observable, we first take the average w.r.t. the Gibbs measure, followed by the average w.r.t. the external conditions. As a consequence, we use the bracket notation to differentiate the randomness w.r.t. the Gibbs measure.

A fundamental quantity of any spin system is its free energy, given by

$$h_n(\beta) = \frac{1}{n} \ln Z.$$

[5]One of the main postulates of equilibrium statistical physics is that the average of the observables can be computed from the Boltzmann distribution.

The behavior of the system is completely determined by knowing its free energy. If, as a function of the inverse temperature β or some other parameter, the free energy is non-analytic, then one says that there is a phase transition in the behavior of the system. For example, we will see later that for LDPC codes, the free energy is represented by the conditional entropy. At the MAP threshold ϵ_{MAP} there is a phase transition in the system. The decoder is no longer able to perform successful decoding, with high probability, above ϵ_{MAP}, whereas for all noise values below ϵ_{MAP} the decoding is successful, w.h.p.

The Hamiltonian can take many forms. For example, $H(\underline{x}) = -\sum_i l_i x_i$ denotes a system without interactions between the particles. For this "free" model, the free-energy is easily evaluated by exchanging the sum and product in (1.20). Let us give some examples of models in statistical physics.

Perhaps the simplest mathematical model is the Ising model, wherein particles are imagined to be positioned on a cube in $\mathbb{Z}^d$ and have nearest neighbor interactions. The interactions between the particles are either ferromagnetic (attractive) or anti-ferromagnetic (repulsive). If all interactions are ferromagnetic, the Hamiltonian is given by

$$\mathcal{H}(\underline{x}) = -\sum_{(ij)} x_i x_j - l \sum_i x_i,$$

where (ij) are nearest neighbor pairs. For such a Hamiltonian, the probability is maximized for a configuration in which all the particles are "aligned" to each other. These lattice models are in general hard to solve (the exact solution to the free energy is not known for $d \geq 3$).

A solvable Ising model is the model on the complete graph (CG) (this model is sometimes called as the Curie-Weiss model) wherein the Hamiltonian is given by

$$\mathcal{H}_{\text{CG}}(\underline{x}) = -\frac{1}{n} \sum_{i<j} x_i x_j - l \sum_i x_i.$$

In words, the spins do not sit on a lattice and there is a ferro-magnetic interaction between all pairs of spins. This is an archetypal example of a *mean-field model*. The solution for the free energy of this model is given by

$$\lim_{n \to +\infty} h_n[\beta] = \max_{m \in [-1,+1]} h[m; \beta, l] \tag{1.21}$$

where $h[m; \beta, l] = \frac{-\beta}{2}(1 - m^2) + \beta l m - \frac{1+m}{2} \log(\frac{1+m}{2}) - \frac{1-m}{2} \log(\frac{1-m}{2})$. The maximum of $h[m; \beta, l]$ satisfies the equation

$$m_* = \tanh(\beta(m_* + l)).$$

Details on how this solution is obtained by mean-field analysis can be found in the recent book [74] and also in [68].

Our interest is in systems where the interactions between the particles are random. Such systems have also been modeled and studied in statistical physics.

A canonical model is the Edwards-Anderson (EA) model. In this model, the spins again sit on a d-dimensional lattice. However, now the interactions between the spins are random. The Hamiltonian of this model is given by

$$\mathcal{H}_{\mathrm{EA}}(\underline{x}) = -\sum_{(ij)} l_{ij} x_i x_j - l \sum_i x_i.$$

Above, the interactions l_{ij} are i.i.d. and distributed as $c(l_{ij})$. The typical choice for $c(l_{ij})$ is the Gaussian distribution.

The geometric structure (lattice) makes it hard to determine the free energy of the EA model. The mean-field simplification of this model is known as the Sherrington-Kirkpatrick (SK) model. The SK model is defined on a complete graph. The Hamiltonian of the SK model is given by

$$\mathcal{H}_{\mathrm{SK}}(\underline{x}) = \frac{-1}{\sqrt{n}} \sum_{i<j} l_{ij} x_i x_j - l \sum_i x_i.$$

with $\mathbb{P}(l_{ij}) = \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{l_{ij}^2}{2\epsilon^2}}$. We do not present here the replica solution for this model. The solution can be found in the book [71]. The important feature of the solution is that, the replica solution for the SK model is also a variational problem similar to (1.21) (although much more complicated). The SK model belongs the family of more general *fully-connected* spin-glass models, known as p-spin glass models.

We remark here that the replica method is based on mathematically unjustified manipulations, hence there is no guarantee that they provide the correct solution. However, it is believed that the predictions of replica theory are correct for mean-field models. Instances where the replica solution has been proved to give the correct answer includes the SK model. This was shown by Talagrand in [75] by using the second interpolation method of Guerra and Toninelli [31].

The non-triviality in analyzing spin-glasses is due to frustration (one cannot "satisfy" all the interactions simultaneously). Hence there are possibly many (exponential in n) states which have very similar energy. This results in a complicated energy landscape.

The fully-connected models are unrealistic, since they involve diverging number of interactions for each spin. A more realistic model is the *dilute spin-glass* model. In this model, the spins are assumed to sit on the nodes of a random graph, with finite average degrees. Many problems in computer and communication science can be modeled as a dilute spin-glass[6].

In the next section we show a bit more precisely, the mapping of the decoding problem of sparse graph codes to the language of statistical mechanics.

[6]Since the energy landscape of such problems is very complicated, it makes the search for optimal solutions or ground states very difficult.

1.7.2 LDPC Code as a Dilute Spin-Glass

Consider transmission over a general BMS channel, with transition probability given by $p_{Y|X}(y \mid x)$, using a fixed LDPC code or a LDPC code chosen u.a.r. from an ensemble of codes. Let $\underline{x}$ denote the transmitted codeword of block-length n and let $\underline{y}$ denote the received message. Let us re-write the a posteriori distribution $p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y})$ as a Boltzmann measure (see (1.19)). Recall from (1.6) that,

$$p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y}) = \frac{1}{|C| p_{\underline{Y}}(\underline{y})} \prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid x_i) \prod_{a=1}^{m} \mathbb{1}\Big(\prod_{k \in \partial a} x_k = +1 \Big).$$

Using

$$\begin{aligned} p_{\underline{Y}}(\underline{y}) &= \sum_{\underline{x}} \prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid x_i) p_{\underline{X}}(\underline{x}) \\ &= \frac{1}{|C|} \sum_{\underline{x}} \prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid x_i) \prod_{a=1}^{m} \mathbb{1}\Big(\prod_{k \in \partial a} x_k = +1 \Big) \end{aligned}$$

we get

$$p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y}) = \frac{1}{Z} \prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid x_i) \prod_{a=1}^{m} \mathbb{1}\Big(\prod_{k \in \partial a} x_k = +1 \Big), \tag{1.22}$$

where $Z = \sum_{\underline{x}} \prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid x_i) \prod_{a=1}^{m} \mathbb{1}(\prod_{k \in \partial a} x_k = +1)$. A simple calculation shows that

$$\mathbb{1}\Big(\prod_{k \in \partial a} x_k = +1 \Big) = \frac{(1 + \prod_{k \in \partial a} x_k)}{2} = \frac{(1 + x_{\partial a})}{2},$$

and

$$p_{Y_i|X_i}(y_i \mid x_i) = \frac{1}{2} p_{Y_i|X_i}(y_i \mid +1) \frac{1 + e^{-l_i}}{\cosh \frac{l_i}{2}} e^{\frac{l_i}{2} x_i}.$$

Substituting above in (1.22) we get

$$p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y}) = \frac{1}{Z} \sum_{\underline{x}} \prod_{a=1}^{m} \frac{(1 + x_{\partial a})}{2} \prod_{i=1}^{n} e^{\frac{l_i}{2} x_i}, \tag{1.23}$$

and $Z = \sum_{\underline{x}} \prod_{a=1}^{m} \frac{(1+x_{\partial a})}{2} \prod_{i=1}^{n} e^{\frac{l_i}{2} x_i}$. Realizing that $\frac{(1+x_{\partial a})}{2} = \lim_{l_a \to +\infty} e^{l_a (x_{\partial a} - 1)}$, we see that the a posteriori distribution for LDPC codes represents a spin-glass with the interactions between spins (code-bits) given by the code constraints and the external conditions given by the LLR values l_i, governed by the channel distribution law $c_L(l_i)$ (under the assumption that the all-one codeword is transmitted), for $1 \leq i \leq n$.

Free Energy

In this section we formulate the conditional entropy as free energy. Details of the computation can be found in [29] and [65]. In the Gibbs measure (1.23), the channel randomness is contained in the vector of LLR values $\underline{l} = (l_1, l_2, \ldots, l_n)$. As a result, we replace the expectation $\mathbb{E}_{\underline{Y}|\underline{1}}$ with $\mathbb{E}_{\underline{l}}$ in (1.14). Recall that the distribution of each l_i under the assumption of the all-one codeword transmission is denoted by $c_L(l_i)$. For the ease of exposition we use $\mu(\underline{x})$ to denote the aposteriori distribution $p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y})$.

Using (1.23), we can write the conditional entropy as

$$\begin{aligned}
H(\underline{X} \mid \underline{Y}) &\overset{(1.14)}{=} -\mathbb{E}_{\underline{l}}\Big[\sum_{\underline{x}} \mu(\underline{x}) \ln \mu(\underline{x})\Big] \\
&= -\mathbb{E}_{\underline{l}}\Big[\sum_{\underline{x}} \mu(\underline{x}) \ln \Big(\frac{1}{Z}\prod_{a=1}^{m}\frac{1}{2}(1+x_{\partial a})\prod_{i=1}^{n} e^{\frac{l_i}{2}x_i}\Big)\Big] \\
&\overset{(a)}{=} \mathbb{E}_{\underline{l}}[\ln Z] - \mathbb{E}_{\underline{l}}\Big[\sum_{\underline{x}} \mu(\underline{x}) \ln \Big(\prod_{a}\frac{1}{2}(1+x_{\partial a})\Big)\Big] \\
&\qquad - \mathbb{E}_{\underline{l}}\Big[\sum_{\underline{x}} \mu(\underline{x}) \ln \Big(\prod_{i=1}^{n} e^{\frac{l_i}{2}x_i}\Big)\Big] \\
&\overset{(b)}{=} \mathbb{E}_{\underline{l}}[\ln Z] - \mathbb{E}_{\underline{l}}\Big[\sum_{\underline{x}} \mu(\underline{x}) \sum_{i=1}^{n}\frac{l_i}{2}x_i\Big] \\
&\overset{(c)}{=} \mathbb{E}_{\underline{l}}[\ln Z] - \sum_{i=1}^{n}\int_{-\infty}^{+\infty} dl_i c_L(l_i)\frac{l_i}{2}.
\end{aligned} \tag{1.24}$$

Above, equality (a) follows from the fact that Z does not depend on $\underline{x}$. Since $\mu(\underline{x}) = 0$ for binary strings which are not codewords, i.e., $n-$tuples for which $\prod_a \frac{1}{2}(1 + x_{\partial a}) = 0$, and since for binary strings which are codewords we have $\prod_a \frac{1}{2}(1 + x_{\partial a}) = 1$, we must have $\mu(\underline{x}) \ln \prod_a \frac{1}{2}(1 + x_{\partial a}) = 0$ and (b) follows. The equality (c) is shown in [65].

Thus we see that the evaluation of the conditional entropy essentially boils down to the computation of the logarithm of the partition function, or the free energy.

1.7.3 LDGM Code as a Dilute Spin-Glass

Now consider transmission over a general BMS channel using an LDGM code, perhaps picked u.a.r. from an ensemble of codes. Let $\underline{u}$ denote the information word, $\underline{x}$ denote the codeword and $\underline{y}$ denote the received word. The information word length equals m and the block-length is n. The a posteriori distribution,

$p_{\underline{U}|\underline{Y}}(\underline{u} \mid \underline{y})$ is

$$\begin{aligned} p_{\underline{U}|\underline{Y}}(\underline{u} \mid \underline{y}) &= \frac{p_{\underline{Y}|\underline{U}}(\underline{y} \mid \underline{u})}{\sum_{\underline{u}} p_{\underline{Y}|\underline{U}}(\underline{y} \mid \underline{u})} \\ &= \frac{\prod_{i=1}^{n} p_{Y_i|X_i}(y_i \mid \prod_{a \in \partial i} u_a)}{\sum_{\underline{u}} p_{Y_i|X_i}(y_i \mid \prod_{a \in \partial i} u_a)}. \end{aligned}$$

Again we have

$$p_{Y_i|X_i}(y_i \mid \prod_{a \in \partial i} u_a) = p_{Y_i|X_i}(y_i \mid +1) \frac{1 + e^{-l_i}}{2 \cosh \frac{l_i}{2}} e^{\frac{l_i}{2} u_{\partial i}}$$

where we use $u_{\partial i}$ to denote $\prod_{a \in \partial i} u_a$. Putting all together we get

$$p_{\underline{U}|\underline{Y}}(\underline{u} \mid \underline{y}) = \frac{1}{Z} \prod_{i=1}^{n} e^{\frac{l_i}{2} u_{\partial i}}, \tag{1.25}$$

where $Z = \sum_{\underline{u}} \prod_{i=1}^{n} e^{\frac{l_i}{2} u_{\partial i}}$. Thus we conclude that the LDGM coded system represents a spin-glass with the external conditions, l_i distributed by the channel law $c_L(l)$, under the assumption that the all-one codeword is transmitted. It is not hard to see that a generator-matrix code will always have a positive bit error rate for any channel noise. To see this, consider transmission over the BEC with the erasure probability given by $\epsilon \in (0, 1]$. Suppose that all the code-bits neighboring an information bit (a code-bit is a neighbor of an information bit if the information bit participates in the generation of the code-bit) are erased. Then it is impossible to estimate the information bit perfectly. The probability that this happens is given by ϵ^l, where l is the degree of the information bit. Nevertheless, LDGM codes are still useful in other communication scenarios like source coding and they are mathematically easier to analyze. Moreover, most of our ideas can be clearly demonstrated in this case. Intuitively we see this as follows. The a posteriori distribution for the LDGM case differs from the LDPC case from the absence of the term $\frac{1+u_{\partial i}}{2}$. This term is usually called as a "hard constraint" and makes the analysis of LDPC codes more complicated as opposed to the LDGM codes. In the language of statistical mechanics, LDPC codes correspond to *low temperature* or $\beta \to +\infty$ (for the parity-check constraints) whereas LDGM codes correspond to *high temperature* or $\beta \ll +\infty$ (for high noise).

Similar to LDPC codes we can show that the conditional entropy in the case of LDGM can be written as

$$H(\underline{U} \mid \underline{Y}) = \mathbb{E}_{\underline{l}}\big[\ln Z\big] - \sum_{i=1}^{n} \mathbb{E}_{l_i}\left[\frac{l_i}{2}\right].$$

Thus in this case also, by evaluating the free energy, one can determine the conditional entropy.

1.7.4 Replica Solution for the Conditional Entropy

In this section we present the solution of the replica method applied to the conditional entropy. We do not present the derivation here. The interested reader can refer to [28] for a detailed analysis.

Let V be a random variable with density $d_V(v)$ satisfying the symmetry condition $d_V(v) = e^v d_V(-v)$. Also, let

$$\begin{aligned} U &= 2\tanh^{-1}\left[\prod_{i=1}^{k-1}\tanh\frac{V_i}{2}\right] \qquad \text{LDPC}, \\ U &= 2\tanh^{-1}\left[\tanh\frac{l}{2}\prod_{i=1}^{k-1}\tanh\frac{V_i}{2}\right] \qquad \text{LDGM}, \end{aligned} \tag{1.26}$$

where V_i are i.i.d copies of V and k is the (random with distribution given by $P(x)$) degree of a check node. We denote by U_b, $b = 1, ..., \ell$ i.i.d. copies of U where ℓ is the (random with distribution given by $\Lambda(x)$) degree of variable nodes. Recall that l is the channel LLR .

Notice that in the belief propagation (BP) decoding algorithm U appears as the check-to-variable node message and V appears as the variable-to-check node message. Define the functional[7] (we view it as a functional of the probability distribution d_V)

$$\begin{aligned} h_{RS}[d_V;\Lambda,P] =& \mathbb{E}_{l,d,U_b}\left[\ln\left(e^{\frac{l}{2}}\prod_{b=1}^{\ell}(1+\tanh\frac{U_b}{2})+e^{-\frac{l}{2}}\prod_{b=1}^{\ell}(1-\tanh\frac{U_b}{2})\right)\right] \\ &+\frac{\Lambda'(1)}{P'(1)}\mathbb{E}_{k,V_i}\left[\ln(1+\prod_{i=1}^{k}\tanh\frac{V_i}{2})\right] \\ &-\Lambda'(1)\mathbb{E}_{V,U}\left[\ln(1+\tanh\frac{V}{2}\tanh\frac{U}{2})\right]-\frac{\Lambda'(1)}{P'(1)}\ln 2. \end{aligned}$$

The above functional, $h_{RS}[d_V;\Lambda,P]$, is known as the replica solution (or functional). The main conjecture of statistical physics is

Conjecture 1.1. *Consider communication over any general BMS channel using any standard LDPC$(\Lambda(x),P(x))$ or LDGM$(\Lambda(x),P(x))$ code ensemble. Then*

$$\lim_{n\to+\infty}\mathbb{E}_C[h_n] = \sup_{d_V} h_{RS}[d_V;\Lambda,P].$$

Note that in this case also the replica solution is a variational formula similar to the solution of the Curie-Weiss model (c.f. (1.21)).

[7] The subscript *RS* stands for "replica symmetric" because this functional has been obtained from the replica symmetric ansatz for an appropriate spin glass, see for example [65], [66]

It is not hard to see that the critical points of $h_{RS}[d_V; \Lambda, P]$ satisfies the density evolution equations [29]. This observation provides a bridge between the MAP and BP analysis of sparse graph codes. It also provides another interpretation to the work of Measson et al. in [24]. More precisely, consider the Figure 1.6. In this figure, the average MAP-GEXIT equals the average BP-GEXIT till the MAP threshold, ϵ_{MAP}. Below the BP threshold, ϵ_{BP}, both of them are trivially equal. In between the two thresholds, there is a difference between the average BP and MAP estimates. One can explain this difference via the replica solution as follows. It is not hard to show, for the case of transmission over the BEC, that the replica solution between ϵ_{BP} and ϵ_{MAP} is maximized at the stable fixed-point of density evolution *not obtained* by BP. On the other hand, for all the remaining noise values, the replica solution is maximized at the BP fixed-point of the density evolution equation. Hence for such noise values, the average BP and MAP estimates are equal.

Many discrete optimization problems in computer science like the random k-satisfiability, graph partitioning, vertex covering, graph matching have been analyzed by methods of statistical physics [76], [71], [74], [77], [78], [79], [80], [81], [82]. The discrete optimization problems mentioned above are also modeled by the dilute spin-glass model. One is then interested in the studying the behavior of the system under certain "external" conditions. For example, in the random k-satisfiability problem one is interested in knowing if the given Boolean formula is satisfiable when the ratio of number of constraints to variable is changed [77], [76], [83]. Replica method is then employed to compute the free energy of the system. This allows to make predictions on the values of phase transition thresholds [77]. For example, in the random 3-satisfiability problem, the ratio of number of constraints to variables greater than $\alpha_c \approx 4.27$, signals the onset of the unsatisfiable phase, with high probability.

The process of rigorization of these remarkable yet unproven results, for dilute spin-glasses, was undertaken recently. In [29], Montanari proved that the replica solution is a lower bound, for any symmetric $d_V(v)$, on the asymptotic per-bit conditional entropy when transmitting over any BMS channel. He shows this result only for a special class of sparse graph codes. The requirement of the class of codes is that the degree distribution $P(x)$ needs to be convex. In [84], Macris removed this restriction of convexity and showed that for Poisson-LDPC code ensembles (see Section 2.5 for the definition), the replica solution is a lower bound on the conditional entropy, when transmitting over the BIAWGNC. In the next chapter we will extend this result to a more general class of BMS channels and *any* standard LDPC or LDGM code ensemble. In [79], Franz and Leone also use the interpolation method to show that the replica solution is a tight lower bound to the free energy of finite-temperature random k-satisfiability problem. We also mention the work of Talagrand in [85], where he proves the replica solution is exact for the high temperature random k-satisfiability problem, using the decay of correlations. We point out that the techniques used to show decay of correlations in [85] are different from those used in this thesis.

In the next section we delineate the contribution of this thesis.

1.8 Contributions and Organization of this Thesis

The focus of this thesis is on the analysis of sparse graph codes using methods from statistical physics. One of the main results in this thesis is the optimality of the BP estimate in appropriate ranges of noise, with high probability, when transmitting over general BMS channels using either LDGM or LDPC code ensembles. Another interpretation of this result is that the replica predictions of statistical physics are correct. We show this by employing various methods of statistical mechanics like the *interpolation method* and the *cluster expansion* techniques.

We also develop simple sum-rules which relate the MAP and BP estimates via correlations between code-bits. Another main result of this thesis is that the average (over the noise realizations) correlation between any two code-bits decays exponentially with their graph distance, when we consider transmission over general BMS channels using either LDPC or LDGM code ensembles with bounded maximum degrees. This has a pleasing intuitive consequence that the "far-away" code-bits do not influence each other. We then use the decay of correlations to show the correctness of BP estimate, with high probability. In other words, we show that the asymptotic average MAP-GEXIT can be computed via the density evolution equations!

The thesis is organized as follows:

In Chapter 2 we show that the replica solution of the average per-bit conditional entropy is a lower bound when we consider transmission over general BMS channels using *any* sparse graph code ensemble. We use the *interpolation method*, first developed by Guerra [86], [30] in the context of spin-glass theory and used in the context of coding by Montanari [29]. In the proofs we employ tools of statistical physics, namely, the Griffiths-Kelly-Sherman inequality and the Nishimori identities.

In Chapter 3, we show that the replica solution for the conditional entropy is exact when we consider transmission over the BEC. Although this result is not new, the method is completely different and utilizes the *second interpolation method* developed by Guerra and Toninelli [31]. We stress that the methods used are non-combinatorial as opposed to the existing techniques. Thus there would seem to be some hope to extend this method to the general channels.

In Chapter 4 we study the correlations of the MAP decoder, between any two code-bits of a fixed LDGM code or a code picked from a special class of LDPC code ensemble. We show that if the channel noise is high enough, then the average correlation between two code-bits decays exponentially with their graph distance. The channels that we consider belong to a fairly general

class of BMS channels, which include the important case of the BIAWGNC. An important consequence of the main result is that the asymptotic average MAP-GEXIT function can be evaluated by density evolution equations for the BP decoder. Another interpretation of this result is that the replica computations are exact. The techniques we use are the high-temperature cluster expansions of statistical physics.

In Chapter 5 we look at the much tougher case of LDPC codes at low noise. This represents truly a low-temperature spin-glass for which there are almost no rigorous results in the literature, according to our knowledge. More precisely, we consider the transmission over a class of general BMS channels, which includes the BIAWGNC. We use *any* fixed LDPC code, with bounded maximum degrees, for transmission. We again show that the average correlations of the MAP decoder, between any two code-bits decays exponentially with their graph distance, if the channel noise is low enough. As a result we obtain the exactness of the BP estimate, with high probability, in the low noise regime. For regular LDPC code ensembles, the content of the previous conclusion might be trivial, since the BP estimate is zero. Nevertheless for LDPC code ensemble, with a non-zero fraction of degree one variable nodes, the content of the result is non-trivial. To prove these results we combine the duality of LDGM-LDPC codes with a suitable cluster expansion technique of statistical physics.

Lower Bounds on Conditional Entropy: The Interpolation Method

2

2.1 Introduction and Main Results

The replica solution to the normalized conditional entropy introduced in Section 1.7 is conjectured to be exact. In this chapter we give a partial proof of this conjecture. More precisely, we prove that the replica solution is a lower bound for transmission using LDPC or LDGM code ensembles, with *any* edge degree distributions $(\lambda(x), \rho(x))$.

Our proof holds for the BEC(ϵ) and BIAWGNC(ϵ) for *all* values of the channel noise parameter ϵ. Further, for general BMS(ϵ) channels, the proof holds in the high noise regime. Even though all the results presented in this chapter hold for LDPC as well as LDGM code ensembles, we restrict ourselves to LDPC code ensembles for the sake of ease of exposition.

A promising approach towards a general proof of the exactness of the replica formulas is the so-called *interpolation method* first developed in the context of the Sherrington-Kirkpatrick model [31], [86], [87] for magnetic systems. Consider an LDPC(n, Λ, P) ensemble where $\Lambda(x) = \sum_l \Lambda_l x^l$ and $P(x) = \sum_k P_k x^k$ are the variable and check degree distributions from the node perspective. We will always assume that the maximal degrees are finite. Montanari [29] (see also the related work of Franz-Leone [79] and Talagrand- Pachenko [88]) developed the interpolation method for such a coding system and derived a lower bound[1] for the conditional entropy for ensembles with any polynomial $\Lambda(x)$, but with $P(x)$ *restricted to be convex for* $-e \leq x \leq e$ *(in particular if the check degree is constant this means it has to be even).*

In the present chapter *we drop the convexity requirement for* $P(x)$ in the cases of the BEC and BIAWGNC with any noise level and in the case of

[1] Called "variational bound" in the sequel.

general BMS channels in a high noise regime. In other words we prove that the replica solution is a lower bound on the per-bit conditional entropy, in the limit $n \to +\infty$, for *any standard regular* (so odd degrees are allowed) or *irregular* code ensemble.

We introduce a new tool in the form of a relationship between the second derivative of the conditional entropy with respect to the noise and correlations functions of code-bits. These correlation functions are shown to be intimately related to the mutual information between two code-bits. The formulas are somewhat similar to those for GEXIT functions [23] which relate the first derivative of conditional entropy to soft bit estimates. By combining these relations with the interpolation method we are able to control the fluctuations of the so-called "overlap" parameters. This part of our analysis is crucial for proving the general lower bound on the per-bit conditional entropy and relies heavily on channel symmetry. The next section states our main result, namely the lower bounding of the conditional entropy.

2.1.1 Variational Bound on the Conditional Entropy

Let $p_{Y|X}(y \mid x)$ be the transition probability of a BMS(ϵ) channel where ϵ is the noise parameter (understood to vary in the appropriate range, see 1.2). We will work in terms of both the LLR l (see Definition 1.3) and difference variables

$$t = p_{Y|X}(y \mid +1) - p_{Y|X}(y \mid -1) = \tanh \frac{l}{2}.$$

Recall that $c_L(l)$ denotes the distribution of l, assuming that the all-one codeword is transmitted. It will also be convenient to use $c_D(t)$ for the distributions of t, assuming that the all-one codeword is transmitted (that is to say that $c_L(l)dl = c_D(t)dt = p_{Y|X}(y \mid +1)dy$).

For general BMS(ϵ) channels, we show that the replica solution is a lower bound for *high* noise values.

Definition 2.1 (High Noise Regime: H)**.** *For $p \in \mathbb{N}$ let*

$$m_0^{(2p)} = \mathbb{E}[t^{2p}], \qquad m_1^{(2p)} = \frac{d}{d\epsilon}\mathbb{E}[t^{2p}], \quad \textit{and} \quad m_2^{(2p)} = \frac{d^2}{d\epsilon^2}\mathbb{E}[t^{2p}]. \tag{2.1}$$

We say that a BMS(ϵ) *channel is in the high noise regime if the series expansions*

$$\sum_{p\geq 0}(p+1)m_0^{(2p)}, \qquad \sum_{p\geq 1}\Big(\frac{5}{2}\Big)^{2p}|m_1^{(2p)}|, \quad \textit{and} \qquad \sum_{p\geq 1}\frac{|m_2^{(2p)}|}{2p(2p-1)}, \tag{2.2}$$

are convergent and if

$$(\sqrt{2}-1)\Big(\frac{5}{2}\Big)^{2}|m_1^{(2)}| > \sum_{p\geq 2}\Big(\frac{5}{2}\Big)^{2p}|m_1^{(2p)}|. \tag{2.3}$$

Example 2.1 (High Noise BSC(ϵ)). *For the BSC(ϵ) we have* $l \in \{\ln\frac{1-\epsilon}{\epsilon}, -\ln\frac{1-\epsilon}{\epsilon}\}$. *Hence* $\tanh\frac{l}{2} = \left(\frac{e^{l/2}-e^{-l/2}}{e^{l/2}+e^{-l/2}}\right) \in \{(1-2\epsilon), -(1-2\epsilon)\}$. *Thus we have* $m_0^{(2p)} = \mathbb{E}[t^{2p}] = (1-2\epsilon)^{2p}$ *and consequently,* $m_1^{(2p)} = -4p(1-2\epsilon)^{2p-1}$ *and* $m_2^{(2p)} = 8p(2p-1)(1-2\epsilon)^{2p-2}$. *It is easy to see that for* $\epsilon < 1/2$, *the first series* $\sum_{p\geq 0}(p+1)m_0^{(2p)}$ *is convergent. From the expression for* $|m_2^{(2p)}|$ *above, we also have that* $\sum_{p\geq 1}\frac{|m_2^{(2p)}|}{2p(2p-1)}$ *equals* $4\sum_{p\geq 1}(1-2\epsilon)^{2p-2}$. *The last expression also converges for any* $\epsilon < 1/2$. *Let us now consider the middle series. Rearranging, the series is given by*

$$\sum_{p\geq 1}\Big(\frac{5}{2}\Big)^{2p}|m_1^{(2p)}| = \frac{4}{1-2\epsilon}\Big(\frac{5}{2}(1-2\epsilon)\Big)^2\sum_{p\geq 1}p\Big(\Big[\frac{5}{2}(1-2\epsilon)\Big]^2\Big)^{p-1},$$

which converges if and only if $\epsilon > \frac{3}{10} = 0.3$. *The condition* (2.3) *implies*

$$(\sqrt{2}-1)\Big(\frac{5}{2}\Big)^2 4(1-2\epsilon) > \sum_{p\geq 2}\Big(\frac{5}{2}\Big)^{2p}4p(1-2\epsilon)^{2p-1}.$$

A little algebra shows that this condition is equivalent to

$$\sqrt{2} > \sum_{p\geq 1}p\Big(\Big(\frac{5}{2}\Big)^2(1-2\epsilon)^2\Big)^{p-1} = \frac{1}{\Big(1-\Big(\frac{5}{2}\Big)^2(1-2\epsilon)^2\Big)^2}.$$

Thus for $\epsilon > \frac{1}{2} - \frac{1}{5}\sqrt{1-\frac{1}{2^{1/4}}} \approx 0.42$, *all the high noise conditions* (2.2), (2.3) *are satisfied.*

Any channel *with bounded log-likelihoods*, such that the support of $c_L(l)$ shrinks to zero as $\epsilon \to \epsilon_{\max}$, satisfies H for a regime of sufficiently high noise.

Note that the BEC(ϵ) which has mass at $l = +\infty$ does not satisfy this condition since $\mathbb{E}[t^{2p}] = 1-\epsilon$. Also, for the BIAWGNC, the second sequence in (2.2) does not converge, because $|m_1^{(2p)}|$ does not decay fast enough. One can check that $|m_1^{(2p)}|$, for large p, decays like $\Theta\big(\frac{1}{(2p)^{\ln 2p}}\big)$, which does not kill the exponential growth of $\big(\frac{5}{2}\big)^{2p}$ for large p and the second series in (2.2) fails to converge for any value of the noise variance.

However, for these two channels the variational bound holds for all noise levels. For the BEC, the positivity of the support LLR values bails us out and the gaussian integration by parts (see Appendix 2.A.3) comes to the rescue in the case of the BIAWGNC.

Recall that $h_n = \frac{1}{n}H(\underline{X} \mid \underline{Y})$ denotes the per-bit conditional entropy and $h_{RS}[d_V; \Lambda, P]$ denotes the replica solution (see Section 1.7 for details). We have the following theorem.

Theorem 2.1 (Variational Bound). *Assume communication using a standard irregular code ensemble* $\mathcal{C} = \mathrm{LDPC}(n, \Lambda, P)$ *through a BEC(ϵ) or BIAWGNC(ϵ)*

with any noise level or a BMS(ϵ) channel satisfying H. For all ϵ in the above ranges we have,

$$\liminf_{n\to+\infty} \mathbb{E}_{\mathcal{C}}[h_n] \geq \sup_{d_V} h_{RS}[d_V;\Lambda,P],$$

where the supremum is over symmetric distributions $d_V(v) = e^{-v}d_V(-v)$.

Let us note that this theorem already appears in [84] for the special case of the BIAWGNC and Poisson-LDPC code ensemble ($\Lambda(x) = e^{\mathbf{l}_{\text{avg}}(x-1)}$).

2.1.2 Second Derivative of the Conditional Entropy

Our proof of the variational bound uses integral formulas for the first and second derivatives of $\mathbb{E}_{\mathcal{C}}[h_n]$ with respect to the noise parameter ϵ. The first derivative formula is given in Theorem 1.2. We provide the proof of this theorem in Section 2.3. The second derivative formula is also valid for any fixed linear code and not necessarily restricted to sparse graph codes. To give the formulation for a fixed linear code it is convenient to introduce a noise vector $\underline{\epsilon} = (\epsilon_1, ..., \epsilon_n)$ and a BMS($\underline{\epsilon}$) channel with noise level ϵ_i when bit x_i is sent. When all noise levels are set to the same value ϵ we get back our original channel BMS(ϵ). The distributions of the likelihoods l_i and difference domain t_i depend on ϵ_i. In order to keep the notation simple we do not explicitly indicate the ϵ_i dependence and still denote them as $c_L(l_i)$ and $c_D(t_i)$ respectively.

For ease of notation we drop "MAP" and the dependence on $\underline{Y}$ from the soft bit MAP estimate in (1.5). Thus

$$\phi_i = \ln\left[\frac{p_{X_i|\underline{Y}}(+1 \mid \underline{y})}{p_{X_i|\underline{Y}}(-1 \mid \underline{y})}\right].$$

For the purpose of our exposition, we introduce the soft MAP estimates of bit X_i in the difference domain,

$$T_i = p_{X_i|\underline{Y}}(+1 \mid \underline{y}) - p_{X_i|\underline{Y}}(-1 \mid \underline{y}) = \tanh\left(\frac{\phi_i}{2}\right). \tag{2.4}$$

We also introduce the soft estimate for the modulo-2 sum $X_i \oplus X_j$, again, in both the $L-$domain and the difference domain. Thus

$$\begin{aligned}
\phi_{ij} &= \ln\left[\frac{p_{X_i\oplus X_j|\underline{Y}}(+1 \mid \underline{y})}{p_{X_i\oplus X_j|\underline{Y}}(-1 \mid \underline{y})}\right], \\
T_{ij} &= p_{X_i\oplus X_j|\underline{Y}}(+1 \mid \underline{y}) - p_{X_i\oplus X_j|\underline{Y}}(-1 \mid \underline{y}) = \tanh\left(\frac{\phi_{ij}}{2}\right).
\end{aligned} \tag{2.5}$$

Recall that the notations $\underline{v}^{\sim i}$ (resp. $\underline{v}^{\sim ij}$) mean that component v_i (resp. v_i and v_j) are omitted from the vector $\underline{v}$.

Lemma 2.1 (Correlation Formula)**.** *For any* BMS($\underline{\epsilon}$) *channel and any fixed linear code we have*

$$\begin{aligned}\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i \partial \epsilon_j} =& \delta_{ij} \int_{-1}^{+1} dt_i \frac{\partial^2 c_D(t_i)}{\partial \epsilon_i^2} g_1(t_i) \\ &+ (1-\delta_{ij}) \int_{-1}^{+1}\int_{-1}^{+1} dt_i dt_j \frac{\partial c_D(t_i)}{\partial \epsilon_i}\frac{\partial c_D(t_j)}{\partial \epsilon_j} g_2(t_i,t_j) \end{aligned} \tag{2.6}$$

with

$$g_2(t_i,t_j) = \mathbb{E}_{\underline{t}\sim ij}\left[\ln\left(\frac{1-t_iT_i-t_jT_j+t_it_jT_{ij}}{1-t_iT_i-t_jT_j+t_it_jT_iT_j}\right)\right]$$

and

$$g_1(t_i) = -\mathbb{E}_{\underline{t}\sim i}\left[\ln\left(\frac{1-t_iT_i}{1-t_i}\right)\right],$$

and where $\delta_{ij}=1$ *if* $i=j$ *and is zero else. Here* $\mathbb{E}_{\underline{t}\sim i}$ *denotes the expectation w.r.t. all the random variables* $\underline{t}=(t_1,t_2,\ldots,t_n)$ *except for* t_i*. Also,* $\mathbb{E}_{\underline{t}\sim ij}$ *denotes the expectation w.r.t. all* $\underline{t}$ *except* t_i *and* t_j*.*

For the case of a BMS(ϵ) channel and a linear code ensemble, using the symmetry of sites argument, the second derivative of the average per-bit conditional entropy is given by,

$$\begin{aligned}\frac{d^2\mathbb{E}_{\mathcal{C}}[h_n]}{d\epsilon^2} =& \int_{-1}^{+1} dt_i \frac{\partial^2 c_D(t_i)}{\partial \epsilon^2}\mathbb{E}_{\mathcal{C}}[g_1(t_i)] \\ &+ \sum_{\substack{j=1\\ j\neq i}}^{n}\int_{-1}^{+1}\int_{-1}^{+1} dt_i dt_j \frac{\partial c_D(t_i)}{\partial \epsilon}\frac{\partial c_D(t_j)}{\partial \epsilon}\mathbb{E}_{\mathcal{C}}[g_2(t_i,t_j)]. \end{aligned} \tag{2.7}$$

For the BEC these formulas simplify to

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial \ln \epsilon_i \partial \ln \epsilon_j} = (1-\delta_{ij})\ln 2 \mathbb{E}_{\underline{t}}\big[T_{ij}-T_iT_j\big],$$

and

$$\frac{d^2\mathbb{E}_{\mathcal{C}}[h_n]}{d(\ln\epsilon)^2} = \ln 2 \sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{E}_{\mathcal{C},\underline{t}}\big[T_{ij}-T_iT_j\big]. \tag{2.8}$$

For the BIAWGNC we have

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i^{-2} \partial \epsilon_j^{-2}} = \frac{1}{2}\mathbb{E}_{\underline{t}}[\big(T_{ij}-T_iT_j\big)^2], \tag{2.9}$$

and

$$\frac{d^2\mathbb{E}_{\mathcal{C}}[h_n]}{d(\epsilon^{-2})^2} = \frac{1}{2}\sum_{j=1}^{n}\mathbb{E}_{\mathcal{C},\underline{t}}[\big(T_{ij} - T_iT_j\big)^2]. \tag{2.10}$$

We take the derivatives with respect to the logarithm of the noise in (2.8) and the inverse of the noise variance in (2.10) to get a clean expression for the second derivative of the conditional entropy in terms of the correlation. We stress that it serves no other purpose but convenience of exposition. We will provide a proof of Lemma 2.1 along with the simplifications in the case of BEC and BIAWGNC in Section 2.4. Formulas (2.8) and (2.10) involve the "correlation" $(T_{ij} - T_iT_j)$ for bits X_i and X_j. The general formula (2.7) can also be recast in terms of powers of such correlations by expanding the logarithm (see Section 2.4). Loosely speaking, in the infinite block length limit $n \to +\infty$, the second derivative will be well defined only if the correlations have sufficient decay with respect to the graph distance (the minimal length among all paths joining i and j on the Tanner graph). We expect the first derivative to be smooth everywhere except at the phase transition points. Thus we expect good decay properties for all noise levels except at the phase transition thresholds where, in the limit $n \to +\infty$, the first derivative generally has bounded discontinuities. Thus the second derivative cannot be uniformly bounded in n at such discontinuities. Before going on to prove the correlation formula and the variational bound we take a slight detour and show how the correlation between bits conveys the mutual information between them, implying that if the correlation is weak then the two bits are weakly dependent.

2.1.3 Relationship to Mutual Information

The correlation $(T_{ij} - T_iT_j)$ is a measure of the independence of two code-bits, thus it is natural to expect that it is related to the mutual information $I(X_i; X_j \mid \underline{Y})$. We do not pursue this issue in all details because it will not be used in the rest of the thesis, but wish to briefly state the main relations which follow naturally form the previous formulas. See [89] for related discussions.

The BEC($\underline{\epsilon}$). Take $i \neq j$. The chain rule implies $H(\underline{X} \mid \underline{Y}) = H(X_iX_j \mid \underline{Y}) + H(\underline{X}^{\sim ij} \mid X_iX_j\underline{Y})$. Given X_i and X_j we do not require Y_i, Y_j for the estimation and hence we have $H(\underline{X}^{\sim ij} \mid X_iX_j\underline{Y}) = H(\underline{X}^{\sim ij} \mid X_iX_j\underline{Y}^{\sim ij})$. Further, since $H(\underline{X}^{\sim ij} \mid X_iX_j\underline{Y}^{\sim ij})$ does not depend on ϵ_i, ϵ_j we have

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial\epsilon_i\partial\epsilon_j} = \frac{\partial^2 H(X_iX_j \mid \underline{Y})}{\partial\epsilon_i\partial\epsilon_j}.$$

Since we are transmitting over an erasure channel, it is not difficult to see that the conditional entropy is non-zero except when Y_i, Y_j are perfectly known.

Thus,

$$H(X_iX_j \mid \underline{Y}) = \epsilon_i\epsilon_j H(X_iX_j \mid \underline{Y}^{\sim ij}) + \epsilon_i(1-\epsilon_j)H(X_i \mid X_j\underline{Y}^{\sim ij}) + (1-\epsilon_i)\epsilon_j H(X_j \mid X_i\underline{Y}^{\sim ij}).$$

In the above expression whenever Y_i (resp. Y_j) is known, we have uncertainty only in X_j (resp. X_i) and we replace Y_i (resp. Y_j) with X_i (resp. X_j). Also the three conditional entropies above are independent of the channel parameters ϵ_i and ϵ_j. Thus,

$$\begin{aligned}\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial\epsilon_i\partial\epsilon_j} &= H(X_iX_j \mid \underline{Y}^{\sim ij}) - H(X_i \mid X_j\underline{Y}^{\sim ij}) - H(X_j \mid X_i\underline{Y}^{\sim ij}) \\ &\overset{\text{Using } H(U,V)=H(U)+H(V|U)}{=} H(X_j \mid \underline{Y}^{\sim ij}) - H(X_j \mid X_i\underline{Y}^{\sim ij}) \\ &= I(X_i; X_j \mid \underline{Y}^{\sim ij}) = \frac{1}{\epsilon_i\epsilon_j} I(X_i; X_j \mid \underline{Y}).\end{aligned}$$

Summarizing, we have obtained for $i \neq j$,

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial \ln\epsilon_i \partial \ln\epsilon_j} = I(X_i; X_j \mid \underline{Y})$$

We have previously shown (see (2.8)) that the second derivative above, equals the correlation between code-bits i and j.

The BIAWGNC($\underline{\epsilon}$). Take $i \neq j$. From (2.5) we note that

$$\begin{aligned}T_{ij} &= p_{X_i,X_j|\underline{Y}}(+1,+1 \mid \underline{y}) + p_{X_i,X_j|\underline{Y}}(-1,-1 \mid \underline{y}) \\ &\quad - p_{X_i,X_j|\underline{Y}}(+1,-1 \mid \underline{y}) - p_{X_i,X_j|\underline{Y}}(-1,+1 \mid \underline{y}) \\ &= \sum_{x_i,x_j\in\{\pm1\}} x_ix_j p_{X_i,X_j|\underline{Y}}(x_i,x_j \mid \underline{y}),\end{aligned}$$

from which it follows that

$$\begin{aligned}(T_{ij} - T_iT_j)^2 &= \Big(\sum_{x_i,x_j\in\{\pm1\}} x_ix_j\Big(p_{X_i,X_j|\underline{Y}}(x_i,x_j \mid \underline{y}) - p_{X_i|\underline{Y}}(x_i \mid \underline{y})p_{X_j|\underline{Y}}(x_j \mid \underline{y})\Big)\Big)^2 \\ &\leq \Big(\sum_{x_i,x_j\in\{\pm1\}} \Big|p_{X_i,X_j|\underline{Y}}(x_i,x_j \mid \underline{y}) - p_{X_i|\underline{Y}}(x_i \mid \underline{y})p_{X_j|\underline{Y}}(x_j \mid \underline{y})\Big|\Big)^2 \\ &\leq 4\sum_{x_i,x_j} \Big|p_{X_i,X_j|\underline{Y}}(x_i,x_j \mid \underline{y}) - p_{X_i|\underline{Y}}(x_i \mid \underline{y})p_{X_j|\underline{Y}}(x_j \mid \underline{y})\Big|^2,\end{aligned}$$

where in the last inequality we use $(a+b+c+d)^2 \leq 4(a^2+b^2+c^2+d^2)$. Applying the Pinsker inequality [90]

$$\frac{1}{2}\sum_x |P(x) - Q(x)|^2 \leq D(P\|Q),$$

where $D(P\|Q)$ is the Kullback-Leibler divergence of the two distributions $P = p_{X_i,X_j|\underline{y}}$ and $Q = p_{X_i|\underline{y}} p_{X_j|\underline{y}}$. We get for $i \neq j$

$$(T_{ij} - T_i T_j)^2 \leq 8I(X_i; X_j \mid \underline{y}).$$

Above we use the standard fact that $D(P\|Q) = I(X_i; X_j \mid \underline{y})$, in our case. Averaging over the outputs we get

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i^{-2} \partial \epsilon_j^{-2}} = \mathbb{E}_{\underline{t}}[(T_{ij} - T_i T_j)^2] \leq 8I(X_i; X_j \mid \underline{Y}),$$

where the equality above follows from (2.10).

Highly noisy BMS channels. From the high noise expansion (see Section 2.4 and the above remarks), we can derive an inequality like the preceding one, which holds in the high noise regime for general BMS channels. The number 8 gets replaced by some suitable factor which depends on the channel noise.

2.1.4 Organization of the Chapter

The statistical mechanics formulation is convenient to perform many of the necessary calculations, but also the interpolation method is best formulated in this framework. Thus we briefly recall it in Section 2.2 as well as a few connections to the information theoretic language. Section 2.3 gives a derivation of Theorem 1.2 (the operational formula for the MAP-GEXIT function as shown in Chapter 1), along with its simplifications for the BEC(ϵ) and BIAWGNC(ϵ). Section 2.4 contains the derivation of the correlation formula (Lemma 2.1). The interpolation method that is used to prove the variational bound (Theorem 2.1) is presented in Section 2.5. The main new ingredient of the proof is an estimate (see Lemma 2.2 in Section 2.5) on the fluctuations of overlap parameters. The proof of Lemma 2.2 is the object of Section 2.6. The appendices contain technical calculations involved in the proofs and describe some tools which are used throughout this chapter and the rest of the thesis.

2.2 Statistical Mechanics Formulation

Consider a fixed code belonging to the ensemble $\mathcal{C} = LDPC(n, \Lambda(x), P(x))$. Recall that the posterior distribution $p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y})$ used in MAP decoding can be viewed as the Gibbs measure of a particular random spin system. We call code-bit $x \in \{+1, -1\}$ a spin, alluding to the statistical mechanics language. Given a set $A \subseteq \{1, ..., n\}$, we use the notation $x_A = \prod_{i \in A} x_i$. It will be clear from the context if the subscript is a set or a single bit. For a uniform prior over the code words and a BMS(ϵ) channel, Bayes rule implies $p_{\underline{X}|\underline{Y}}(\underline{x} \mid \underline{y}) = \mu(\underline{x})$ with

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{a=1}^{m} \frac{1}{2}(1 + x_{\partial a}) \prod_{i=1}^{n} e^{\frac{l_i}{2} x_i}, \tag{2.11}$$

where $\prod_a$ is a product over all check nodes of the given code, and $x_{\partial a} = \prod_{i\in\partial a} x_i$ is the product of the spins (mod 2 sum of the bits) corresponding to the variable nodes i that are connected to a check a. Z is the normalization factor or "partition function" and $\ln Z$ is the "pressure" associated to the Gibbs measure $\mu(\underline{x})$ (see 1.7 for details).

Recall that from (1.24) the conditional entropy is given by

$$H(\underline{X} \mid \underline{Y}) = \mathbb{E}_{\underline{l}}[\ln Z] - \sum_{i=1}^{n} \int_{-\infty}^{+\infty} dl_i c_L(l_i) \frac{l_i}{2}. \tag{2.12}$$

Expectations with respect to $\mu(\underline{x})$ for a fixed graph and a fixed channel output are denoted by the Gibbs bracket $\langle - \rangle$. More precisely, for any $A \subseteq \{1, ..., n\}$,

$$\langle x_A \rangle = \sum_{x^n} x_A \mu(x^n), \qquad x_A = \prod_{i\in A} x_i.$$

Details on the above formalism can be found for example in [84].

Using the new notation, the soft estimate of the bit x_i (c.f. (2.4)) is

$$T_i = \langle x_i \rangle. \tag{2.13}$$

We will also need soft estimates for $x_i \oplus x_j$, $i \neq j$. In the statistical mechanics formalism they are simply expressed as (c.f. (2.5))

$$T_{ij} = \langle x_i x_j \rangle. \tag{2.14}$$

In particular, the correlation between bits x_i and x_j becomes $(T_{ij} - T_i T_j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$, which is the usual notion of spin-spin correlation in statistical mechanics.

In Section 2.4 (and Appendices 2.B and 2.C) the algebraic manipulations are best performed in terms of "extrinsic" soft bit estimates. We will need many variants, the simplest one being the estimate of x_i when observation y_i is not available

$$T_i^{\sim i} = \tanh\left(\frac{\Phi_i^{\sim i}}{2}\right) = p_{X_i|\underline{Y}^{\sim i}}(+1 \mid \underline{y}^{\sim i}) - p_{X_i|\underline{Y}^{\sim i}}(-1 \mid \underline{y}^{\sim i}).$$

The second is the estimate of x_i when both y_i and y_j are not available

$$T_i^{\sim ij} = \tanh\left(\frac{\Phi_i^{\sim ij}}{2}\right) = p_{X_i|\underline{Y}^{\sim ij}}(+1 \mid \underline{y}^{\sim ij}) - p_{X_i|\underline{Y}^{\sim ij}}(-1 \mid \underline{y}^{\sim ij}).$$

Finally, we will also need the extrinsic estimate of the mod 2 sum $X_i \oplus X_j$ when both y_i and y_j are not available,

$$T_{ij}^{\sim ij} = \tanh\left(\frac{\Phi_{ij}^{\sim ij}}{2}\right) = p_{X_i\oplus X_j|\underline{Y}^{\sim ij}}(+1 \mid \underline{y}^{\sim ij}) - p_{X_i\oplus X_j|\underline{Y}^{\sim ij}}(-1 \mid \underline{y}^{\sim ij}).$$

It is practical to work in terms of a modified Gibbs average $\langle x_A \rangle_{\sim i}$ which means that $l_i = 0$, in other words y_i is not available. Similarly, we introduce the averages $\langle x_A \rangle_{\sim ij}$, in other words the Gibbs average when both y_i and y_j are unavailable. One has

$$T_i^{\sim i} = \langle x_i \rangle_{\sim i}, \qquad T_i^{\sim ij} = \langle x_i \rangle_{\sim ij}, \qquad T_{ij}^{\sim ij} = \langle x_i x_j \rangle_{\sim ij}.$$

The extrinsic brackets $\langle - \rangle_{\sim i}$ and $\langle - \rangle_{\sim ij}$ are related to the usual ones $\langle - \rangle$ by the following formulas derived in Appendix 2.B,

$$\langle x_i \rangle_{\sim i} = \frac{\langle x_i \rangle - t_i}{1 - \langle x_i \rangle t_i}, \tag{2.15}$$

$$\langle x_i \rangle_{\sim ij} = \frac{\langle x_i \rangle - t_i - \langle x_i x_j \rangle t_j + t_i t_j \langle x_j \rangle}{1 - \langle x_i \rangle t_i - \langle x_j \rangle t_j + \langle x_i x_j \rangle t_i t_j}, \tag{2.16}$$

$$\langle x_i x_j \rangle_{\sim ij} = \frac{\langle x_i x_j \rangle - t_i \langle x_j \rangle - \langle x_i \rangle t_j + t_i t_j}{1 - \langle x_i \rangle t_i - \langle x_j \rangle t_j + \langle x_i x_j \rangle t_i t_j}, \tag{2.17}$$

where recall that $t_i = \tanh \frac{l_i}{2}$ and $t_j = \tanh \frac{l_j}{2}$ are the difference domain random variables.

2.3 First Derivative Formula

We give a quick proof of Theorem 1.2 here. A derivation can be found in [23], [34], [91]. We assume that transmission takes place over a smooth BMS(ϵ) channel. Basic information theory gives us,

$$\begin{aligned} H(\underline{X} \mid \underline{Y}) &= H(X_i \mid \underline{Y}) + H(\underline{X}^{\sim i} \mid X_i, \underline{Y}) \\ &= H(X_i \mid \underline{Y}) + H(\underline{X}^{\sim i} \mid X_i, \underline{Y}^{\sim i}). \end{aligned}$$

Since $H(\underline{X}^{\sim i} \mid X_i, \underline{Y}^{\sim i})$ is independent of ϵ_i we have

$$\frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i} = \frac{\partial H(X_i \mid \underline{Y})}{\partial \epsilon_i}.$$

Note that the channel family is smooth and that only the distribution $c_L(l)$ depends on ϵ. Using (1.24) and the smoothness of the channel (c.f. Definition 1.4 in Chapter 1), we have

$$\begin{aligned} \frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i} &= \frac{\partial}{\partial \epsilon_i} \mathbb{E}_{\underline{l}}[\ln Z] - \int_{-\infty}^{+\infty} dl_i \frac{\partial c_L(l_i)}{\partial \epsilon_i} \frac{l_i}{2} \\ &= \mathbb{E}_{\underline{l}^{\sim i}} \int_{-\infty}^{+\infty} dl_i \frac{\partial c_L(l_i)}{\partial \epsilon_i} \Big(\ln Z - \frac{l_i}{2} \Big), \end{aligned} \tag{2.18}$$

where we recall that $\mathbb{E}_{\underline{l}^{\sim i}}$ denotes the expectation w.r.t. all LLR values $\underline{l}$ except l_i. Let

$$Z_{\sim i} = \sum_{\underline{x}} \prod_{a \in C} \frac{1}{2}(1 + x_{\partial a}) \prod_{k \neq i} e^{\frac{l_k}{2} x_k}, \tag{2.19}$$

be the partition function for the Gibbs bracket $\langle - \rangle_{\sim i}$ introduced in Section 2.2 and consider

$$\ln\left(\frac{Z}{Z_{\sim i}}\right) = \ln\left(\langle e^{\frac{l_i}{2}x_i}\rangle_{\sim i}\right). \tag{2.20}$$

Consider the identity,

$$e^{\frac{l_i}{2}x_i} = e^{\frac{l_i}{2}}\frac{1+t_i x_i}{1+t_i}, \tag{2.21}$$

which holds for all real l_i, $t_i = \tanh\frac{l_i}{2}$ and for $x_i \in \{+1, -1\}$. Using the above identity and (2.20) we get

$$\ln Z - \frac{l_i}{2} = \ln Z_{\sim i} + \ln\left(\frac{1+t_i\langle x_i\rangle_{\sim i}}{1+t_i}\right). \tag{2.22}$$

When we replace this expression in the integral (2.18) we see that the contribution of $\ln Z_{\sim i}$ vanishes because this latter quantity is independent of l_i. Indeed,

$$\int_{-\infty}^{+\infty} dl_i \frac{\partial c_L(l_i)}{\partial \epsilon_i} \ln Z_{\sim i} = \ln Z_{\sim i}\left(\frac{\partial}{\partial \epsilon_i}\int_{-\infty}^{+\infty} dl_1 c_L(l_i)\right) = 0, \tag{2.23}$$

since $c_L(l_i)$ is a probability distribution and hence, normalized. Then, using (2.15) leads to

$$\frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i} = \int_{-\infty}^{+\infty} dl_i \frac{\partial c_L(l_i)}{\partial \epsilon_i}\mathbb{E}_{\underline{l}^{\sim i}}\left[\ln\left(\frac{1+t_i\langle x_i\rangle_{\sim i}}{1+t_i}\right)\right] \tag{2.24}$$

$$= -\int_{-1}^{+1} dt_i \frac{\partial c_D(t_i)}{\partial \epsilon_i}\mathbb{E}_{\underline{t}^{\sim i}}\left[\ln\left(\frac{1-t_i\langle x_i\rangle}{1-t_i}\right)\right]. \tag{2.25}$$

Using (2.13) we obtain,

$$\frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i} = \int_{-1}^{+1} dt_i \frac{\partial c_D(t_i)}{\partial \epsilon_i} g_1(t_i).$$

Here

$$g_1(t_i) = -\mathbb{E}_{\underline{t}^{\sim i}}\left[\ln\left(\frac{1-t_i T_i}{1-t_i}\right)\right].$$

We will use this expression to derive the formula for the first derivative of the average (w.r.t. the ensemble) per-bit conditional entropy. Using

$$\frac{dH(\underline{X} \mid \underline{Y})}{d\epsilon} = \sum_{i=1}^{n} \frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i}\bigg|_{\epsilon_i=\epsilon},$$

and averaging over the code ensemble $\mathcal{C}$, we get for the average per-bit conditional entropy h_n,

$$\frac{d\mathbb{E}_{\mathcal{C}}[h_n]}{d\epsilon} = \int_{-1}^{+1} dt_1 \frac{\partial c_D(t_1)}{\partial \epsilon}\mathbb{E}_{\mathcal{C}}[g_1(t_1)].$$

For the BEC and the BIAWGNC these general formulas take a simpler form.

Example 2.2 (MAP-GEXIT Equals Erasure Probability for BEC(ϵ))**.** *For the BEC we use the extrinsic formula*[2] *in* (2.24) *and* $c_L(l_i) = \epsilon_i\delta_0(l_i)+(1-\epsilon_i)\delta_\infty(l_i)$ *to get*

$$\begin{aligned}\frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i} &= \int_{-\infty}^{+\infty} dl_i\big(\delta_0(l_i) - \delta_\infty(l_i)\big)\mathbb{E}_{\underline{l}^{\sim i}}\left[\ln\left(\frac{1+t_i\langle x_i\rangle_{\sim i}}{1+t_i}\right)\right] \\ &= -\mathbb{E}_{\underline{l}^{\sim i}}\left[\ln\left(\frac{1+\langle x_i\rangle_{\sim i}}{2}\right)\right],\end{aligned}$$

where we used that $t_i = 0$ *when* $l_i = 0$*, rendering the logarithm to be zero. As shown in Appendix 2.A.2,* $\langle x_i\rangle_{\sim i} \in \{0,1\}$*. Intuitively, we can argue the previous statement as follows. When we transmit over the BEC, the soft estimate of the bit (under MAP decoding) is either an erasure or it is* $+1$ *(under the assumption of the all-one codeword transmission). This implies we must have* $\langle x_i\rangle_{\sim i} \in \{0,1\}$*. As a result, we can write* $\ln(1+\langle x_i\rangle_{\sim i}) = \langle x_i\rangle_{\sim i}\big(\ln 2\big)$*. Substituting we get,*

$$\frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i} = \ln 2(1 - \mathbb{E}_{\underline{l}^{\sim i}}[\langle x_i\rangle_{\sim i}]). \tag{2.26}$$

Notice that $(1-\mathbb{E}_{\underline{l}^{\sim i}}[\langle x_i\rangle_{\sim i}])$ *is equal to the erasure probability of the bit MAP decoder given the observations* $\underline{l}^{\sim i}$*. Hence we can write* $\frac{\partial}{\partial\epsilon_i}H(\underline{X} \mid \underline{Y}) = \mathbb{P}(\hat{x}_{i,MAP}(\underline{y}^{\sim i}) = *)\ln 2$*, which is amongst the various characterization of the MAP-GEXIT as shown in [23]. If the output of the bit* i *is not an erasure then* $l_i = \infty$*. In this case* $\langle x_i\rangle = 1$*. If* $l_i = 0$*, then* $\langle x_i\rangle = \langle x_i\rangle_{\sim i}$*. Thus we get* $\mathbb{E}_{\underline{l}}[\langle x_i\rangle] = \epsilon\mathbb{E}_{\underline{l}}[\langle x_i\rangle_{\sim i}] + (1-\epsilon)$*. Combining all we have*

$$\frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \ln \epsilon_i} = \ln 2(1 - \mathbb{E}_{\underline{t}}[T_i]), \tag{2.27}$$

and

$$\frac{d\mathbb{E}_{\mathcal{C}}[h_n]}{d\ln\epsilon} = \ln 2(1 - \mathbb{E}_{\mathcal{C},\underline{t}}[T_1]). \tag{2.28}$$

Example 2.3 (MAP-GEXIT for BIAWGNC(ϵ))**.** *For the BIAWGNC, we use* (1.24) *and use the gaussian integration by parts formula with* $f(\underline{l},\underline{x}) = \ln Z$ *and* $f(\underline{l},\underline{x}) = \frac{l_i}{2}$ *(c.f.* (2.77) *in Appendix 2.A.3) to get*

$$\frac{\partial H(X_i \mid \underline{Y})}{\partial \epsilon_i} = -4\epsilon_i^{-3}\mathbb{E}_{\underline{l}}\Big[\Big(\frac{\partial}{\partial l_i} + \frac{\partial^2}{\partial l_i^2}\Big)\ln Z\Big] - 2\epsilon_i^{-3}.$$

From the definition of Gibbs averages and the partition function (2.11) *we get*

$$\frac{\partial \ln Z}{\partial l_i} = \frac{1}{2}\langle x_i\rangle, \quad \frac{\partial^2 \ln Z}{\partial l_i^2} = \frac{1}{4}(\langle x_i^2\rangle - \langle x_i\rangle^2) = \frac{1}{4}(1 - \langle x_i\rangle^2).$$

[2] In this case the ratio in the logarithm may take the ambiguous value $\frac{0}{0}$ but the formula is to be interpreted according to (2.27). We will see in Section 2.2 that in terms of extrinsic soft bit estimates there is an analogous expression that is unambiguous.

Using the Nishimori identities in Appendix 2.A.2 we get $\mathbb{E}_{\underline{l}}[\langle x_i \rangle] = \mathbb{E}_{\underline{l}}[\langle x_i \rangle^2]$ *and thus we obtain,*

$$\frac{\partial H(X_i \mid \underline{Y})}{\partial \epsilon_i} = \epsilon_i^{-3} \mathbb{E}_{\underline{l}}[1 - \langle x_i \rangle]. \tag{2.29}$$

Using (2.13) *and writing in terms of inverse square noise we get*

$$\frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i^{-2}} = -\frac{1}{2}(1 - \mathbb{E}_{\underline{l}}[T_i]), \tag{2.30}$$

and

$$\frac{d\mathbb{E}_{\mathcal{C}}[h_n]}{d\epsilon^{-2}} = -\frac{1}{2}(1 - \mathbb{E}_{\mathcal{C},\underline{t}}[T_1]). \tag{2.31}$$

We would like to point out here that the average MAP-GEXIT formulas for the BIAWGNC presented here have been derived in [23], [34]. Also the derivative w.r.t. the inverse noise variance is just for convenience of exposition.

2.4 Second Derivative – The Correlation Formula

In this section we prove the correlation formula in Lemma 2.1 for the second derivative of the conditional entropy. We will then specialize the formula to the case of transmission over the BEC and BIAWGNC. As a result we will observe that, for these two channels the relationship of the second derivative of the conditional entropy w.r.t. ϵ is very straightforward.

2.4.1 Proof of the Correlation Formula in Lemma 2.1

For any BMS($\underline{\epsilon}$) channel and any linear code we have from (2.18)

$$\frac{\partial H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i} = \mathbb{E}_{\underline{l}^{\sim i}}\left[\int_{-\infty}^{+\infty} dl_i \frac{\partial c_L(l_i)}{\partial \epsilon_i}(\ln Z - \frac{l_i}{2})\right].$$

Differentiating once more, we get

$$\frac{\partial^2}{\partial \epsilon_i \partial \epsilon_j} H(\underline{X} \mid \underline{Y}) = \delta_{ij} S_1 + (1 - \delta_{ij}) S_2, \tag{2.32}$$

where

$$S_1 = \mathbb{E}_{\underline{l}^{\sim i}}\left[\int_{-\infty}^{+\infty} dl_i \frac{\partial^2 c_L(l_i)}{\partial \epsilon_i^2}\Big(\ln Z - \frac{l_i}{2}\Big)\right], \tag{2.33}$$

and

$$S_2 = \mathbb{E}_{\underline{l}^{\sim ij}}\left[\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i}\frac{\partial c_L(l_j)}{\partial \epsilon_j}\Big(\ln Z - \frac{l_i}{2}\Big)\right].$$

Recall that $\mathbb{E}_{\underline{l}^{\sim ij}}$ denotes the expectation w.r.t. all the LLR values $\underline{l}$ except l_i and l_j. Consider first S_1. Recall from (2.22) that

$$\ln Z - \frac{l_i}{2} = \ln Z_{\sim i} + \ln\left(\frac{1+t_i\langle x_i\rangle_{\sim i}}{1+t_i}\right).$$

When we replace this expression in the integral (2.33) we again see that the contribution of $\ln Z_{\sim i}$ vanishes because this later quantity is independent of l_i. Again, using (2.15) leads to

$$S_1 = \int_{-1}^{+1} dt_i \frac{\partial^2 c_D(t_i)}{\partial \epsilon_i^2} \mathbb{E}_{\sim i}\left[\ln\left(\frac{1+t_i\langle x_i\rangle_{\sim i}}{1+t_i}\right)\right] \tag{2.34}$$

$$= -\int_{-1}^{+1} dt_i \frac{\partial^2 c_D(t_i)}{\partial \epsilon_i^2} \mathbb{E}_{\underline{t}^{\sim i}}\left[\ln\left(\frac{1-t_i\langle x_i\rangle}{1-t_i}\right)\right], \tag{2.35}$$

which (because of (2.13)) coincides with the first term in the correlation formula provided by Lemma 2.1.

Consider next the term S_2. Notice that

$$\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i}\frac{\partial c_L(l_j)}{\partial \epsilon_j}\Big(\frac{l_j}{2}\Big) = \int_{-\infty}^{+\infty} dl_j \frac{\partial c_L(l_j)}{\partial \epsilon_j}\frac{l_j}{2}\Big(\frac{\partial}{\partial \epsilon_i}\int_{-\infty}^{+\infty} dl_i c_L(l_i)\Big) = 0.$$

Similarly,

$$\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i}\frac{\partial c_L(l_j)}{\partial \epsilon_j}\Big(\frac{l_i}{2}\Big) = \int_{-\infty}^{+\infty} dl_i \frac{\partial c_L(l_i)}{\partial \epsilon_i}\frac{l_i}{2}\Big(\frac{\partial}{\partial \epsilon_j}\int_{-\infty}^{+\infty} dl_j c_L(l_j)\Big) = 0.$$

Thus we can re-write S_2 as

$$S_2 = \mathbb{E}_{\underline{l}^{\sim ij}}\Big[\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i}\frac{\partial c_L(l_j)}{\partial \epsilon_j}\Big(\ln Z - \frac{l_i}{2} - \frac{l_j}{2}\Big)\Big].$$

The above manipulations are done to symmetrize the expression for S_2. This helps us in deriving the correlation formula for the second derivative.

Let $Z_{\sim ij} = \sum_{\underline{x}} \prod_{c\in\mathcal{C}} \frac{1}{2}(1+x_{\partial c}) \prod_{k\neq i,j} e^{\frac{l_k}{2}x_k}$ be the partition function for the Gibbs bracket $\langle\cdot\rangle_{\sim ij}$, and consider

$$\ln\Big(\frac{Z}{Z_{\sim ij}}\Big) = \ln\big(\langle e^{\frac{l_i}{2}x_i + \frac{l_j}{2}x_j}\rangle_{\sim ij}\big).$$

Using again (2.21) we get

$$\ln Z - \frac{l_i}{2} - \frac{l_j}{2} = \ln Z_{\sim ij} + \ln\left(\frac{1+t_i\langle x_i\rangle_{\sim ij} + t_j\langle x_j\rangle_{\sim ij} + t_it_j\langle x_ix_j\rangle_{\sim ij}}{1+t_i+t_j+t_it_j}\right).$$

As before the contribution of $\ln Z_{\sim ij}$ vanishes because it is independent of l_i and l_j. Since $\ln(1+t_i\langle x_i\rangle_{\sim ij})$ (resp. $\ln(1+t_j\langle x_j\rangle_{\sim ij})$) is independent of l_j (resp.

l_i), we have

$$\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i} \frac{\partial c_L(l_j)}{\partial \epsilon_j} \ln(1 + t_i \langle x_i \rangle_{\sim ij}) = \int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i} \frac{\partial c_L(l_j)}{\partial \epsilon_j} \ln(1 + t_j \langle x_j \rangle_{\sim ij}) = 0. \tag{2.36}$$

Similarly, notice that

$$\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i} \frac{\partial c_L(l_j)}{\partial \epsilon_j} \ln(1 + t_i) = \int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i} \frac{\partial c_L(l_j)}{\partial \epsilon_j} \ln(1 + t_j) = 0. \tag{2.37}$$

Using the above identities (2.36) and (2.37) this leads to the following formula for S_2,

$$S_2 = \mathbb{E}_{\underset{\sim}{l}^{\sim ij}} \left[\int_{-1}^{+1} dt_i dt_j \frac{\partial c_D(t_i)}{\partial \epsilon_i} \frac{\partial c_D(t_j)}{\partial \epsilon_j} \times \ln \left(\frac{1 + t_i \langle x_i \rangle_{\sim ij} + t_j \langle x_j \rangle_{\sim ij} + t_i t_j \langle x_i x_j \rangle_{\sim ij}}{1 + t_i \langle x_i \rangle_{\sim ij} + t_j \langle x_j \rangle_{\sim ij} + t_i t_j \langle x_i \rangle_{\sim ij} \langle x_j \rangle_{\sim ij}} \right) \right]. \tag{2.38}$$

To get the formulas in terms of usual averages we use the relations (2.16), (2.17). Hence

$$S_2 = \mathbb{E}_{\underset{\sim}{l}^{\sim ij}} \left[\int_{-1}^{+1} dt_i dt_j \frac{\partial c_D(t_i)}{\partial \epsilon_i} \frac{\partial c_D(t_j)}{\partial \epsilon_j} \ln \left(\frac{1 - t_i \langle x_i \rangle - t_j \langle x_j \rangle + t_i t_j \langle x_i x_j \rangle}{1 - t_i \langle x_i \rangle - t_j \langle x_j \rangle + t_i t_j \langle x_i \rangle \langle x_j \rangle} \right) \right]. \tag{2.39}$$

Because of (2.13) and (2.14) this coincides with the second term in the correlation formula. The lemma now follows from (2.32), (2.35) and (2.39).

2.4.2 Expressions in terms of the Spin-Spin Correlation

An important step in our analysis is the second derivative of the conditional entropy formula. Although we made progress in explicitly deriving a formula for the second derivative in the previous section, it still seems unusable. The relationship between the second derivative and the correlation is not clear from (2.39). The reason for this is the presence of the logarithm in the expression. We show later that a miraculous manipulation of the second derivative formula (2.6) reveals a nice relationship between the second derivative and spin-spin correlation which will facilitate the analysis. As a first step, we consider the BEC and BIAWGNC for which the computations are easier and we also show how we use the Nishimori identities (see Appendix 2.A.2) for algebraic manipulations.

Example 2.4 (Correlation Formula for the BEC)**.** *From* $c_D(t) = (1-\epsilon)\delta(t-1)+\epsilon\delta(t)$*, the second derivative in terms of extrinsic quantities (formulas* (2.34) *and* (2.38)*) reduces to*

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial\epsilon_i \partial\epsilon_j} = (1-\delta_{ij})\mathbb{E}_{\underline{t}^{\sim ij}}\left[\ln\left(\frac{1+\langle x_i\rangle_{\sim ij}+\langle x_j\rangle_{\sim ij}+\langle x_i x_j\rangle_{\sim ij}}{1+\langle x_i\rangle_{\sim ij}+\langle x_j\rangle_{\sim ij}+\langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}}\right)\right].$$

We remark that the term S_1 *vanishes because* $\frac{\partial^2 c_D(t_i)}{\partial\epsilon_i^2}$ *equals zero when transmitting over the BEC. From Appendix 2.A.1, when transmitting over the BEC, the Gibbs average* $\langle x_A\rangle_{\sim ij}$ *equals either* 0 *or* 1*. As a result, it is not hard to see that*

$$\begin{aligned}\ln\big(1+\langle x_i\rangle_{\sim ij}+\langle x_j\rangle_{\sim ij}+\langle x_i x_j\rangle_{\sim ij}\big) &= (\ln 2)\big(\langle x_i\rangle_{\sim ij}+\langle x_j\rangle_{\sim ij}+\langle x_i x_j\rangle_{\sim ij}\big)\\ &+ (\ln 3 - 2\ln 2)\big(\langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}+\langle x_i\rangle_{\sim ij}\langle x_i x_j\rangle_{\sim ij}+\langle x_j\rangle_{\sim ij}\langle x_i x_j\rangle_{\sim ij}\big)\\ &+ (5\ln 2 - 3\ln 3)\langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}\langle x_i x_j\rangle_{\sim ij},\end{aligned}$$

and

$$\ln\big(1+\langle x_i\rangle_{\sim ij}+\langle x_j\rangle_{\sim ij}+\langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}\big) = (\ln 2)\big(\langle x_i\rangle_{\sim ij}+\langle x_j\rangle_{\sim ij}\big).$$

The difference of the two logarithms is simplified using the following Nishimori identities (see Appendix 2.A.2),

$$\begin{aligned}\mathbb{E}_{\underline{l}}[\langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}] = \mathbb{E}_{\underline{l}}[\langle x_i\rangle_{\sim ij}\langle x_i x_j\rangle_{\sim ij}] &= \mathbb{E}_{\underline{l}}[\langle x_j\rangle_{\sim ij}\langle x_i x_j\rangle_{\sim ij}]\\ &= \mathbb{E}_{\underline{l}}[\langle x_i x_j\rangle_{\sim ij}\langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}]. \quad (2.40)\end{aligned}$$

Finally we obtain the simple expression

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial\epsilon_i \partial\epsilon_j} = \ln 2(1-\delta_{ij})\mathbb{E}_{\underline{l}}\big[\langle x_i x_j\rangle_{\sim ij} - \langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}\big].$$

It is not hard to see that

$$\mathbb{E}_{\underline{l}}\big[\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle\big] = \epsilon_i\epsilon_j\mathbb{E}_{\underline{l}}\big[\langle x_i x_j\rangle_{\sim ij} - \langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}\big].$$

Indeed, whenever $l_i = +\infty$ *we have* $x_i = +1$*. As a consequence we have* $\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle = \langle x_j\rangle - \langle x_j\rangle = 0$*. We argue in a similar way when* $l_j = +\infty$*. Putting everything together we finally get,*

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial\epsilon_i \partial\epsilon_j} = \frac{\ln 2}{\epsilon_i\epsilon_j}(1-\delta_{ij})\mathbb{E}_{\underline{l}}\big[T_{ij} - T_i T_j\big].$$

Let us point out that the second GKS inequality (see Appendix 2.A.1) implies that $\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle \geq 0$*.*

Example 2.5 (Correlation Formula for the BIAWGNC)**.** *Using* (2.29) *and the gaussian integration by parts formula with* $f(\underline{l},\underline{x}) = \epsilon_i^{-3}\mathbb{E}_{\underline{l}}[1-\langle x_i\rangle]$*, we get*

$$\frac{\partial^2 H(\underline{X}\mid\underline{Y})}{\partial\epsilon_i\partial\epsilon_j} = -4\epsilon_j^{-3}\Big[\Big(\frac{\partial}{\partial l_j}+\frac{\partial^2}{\partial l_j^2}\Big)\epsilon_i^{-3}\mathbb{E}_{\underline{l}}[1-\langle x_i\rangle]\Big].$$

A simple calculation shows that

$$\frac{\partial\langle x_i\rangle}{\partial l_j} = \frac{1}{2}(\langle x_ix_j\rangle - \langle x_i\rangle\langle x_j\rangle), \qquad \frac{\partial\big(\langle x_ix_j\rangle\big)}{\partial l_j} = \frac{1}{2}(\langle x_i\rangle - \langle x_ix_j\rangle\langle x_j\rangle),$$

and

$$\frac{\partial\big(\langle x_i\rangle\langle x_j\rangle\big)}{\partial l_j} = \frac{1}{2}\big(\langle x_ix_j\rangle\langle x_j\rangle - \langle x_i\rangle\langle x_j\rangle^2 + \langle x_i\rangle - \langle x_i\rangle\langle x_j\rangle^2\big).$$

Let us show the calculations for the last derivative in more details. We have

$$\frac{\partial\big(\langle x_i\rangle\langle x_j\rangle\big)}{\partial l_j} = \Big(\frac{\partial\langle x_i\rangle}{\partial l_j}\Big)\langle x_j\rangle + \langle x_i\rangle\Big(\frac{\partial\langle x_j\rangle}{\partial l_j}\Big).$$

For the first term we get

$$\begin{aligned}
\frac{\partial\langle x_i\rangle}{\partial l_j} &= \frac{1}{Z}\sum_{\underline{x}} x_i\prod_a\frac{1}{2}(1+x_{\partial a})\prod_{i\neq j}e^{\frac{l_i}{2}x_i}\frac{\partial}{\partial l_j}e^{\frac{l_j}{2}x_j}\\
&\quad-\frac{1}{Z^2}\Big(\sum_{\underline{x}} x_i\prod_a\frac{1}{2}(1+x_{\partial a})\prod_{i=1}^n e^{\frac{l_i}{2}x_i}\Big)\Big(\sum_{\underline{x}}\prod_a\frac{1}{2}(1+x_{\partial a})\prod_{i\neq j}e^{\frac{l_i}{2}x_i}\frac{\partial}{\partial l_j}e^{\frac{l_j}{2}x_j}\Big)\\
&= \frac{1}{2Z}\sum_{\underline{x}} x_ix_j\prod_a\frac{1}{2}(1+x_{\partial a})\prod_{i=1}^n e^{\frac{l_i}{2}x_i}\\
&\quad-\frac{1}{2Z^2}\Big(\sum_{\underline{x}} x_i\prod_a\frac{1}{2}(1+x_{\partial a})\prod_{i=1}^n e^{\frac{l_i}{2}x_i}\Big)\Big(\sum_{\underline{x}} x_j\prod_a\frac{1}{2}(1+x_{\partial a})\prod_{i=1}^n e^{\frac{l_i}{2}x_i}\Big)\\
&= \frac{1}{2}\langle x_ix_j\rangle - \frac{1}{2}\langle x_i\rangle\langle x_j\rangle.
\end{aligned}$$

Multiplying with $\langle x_j\rangle$ *and performing similar calculations for* $\frac{\partial\langle x_j\rangle}{\partial l_j}$ *we obtain the above identity.*

Now using (2.40) *and the Nishimori identities (see Appendix 2.A.2)*

$$\begin{aligned}
&\mathbb{E}_{\underline{l}}\big[\langle x_ix_j\rangle\big] = \mathbb{E}_{\underline{l}}\big[\langle x_ix_j\rangle^2\big], \quad \mathbb{E}_{\underline{l}}\big[\langle x_i\rangle\langle x_j\rangle^2\big] = \mathbb{E}_{\underline{l}}\big[\langle x_i\rangle^2\langle x_j\rangle^2\big],\\
&\mathbb{E}_{\underline{l}}\big[\langle x_i\rangle\langle x_j\rangle\big] = \mathbb{E}_{\underline{l}}\big[\langle x_i\rangle\langle x_j\rangle\langle x_ix_j\rangle\big], \quad \mathbb{E}_{\underline{l}}\big[\langle x_ix_j\rangle\langle x_j\rangle\big] = \mathbb{E}_{\underline{l}}\big[\langle x_i\rangle\langle x_j\rangle\langle x_ix_j\rangle\big],
\end{aligned}$$

we get

$$\frac{\partial^2 H(\underline{X}\mid\underline{Y})}{\partial\epsilon_i\partial\epsilon_j} = 2\epsilon_j^{-3}\epsilon_i^{-3}\mathbb{E}_{\underline{l}}\Big[\big(\langle x_ix_j - \langle x_i\rangle\langle x_j\rangle\big)^2\Big].$$

Thus in terms of inverse square noise we have

$$\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i^{-2} \partial \epsilon_j^{-2}} = \frac{1}{2}\mathbb{E}_{\underline{l}}[(\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle)^2] = \frac{1}{2}\mathbb{E}_{\underline{l}}\big[\big(T_{ij} - T_i T_j\big)^2\big].$$

Example 2.6 (Correlation Formula for Highly Noisy BMS Channels)**.** *We use the extrinsic form of the correlation formula given by* (2.34) *and* (2.38)*. First we expand the logarithms in S_1 and S_2 in powers of t_i and t_j and then use various Nishimori identities. After a rather tedious algebra (see Appendices 2.C and 2.D) we can organize the expansion in powers of the channel parameters* (2.1)*. In the high noise regime this expansion is absolutely convergent. To lowest order we have*

$$\begin{aligned}
&\frac{\partial^2 H(\underline{X} \mid \underline{Y})}{\partial \epsilon_i \partial \epsilon_j} = \delta_{ij} S_1 + (1 - \delta_{ij}) S_2 \\
&\approx \frac{1}{2}\delta_{ij} {m_2}^{(2)} (\mathbb{E}_{\underline{l}}[\langle x_i \rangle^2] - 1) + \frac{1}{2}(1 - \delta_{ij})[{m_1}^{(2)}]^2 \mathbb{E}_{\underline{l}}\Big[\big(\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle\big)^2\Big] + \dots \\
&= \frac{1}{2}\delta_{ij} {m_2}^{(2)} (\mathbb{E}_{\underline{l}}[T_i^2] - 1) + \frac{1}{2}(1 - \delta_{ij})[{m_1}^{(2)}]^2 \mathbb{E}_{\underline{l}}\Big[\big(T_{ij} - T_i T_j\big)^2\Big] + \dots . \quad (2.41)
\end{aligned}$$

The second derivative of the conditional entropy is directly related to the average square of the code-bit or spin-spin correlation.

2.5 The Interpolation Method

We use the interpolation method in the form developed by Montanari to prove the lower bound for any standard ensemble of LDPC codes. The interpolation method for standard ensembles has also been developed differently in the context of spin-glasses by Franz et al. [92]. But the latter form is more complex and less amenable to analysis in the present context.

As the name suggests, the main idea of the interpolation method is to define a partition function dependent on a scalar interpolating parameter s, where s varies over some interval of $\mathbb{R}$. The system is set up in such a way that for one extreme value of the parameter (say $s = 1$), one recovers the partition function of the system of interest and for the other extreme value of the parameter (say $s = 0$) we obtain a partition function for an "easy" system (usually a system where the Gibbs measure is a product distribution over the constituents of the system).

Using the fundamental theorem of calculus, one then relates the two extreme systems via the derivative of the partition function w.r.t. the parameter s. By carefully designing a "path", which is a function of s, one controls the derivative of the partition function with respect to s and thus give an estimate of the partition function of the system of interest.

As explained in [29], it is difficult to establish directly the bounds for the standard ensembles. Rather, one introduces a "multi-Poisson" ensemble which approximates the standard ensemble. Once the bounds are derived for the Multi-Poisson ensemble they are extended to the standard ensemble by a limiting procedure. The interpolation construction is fairly complicated so that it is helpful to briefly review the simpler pure Poisson case.

2.5.1 Poisson Ensemble

We introduce the ensemble Poisson-LDPC$(n, 1-\mathtt{R}, P(x))$, denoted by $\mathcal{P}$, where n is the block length, $\mathtt{R}$ is the design rate and $P(x) = \sum_k P_k x^k$ is the check degree distribution. A bipartite graph from the ensemble $\mathcal{P}$ is constructed as follows. The graph has n variable nodes. For any k choose, independently, a Poisson number m_k of check nodes with mean equal to $n(1-\mathtt{R})P_k$. Thus the graph has a total of $m = \sum_k m_k$ check nodes. Clearly, m is a Poisson random variable with mean $n(1-\mathtt{R})$. For each check node a of degree k, choose k variable nodes u.a.r. and connect them to a. One can show that the variable node degree distribution concentrates around a Poisson distribution $\Lambda_{\mathcal{P}}(x) = e^{P'(1)(1-\mathtt{R})(x-1)}$ (see [29]). In other words, the fraction of variable nodes with degree l, Λ_l, is Poisson with mean $P'(1)(1-\mathtt{R})$.

The interpolation technique recursively removes the check node constraints, thus increasing the rate of the code. This increase is compensated by adding extra observations, U, distributed as (1.26) where d_V is a trial distribution to be optimized in the final inequality. One can interpret these extra observations as coming from a repetition code whose rate is tuned in a such a way that the total design rate $\mathtt{R}$ remains fixed. More precisely, let $s \in [0,1]$ be an interpolating parameter. At "time" s we consider a Poisson-LDPC$(n, (1-\mathtt{R})s, P) = \mathcal{P}_s$ code. For this code ensemble, the design rate equals $1-(1-\mathtt{R})s$. Besides the usual channel outputs l_i, each node i receives e_i extra i.i.d. observations U_b^i, $b = 1, ..., e_i$, where e_i is Poisson with mean $P'(1)(1-\mathtt{R})(1-s)$ (so the total effective design rate is fixed to $\mathtt{R}$). Indeed, the average variable node degree is equal to $P'(1)\big((1-\mathtt{R})(1-s) + (1-\mathtt{R})s\big) = P'(1)(1-\mathtt{R})$. The interpolating Gibbs measure is given by

$$\mu_s(\underline{x}) = \frac{1}{Z(s)} \prod_a \frac{1}{2}(1 + x_{\partial a}) \prod_{i=1}^{n} e^{(\frac{l_i}{2} + \sum_{b=1}^{e_i} \frac{U_b^i}{2})x_i}. \tag{2.42}$$

Here $\prod_a$ is a product over checks of a given graph in the ensemble $\mathcal{P}_s$. At $s = 1$ one recovers the original measure for $\mathcal{P}$, while at $s = 0$ (no check nodes) we have a simple product measure (corresponding to a repetition code) which is tailored to yield the replica symmetric entropy $h_{RS}[d_V; \Lambda_{\mathcal{P}}, P]$ (up to an extra constant).

The central result of Montanari [29] is the sum-rule

$$\mathbb{E}_{\mathcal{P}}[h_n] = h_{RS}[d_V; \Lambda_{\mathcal{P}}, P] + \int_0^1 R_n(s)ds. \tag{2.43}$$

As a consequence, Montanari in [29] proves that for the special case of convex $P(x)$, the remainder term $R_n(s)$ is always non-negative. This proves that the replica solution is a lower bound for the special case of convex $P(x)$.

Details of the sum-rule and its derivation can be found in [29], [79]. For completeness, we provide a derivation in Appendix 2.A.4. The integral in the sum-rule arises as a result of taking the derivative of the interpolated partition function with respect to the parameter s.

The first term on the right hand side of (2.43), $h_{RS,\mathcal{P}}[d_V; \Lambda_{\mathcal{P}}, P]$, is the replica symmetric functional of Section 2.1 evaluated for the Poisson ensemble. The remainder term $R_n(s)$ is given by

$$R_n(s) = (1 - \mathtt{R}) \sum_{p=1}^{\infty} \frac{1}{2p(2p-1)} \mathbb{E}_s \Big[\big\langle P(Q_{2p}) - P'(q_{2p})(Q_{2p} - q_{2p}) - P(q_{2p}) \big\rangle_{2p,s} \Big]. \tag{2.44}$$

Let us explain the various objects appearing above. Define $q_{2p} = \mathbb{E}_V[(\tanh V)^{2p}]$. Q_{2p} is called the overlap parameter, and is given by

$$Q_{2p} = \frac{1}{n} \sum_{i=1}^{n} x_i^{(1)} x_i^{(2)} \cdots x_i^{(2p)}. \tag{2.45}$$

Here $x_i^{(\alpha)}, \alpha = 1, 2, \ldots, 2p$ are $2p$ independent copies (replicas) of the spin x_i (each replica corresponds to an identical spin system with the same code and LLR realizations) and $\langle - \rangle_{2p,s}$ is the average (Gibbs bracket) associated to the product measure (replica measure)

$$\prod_{\alpha=1}^{2p} \mu_s\big(\underline{x}^{(\alpha)}\big). \tag{2.46}$$

Also we use the short-hand notation $\mathbb{E}_s$ to denote the expectation w.r.t. all the randomness present at time s.

2.5.2 Multi-Poisson Ensemble

In this section we describe the interpolation method technique applied to the Multi-Poisson ensemble introduced by Montanari in [29]. The Multi-Poisson-LDPC(n, Λ, P, γ) ensemble, denoted by $\mathcal{MP}$, is a more elaborate construction which allows to approximate a target standard LDPC(n, Λ, P) ensemble. Its parameters are the block length n, the target variable and check node degree distributions $\Lambda(x)$ and $P(x)$ and the real number γ which controls the closeness to the standard ensemble. We recall that variable and check node degrees have finite maximum degrees. We first describe $\mathcal{MP}$ and then give the construction of the ensemble used by the interpolation method.

The construction of a bipartite graph from the $\mathcal{MP}$ ensemble proceeds via rounds: the process starts with a high rate code and at each round one adds a

very small number of check nodes till one ends up with a code with almost the desired rate and degree distribution. A graph process $\mathcal{G}_t$ is defined for discrete times $t = 0, ..., t_{\max} - 1$, $t_{\max} = \lfloor \Lambda'(1)/\gamma \rfloor - 1$ as follows. For $t = 0$, $\mathcal{G}_0$ has no check nodes and has n variable nodes. The set of variable nodes is partitioned into the subsets $\mathfrak{V}_l$ of cardinality $n\Lambda_l$ for every l and every node $i \in \mathfrak{V}_l$ is decorated with l free sockets. The number $d_i(t)$ keeps track of the number of free sockets on node i once round t is completed. So for $t = 0$, $\mathcal{G}_0$ has no check nodes and each variable node $i \in \mathfrak{V}_l$ has $d_i(0) = l$ free sockets.

At round t, $\mathcal{G}_t$ is constructed from $\mathcal{G}_{t-1}$ as follows. For all k, choose independently a Poisson number m_k^t of check nodes with mean $n\gamma P_k/P'(1)$. Connect each socket of these new degree k check nodes (added at time t) to variable node i according to the probability $w_i(t) = \frac{d_i(t-1)}{\sum_i d_i(t-1)}$. This is the fraction of free sockets present at node i after round $t-1$ was completed. Note that for the Poisson-LDPC ensemble, $w_i(t) = 1/n$, which is uniform distribution over all the variable nodes. Thus in order to get the desired degree distribution, we give a uniform distribution on the free sockets. It is easy to see that at each round or time t, the average degree of the variable nodes increases by γ. Hence at the end of all the rounds, the average variable node degree is given by $\Lambda'(1) - 2\gamma \leq \gamma t_{\max} \leq \Lambda'(1)$.

Once all new check nodes are connected, update the number of free sockets for each variable node $d_i(t) = d_i(t-1) - \Delta_i(t)$. where $\Delta_i(t)$ is the number of times the variable node i was chosen during the round t. For $n \to \infty$ this construction yields graphs with variable degree distributions $\Lambda_\gamma(x)$ (the check degree distribution remains $P(x)$). The variational distance between $\Lambda_\gamma(x)$ and $\Lambda(x)$ tends to zero as $\gamma \to 0$, see [29] for a proof.

The interpolating ensemble now uses two parameters (t_*, s) where $t_* \in \{0, ..., t_{\max} - 1\}$ and $0 \leq s \leq \gamma$. For rounds $0, ..., t_* - 1$ one proceeds exactly as before to obtain a graph $\mathcal{G}_{t_*-1}$. At the next round t_*, one proceeds as before but with γ replaced by s. The decrease in the number of check nodes (on an average) is compensated by adding e_i extra observations for each node i, where e_i is a Poisson random number with mean $n(\gamma - s)w_i(t_*)$. The round is ended by updating the number of free sockets $d_i(t_*) = d_i(t_* - 1) - \Delta_i(t_*) - e_i(t_*)$. Finally, for rounds $t_* + 1, ..., t_{\max}$ no new check node is added but for each variable node i, e_i external observations are added, where e_i is a Poisson random number with mean $n\gamma w_i(t)$. Moreover the free socket counter is updated as $d_i(t) = d_i(t-1) - e_i(t)$. Recall that the external observations are i.i.d. copies of the random variable U (see (1.26)).

The interpolating Gibbs measure $\mu_{t_*,s}(\underline{x})$ has the same form than (2.42) with the appropriate products over checks and extra observations. Let $h_{n,\gamma}$ the conditional entropy of the $\mathcal{MP}$ ensemble $\mathcal{MP}$ (corresponding to the extreme values $t_* = t_{\max}$ and $s = \gamma$). Again, the central result of [29] is the sum-rule

$$\mathbb{E}_{\mathcal{MP}}[h_{n,\gamma}] = h_{RS}[d_V; \Lambda_\gamma, P] + \sum_{t_*=0}^{t_{\max}-1} \int_0^\gamma R_n(t_*, s)ds + o_n(1). \tag{2.47}$$

We will not give the details of this derivation here, although we provide some insights into the appearance of various parameters, in the Appendix 2.A.4. Interested reader is referred to [29]. Explanations on the notation are in order. The first term $h_{RS,\gamma}[d_V;\Lambda_\gamma,P]$ is the replica symmetric functional of 2.1 evaluated for the Multi-Poisson ensemble. The remainder term $R_n(t_*,s)$ is given by

$$R_n(t_*,s)=\sum_{p=1}^{\infty}\frac{1}{(2p)(2p-1)}\mathbb{E}_s\Big[\big\langle P(Q_{2p})-P'(q_{2p})(Q_{2p}-q_{2p})-P(q_{2p})\big\rangle_{2p,t_*,s}\Big], \tag{2.48}$$

where $q_{2p}=\mathbb{E}_V[(\tanh V)^{2p}]$ as before and Q_{2p} are modified overlap parameters

$$Q_{2p}=\sum_{i=1}^{n}w_i(t_*)X_i(t_*)x_i^{(1)}x_i^{(2)}\cdots x_i^{(2p)}. \tag{2.49}$$

Here as before $x_i^{(\alpha)},\alpha=1,2,\ldots,2p$ are $2p$ independent copies (replicas) of the spin x_i and $\langle-\rangle_{2p,t_*,s}$ is the Gibbs bracket associated to the product measure

$$\prod_{\alpha=1}^{2p}\mu_{t_*,s}(\underline{x}^{(\alpha)}).$$

The overlap parameter is now more complicated than in the Poisson case because of the (positive) terms $w_i(t_*)$ and $X_i(t_*)$. Here $X_i(t_*)$ are new i.i.d. random variables whose precise description is quite technical and can be found in [29]. The reader may think of the terms $w_i(t_*)X_i(t_*)$ as behaving like the $\frac{1}{n}$ factor of the pure Poisson ensemble overlap parameter (2.45). More precisely, the only properties (see Appendix E in [29]) that we need are

$$\sum_{i=1}^{n}w_i(t_*)=1\ \text{ (since } w_i(t_*) \text{ is a probability distribution at any } t_*), \tag{2.50}$$

$$\mathbb{P}\big[w_i(t_*)\le\frac{A}{n}\big]\ge 1-e^{-Bn}, \tag{2.51}$$

$$0\le X_i(t_*)\le\mathfrak{X},\qquad \mathbb{E}[\mathfrak{X}^k]\le A_k, \tag{2.52}$$

for any finite k and finite positive constants A, B, A_k independent of n. Finally we use the shorthand $\mathbb{E}_s[-]$ for the expectation with respect to all random variables involved in the interpolation measure. The subscript s is here to remind us that this expectation depends on s, a fact that is important to keep in mind because the remainder involves an integral over s. When we use $\mathbb{E}$ (without the subscript s; as in (2.52) for example) it means that the quantity does not depend on s. In the sequel the replicated Gibbs bracket $\langle-\rangle_{2p,t_*,s}$ is simply denoted by $\langle-\rangle_s$. There will be no risk of confusion because the only property that we us is its linearity.

In [29] it is shown that

$$\mathbb{E}_{\mathcal{C}}[h_n] = \mathbb{E}_{\mathcal{MP}}[h_{n,\gamma}] + O(\gamma^b) + o_n(1), \tag{2.53}$$

where $O(\gamma^b)$ is uniform in n ($b > 0$ a numerical constant) and $o_n(1)$ (depends on γ) tends to 0 as $n \to +\infty$.

In the next section we prove the variational bound on the conditional entropy of the MP ensemble, namely

$$\liminf_{n\to+\infty} \mathbb{E}_{\mathcal{MP}}[h_{n,\gamma}] \geq h_{RS}[d_V; \Lambda_\gamma, P]. \tag{2.54}$$

Note that here $o_n(1)$ again depends on γ. By combining this bound with (2.53) and taking limits

$$\liminf_{n\to+\infty} \mathbb{E}_{\mathcal{C}}[h_n] = \lim_{\gamma\to 0} \liminf_{n\to+\infty} \mathbb{E}_{\mathcal{MP}}[h_{n,\gamma}] \geq \lim_{\gamma\to 0} h_{RS}[d_V; \Lambda_\gamma, P] = h_{RS}[d_V; \Lambda, P]. \tag{2.55}$$

The main Theorem 2.1, for the standard ensemble $(\Lambda(x), P(x))$, then follows by maximizing the right hand side over d_V.

2.5.3 Proof of the Variational Bound (Theorem 2.1) for Multi-Poisson Ensemble

In view of the sum-rule (2.47) it is sufficient to prove that $\liminf_{n\to+\infty} R_n(t_*, s) \geq 0$. In the case of a convex $P(x)$ considered in [29] this is immediate because convexity is equivalent to

$$P(Q_{2p}) - P(q_{2p}) \geq P'(q_{2p})(Q_{2p} - q_{2p}).$$

Note that $P(x) = \sum_k P_k x^k$ is anyway convex for $x \geq 0$ since all $P_k \geq 0$. So if do not assume convexity of the check node degree distribution we have to circumvent the fact that Q_{2p} can be negative. But note

$$\begin{aligned}
\langle Q_{2p}\rangle_s &= \sum_{i=1}^n w_i(t_*) X_i(t_*) \langle x_i^{(1)} x_i^{(2)} \cdots x_i^{(2p)} \rangle_s \\
&= \sum_{i=1}^n w_i(t_*) X_i(t_*) \langle x_i^{(1)} \rangle_s \langle x_i^{(2)} \rangle_s \cdots \langle x_i^{(2p)} \rangle_s \\
&= \sum_{i=1}^n w_i(t_*) X_i(t_*) \langle x_i \rangle_s^{2p} \geq 0,
\end{aligned}$$

where we recall that $w_i(t_*)$ and $X_i(t_*)$ are random variables which just depend on the code parameters and not on the Gibbs measure (because of which there is no Gibbs bracket around $w_i(t_*)$ and $X_i(t_*)$ in the expression above). Therefore we are assured that for any $P(x)$ (i.e not necessarily convex for $x \in \mathbb{R}$) we have

$$P\big(\langle Q_{2p}\rangle_s\big) - P(q_{2p}) \geq P'(q_{2p})\big(\langle Q_{2p}\rangle_s - q_{2p}\big), \tag{2.56}$$

and the proof will follow if we can show that with high probability

$$P(Q_{2p}) \approx P\big(\langle Q_{2p}\rangle_s\big).$$

The following concentration estimate will suffice and is proven in Section 2.6.

Lemma 2.2. *Fix any $0 < \delta < \frac{1}{4}$. On the BEC(ϵ) and BIAWGNC(ϵ) for a.e. ϵ, or on general BMS(ϵ) satisfying H, we have for any integer $p > 0$ and a.e. ϵ,*

$$\lim_{n\to\infty} \int_0^\gamma ds \mathbb{P}_s \left[\left| P(Q_{2p}) - P\big(\langle Q_{2p}\rangle_s\big) \right| > \frac{2p}{n^\delta} \right] = 0. \tag{2.57}$$

Here $\mathbb{P}_s(X)$ is the probability distribution $\mathbb{E}_s\langle \mathbb{I}_X\rangle_s$.

This lemma can presumably be strengthened in two directions. First we conjecture that hypothesis H is not needed (this is indeed the case for the BEC and BIAWGNC). Secondly the statement should hold for all ϵ except at a finite set of threshold values of ϵ where the conditional entropy is not differentiable, and its first derivative is expected to have jumps (except for cycle codes where higher order derivatives are singular). Since we are unable to control the locations of theses jumps our proof only works for Lebesgue almost every ϵ.

We are now ready to complete the proof of the variational bound (2.54).

End of Proof of (2.54). From (2.49) and (2.52)

$$|Q_{2p}| \le \sum_{i=1}^n w_i(t_*) X_i(t_*) \le \mathfrak{X}, \tag{2.58}$$

and

$$\mathbb{E}_s\big[\langle Q_{2p}^k\rangle_s\big] \le A_k. \tag{2.59}$$

Combined with $q_{2p} \le 1$, this implies (since the maximal degree of $P(x)$ is finite) that

$$\mathbb{E}_s\Big[\Big\langle P(Q_{2p}) - P'(q_{2p})(Q_{2p} - q_{2p}) - P(q_{2p})\Big\rangle_s\Big] \le C_1, \tag{2.60}$$

for some positive constant C_1. The only crucial feature here is that this constant does not depend on n and on the number of replicas $2p$ (a more detailed analysis shows that it depends only on the maximum degree of $P(x)$).

Now we split the sum (2.48) into terms with $1 \le p \le n^\delta$ (call this contribution R_A) and terms with $p \ge n^\delta$ (call this contribution R_B), where $\delta > 0$ is the constant of Lemma 2.2. For the second contribution, (2.60) implies

$$R_B \le C_1 \sum_{p \ge n^\delta} \frac{1}{2p(2p-1)} = O(n^{-\delta}). \tag{2.61}$$

For the first contribution we write

$$R_A = \sum_{p\le n^\delta} \frac{1}{2p(2p-1)} \mathbb{E}_s\Big[\big\langle P(Q_{2p})\big\rangle_s - P\big(\langle Q_{2p}\rangle_s\big)\Big] + \sum_{p\le n^\delta} \frac{1}{2p(2p-1)} \mathbb{E}\Big[P\big(\langle Q_{2p})\rangle_s\big) - P'(q_{2p})\big(\langle Q_{2p}\rangle_s - q_{2p}\big) - P(q_{2p})\Big].$$

In the equation above, the second sum is positive due to (2.56). Thus we find

$$\begin{aligned} R_n(t_*, s) &= R_A + R_B \\ &\ge \sum_{p\le n^\delta} \frac{1}{2p(2p-1)} \mathbb{E}[\langle P(Q_{2p})\rangle_s - P(\langle Q_{2p}\rangle_s)] - O(n^{-\delta}). \end{aligned}$$

Below, we use Lemma 2.2 to show that for almost every ϵ in the appropriate range

$$\lim_{n\to+\infty} \int_0^\gamma ds \sum_{p\le n^\delta} \frac{1}{2p(2p-1)} \mathbb{E}_s\Big[\big\langle P(Q_{2p})\big\rangle_s - P\big(\langle Q_{2p}\rangle_s\big)\Big] = 0, \tag{2.62}$$

which implies by Fatou's lemma

$$\liminf_{n\to+\infty} \sum_{t_*=0}^{t_{\max}-1} \int_0^\gamma R_n(t_*, s) ds \ge 0,$$

and thus proves (2.54) for almost every ϵ in the appropriate range.

Let us now prove (2.62). First we set

$$F_{2p} = \Big|\big\langle P(Q_{2p})\big\rangle_s - P\big(\langle Q_{2p}\rangle_s\big)\Big|,$$

and use Cauchy-Schwartz and then (2.59) to obtain

$$\begin{aligned} \mathbb{E}_s[F_{2p}] &= \mathbb{E}_s\Big[F_{2p}\mathbb{1}\Big(F_{2p} \le \frac{2p}{n^\delta}\Big)\Big] + \mathbb{E}_s\Big[F_{2p}\mathbb{1}\Big(F_{2p} \ge \frac{2p}{n^\delta}\Big)\Big] \\ &\le \frac{2p}{n^\delta} + \Big(\mathbb{E}_s\big[F_{2p}^2\big]\Big)^{1/2}\Big(\mathbb{P}_s\Big[F_{2p} \ge \frac{2p}{n^\delta}\Big]\Big)^{1/2} \\ &\le \frac{2p}{n^\delta} + C_2\Big(\mathbb{P}_s\Big[F_{2p} \ge \frac{2p}{n^\delta}\Big]\Big)^{1/2}, \end{aligned}$$

where we used that $|F_{2p}^2|$ is bounded by some positive constant C_2 independent of n and p (depending only on the maximum degree of $P(x)$). Thus using above we get

$$\begin{aligned} &\int_0^\gamma ds \sum_{p\le n^\delta} \frac{1}{2p(2p-1)} \mathbb{E}_s[F_{2p}] \\ &\quad\le \frac{1}{n^\delta} \sum_{p\le n^\delta} \frac{1}{2p-1} + C_2 \sum_{p\le n^\delta} \frac{1}{2p(2p-1)} \int_0^\gamma ds\Big(\mathbb{P}_s\Big[F_{2p} \ge \frac{2p}{n^\delta}\Big]\Big)^{1/2} \\ &\quad\le O\Big(\frac{\ln n^\delta}{n^\delta}\Big) + C_2 \sum_{p\le n^\delta} \frac{\sqrt{\gamma}}{2p(2p-1)} \Big(\int_0^\gamma ds \mathbb{P}_s\Big[F_{2p} \ge \frac{2p}{n^\delta}\Big]\Big)^{1/2}. \end{aligned}$$

In the second inequality we have first permuted the integral with a finite sum and then used Cauchy-Schwartz to get $\int_0^\gamma ds\ 1 \cdot \Big(\mathbb{P}_s\big[F_{2p} \geq \frac{2p}{n^\delta}\big]\Big)^{1/2} \leq \Big(\int_0^\gamma ds\Big)^{1/2}\Big(\int_0^\gamma ds \mathbb{P}_s\big[F_{2p} \geq \frac{2p}{n^\delta}\big]\Big)^{1/2}$. Finally we can apply Lemma 2.2 and Lebesgue's dominated convergence theorem to the last sum over p, to conclude that (2.62) holds.

2.6 Fluctuations of Overlap Parameters

In this section we prove Lemma 2.2. The proofs in this section are done directly for the Multi-Poisson ensemble. We start by a relation between the overlap fluctuation and the spin-spin correlation.

Lemma 2.3. *For any BMS(ϵ) channel there exists a finite constant C_3 independent of n and p (depending only on the maximal check degree) such that for any positive integer p we have*

$$\mathbb{P}_s\left[\Big|P(Q_{2p})-P\big(\langle Q_{2p}\rangle_s\big)\Big| \geq \frac{2p}{n^\delta}\right] \leq \frac{C_3}{p^2 n^{\frac{1}{2}-2\delta}}\left(\sum_{i=1}^n \mathbb{E}_s\Big[\big(\langle x_1 x_i\rangle_s - \langle x_1\rangle_s\langle x_i\rangle_s\big)^2\Big]\right)^{1/2}. \tag{2.63}$$

Proof. Using the identity

$$Q_{2p}^k - \langle Q_{2p}\rangle_s^k = \big(Q_{2p} - \langle Q_{2p}\rangle_s\big)\sum_{l=0}^{k-1} Q_{2p}^{k-l-1}\langle Q_{2p}\rangle_s^l, \tag{2.64}$$

and (2.52) we get

$$\begin{aligned}\big|P(Q_{2p}) - P\big(\langle Q_{2p}\rangle_s\big)\big| &= \big|Q_{2p} - \langle Q_{2p}\rangle_s\big|\ \Big|\sum_k P_k \sum_{l=0}^{k-1} Q_{2p}^{k-l-1}\langle Q_{2p}\rangle_s^l\Big| \\ &\leq \big|Q_{2p} - \langle Q_{2p}\rangle_s\big|\Big(\sum_k k P_k \mathfrak{X}^{k-1}\Big) \\ &\leq P'(\mathfrak{X})\big|Q_{2p} - \langle Q_{2p}\rangle_s\big|. \end{aligned} \tag{2.65}$$

Recall that here $\mathfrak{X}$ is the bound in (2.58). Using (2.65) and applying the Chebycheff inequality we get

$$\begin{aligned}\mathbb{P}_s\left[\Big|P(Q_{2p}) - P\big(\langle Q_{2p}\rangle_s\big)\Big| \geq \frac{2p}{n^\delta}\right] &\leq \mathbb{P}_s\left[P'(\mathfrak{X})\ \Big|Q_{2p} - \langle Q_{2p}\rangle_s\Big| \geq \frac{2p}{n^\delta}\right] \\ &\leq \frac{n^{2\delta}}{4p^2}\mathbb{E}_s\left[P'(\mathfrak{X})^2\Big\langle\Big|Q_{2p} - \langle Q_{2p}\rangle_s\Big|^2\Big\rangle_s\right] \\ &= \frac{n^{2\delta}}{4p^2}\mathbb{E}_s\Big[P'(\mathfrak{X})^2\big(\langle Q_{2p}^2\rangle_s - \langle Q_{2p}\rangle_s^2\big)\Big]. \end{aligned} \tag{2.66}$$

where going from the inequality to the equality above we use $\big\langle |Q_{2p}-\langle Q_{2p}\rangle_s|^2\big\rangle_s = \langle Q_{2p}^2\rangle_s - \langle Q_{2p}\rangle_s^2$. From the definition of the overlap parameters it follows that

$$\begin{aligned}\langle Q_{2p}^2\rangle_s &= \sum_{i,j=1}^{n} w_i(t_*)w_j(t_*)X_i(t_*)X_j(t_*)\big\langle x_i^{(1)}x_j^{(1)}x_i^{(2)}x_j^{(2)}\cdots x_i^{(2p)}x_j^{(2p)}\big\rangle_s \\ &= \sum_{i,j=1}^{n} w_i(t_*)w_j(t_*)X_i(t_*)X_j(t_*)\langle x_ix_j\rangle_s^{2p}, \end{aligned} \tag{2.67}$$

where in the last equality we used the fact that the replicas are identical copies of each other. Also in the last equality above $\langle -\rangle_s$ corresponds to a single system without replicas not to be confused with the Gibbs bracket in the previous line. Arguing similar as above we also have

$$\begin{aligned}\langle Q_{2p}\rangle_s^2 &= \Big(\sum_{i=1}^{n} w_i(t_*)X_i(t_*)\big\langle x_i^{(1)}x_i^{(2)}\cdots x_i^{(2p)}\big\rangle_s\Big)^2 \\ &= \Big(\sum_{i=1}^{n} w_i(t_*)X_i(t_*)\langle x_i\rangle_s^{2p}\Big)^2 \\ &= \sum_{i,j=1}^{n} w_i(t_*)w_j(t_*)X_i(t_*)X_j(t_*)\langle x_i\rangle_s^{2p}\langle x_j\rangle_s^{2p}. \end{aligned} \tag{2.68}$$

Thus we have

$$\begin{aligned}\langle Q_{2p}^2\rangle_s - \langle Q_{2p}\rangle_s^2 &= \sum_{i,j=1}^{n} w_i(t_*)w_j(t_*)X_i(t_*)X_j(t_*)\big(\langle x_ix_j\rangle_s^{2p} - \langle x_i\rangle_s^{2p}\langle x_j\rangle_s^{2p}\big) \\ &\le 2p\sum_{i,j=1}^{n} \mathfrak{X}^2 w_i(t_*)w_j(t_*)\big|\langle x_ix_j\rangle_s - \langle x_i\rangle_s\langle x_j\rangle_s\big|. \end{aligned}$$

Substituting in (2.66) and applying Cauchy-Schwartz to $\sum_{i,j}\mathbb{E}_s[-]$ we get

$$\begin{aligned}\mathbb{P}_s\big[|P(Q_{2p}) - P(\langle Q_{2p}\rangle_s)| \ge \frac{2p}{n^\delta}\big] \le &\frac{n^{2\delta}}{2p}\bigg(\sum_{i,j=1}^{n}\mathbb{E}_s\Big[\mathfrak{X}^4P'(x)^4w_i(t_*)^2w_j(t_*)^2\Big]\bigg)^{1/2} \\ &\times\bigg(\sum_{i,j=1}^{n}\mathbb{E}_s\Big[\big(\langle x_ix_j\rangle_s - \langle x_i\rangle_s\langle x_j\rangle_s\big)^2\Big]\bigg)^{1/2}. \end{aligned}$$

From (2.51), (2.52) it is not hard to see that for any i, j

$$\mathbb{E}_s\big[\mathfrak{X}^4P'(x)^4w_i(t_*)^2w_j(t_*)^2\big] \le \frac{C_3^2}{n^4},$$

where C_3 is independent of n. It follows that

$$\begin{aligned}\mathbb{P}_s\left[\left|P(Q_{2p}) - P\big(\langle Q_{2p}\rangle_s\big)\right| \geq \frac{2p}{n^\delta}\right] \\ &\leq \frac{n^{2\delta-1}}{2p} C_3 \left(\sum_{i,j=1}^{n} \mathbb{E}_s\Big[\big(\langle x_i x_j\rangle_s - \langle x_i\rangle_s\langle x_j\rangle_s\big)^2\Big]\right)^{1/2} \\ &= \frac{n^{2\delta-\frac{1}{2}}}{2p} C_3 \left(\sum_{i=1}^{n} \mathbb{E}_s\Big[\big(\langle x_i x_1\rangle_s - \langle x_i\rangle_s\langle x_1\rangle_s\big)^2\Big]\right)^{1/2}. \qquad (2.69)\end{aligned}$$

In the last equality we have used the symmetry of the ensemble with respect to variable node permutations to eliminate one sum of the sums in the double sum $\sum_{i,j}$. □

Denote by $h_{n,\gamma}(t_*, s)$ the entropy of the $\mu_{t_*,s}$ interpolating measure. Note that this should not be confused with the Multi-Poisson ensemble entropy $h_{n,\gamma}$ (which corresponds to $t_* = t_{\max}$ and $s = \gamma$).

Lemma 2.4. *For the BEC and BIAWGNC with any noise value and for general BMS(ϵ) channels satisfying H we have*

$$\sum_{i=1}^{n} \mathbb{E}_s[(\langle x_1 x_i\rangle_s - \langle x_1\rangle_s\langle x_i\rangle_s)^2] \leq F(\epsilon) + G(\epsilon)\frac{d^2}{d\epsilon^2}\mathbb{E}_s[h_{n,\gamma}(t_*, s)], \qquad (2.70)$$

where $F(\epsilon)$ and $G(\epsilon)$ are two finite constants depending only on the channel parameter.

The proof of Lemma 2.4 is based on the correlation formula of Section 2.1. These are true for any linear code ensemble so they are in particular true for the interpolating (t_*, s) ensemble[3]. For the BEC and BIAWGNC we have already shown the two equalities (2.8) and (2.10): thus the inequality (2.70) is in fact an equality for appropriate values of F and G. The case of general (but highly noisy) BMS channels is presented in Appendix 2.C. A converse inequality can also be proven by the methods of Appendices 2.C and 2.D.

Proof of Lemma 2.2. Note that for all points of the parameter space (ϵ, s) such that the second derivative of the average conditional entropy is bounded uniformly in n the proof immediately follows from (2.69), (2.70) by choosing $\delta < \frac{1}{4}$. However, in the large block length limit $n \to +\infty$, generically the first derivative of the average conditional entropy has jumps for some values of ϵ (these values depend on the interpolation parameter s). This means that for

[3]In fact one has to check that the addition of $\sum_{b=1}^{e_i} U_b^i$ to l_i does not change the derivation and the final formulas. For this it suffices to follow the calculation of Section 2.4. The distribution of the U_a depends on the distribution of V, which is freely distributed and apriori does not depend on the channel noise ϵ, hence our derivations for the correlation formula for the BEC and BIAWGNC follow also for the interpolated ensemble.

these threshold values the second derivative cannot be bounded uniformly in n. Since we cannot control these locations we introduce a test function $\psi(\epsilon)$: non negative, infinitely differentiable and with small enough bounded support (let $\epsilon_a < \epsilon_b$ denote the end-points of the support of $\psi(\epsilon)$ and $\psi(\epsilon_a) = \psi(\epsilon_b) = 0$) included in the range of ϵ satisfying H. We consider the averaged quantity

$$\mathcal{Q} = \int d\epsilon \psi(\epsilon) \int_0^\gamma ds \mathbb{P}_s \left[\left| P(Q_{2p}) - P\big(\langle Q_{2p} \rangle_s\big) \right| \geq \frac{2p}{n^\delta} \right]. \tag{2.71}$$

Since both $\psi(\dot{)}$ and $\mathbb{P}_s[\dot{]}$ are bounded functions, we can exchange the two integrals. Then writing $\psi(\epsilon) = \sqrt{\psi(\epsilon)}\sqrt{\psi(\epsilon)}$, and using Cauchy-Schwartz we find

$$\mathcal{Q} \leq \int_0^\gamma ds \left(\int d\epsilon \psi(\epsilon) \left(\mathbb{P}_s \left[\left| P(Q_{2p}) - P\big(\langle Q_{2p} \rangle_s\big) \right| \geq \frac{2p}{n^\delta} \right] \right)^2 \right)^{1/2}.$$

Combining this inequality with (2.69) and (2.70) we get

$$\begin{aligned}
\mathcal{Q} &\leq \frac{n^{2\delta-\frac{1}{2}}}{2p} C_3 \int_0^\gamma ds \left(\int d\epsilon \psi(\epsilon) \big(F(\epsilon) + G(\epsilon) \frac{d^2}{d\epsilon^2} \mathbb{E}_s[h_{n,\gamma}(t_*, s)] \big) \right)^{1/2} \\
&\overset{(a)}{=} \frac{n^{2\delta-\frac{1}{2}}}{2p} C_3 \int_0^\gamma ds \left(\int d\epsilon \psi(\epsilon) F(\epsilon) + \psi(\epsilon) G(\epsilon) \frac{d}{d\epsilon} \mathbb{E}_s[h_{n,\gamma}(t_*, s)] \Big|_{\epsilon=\epsilon_a}^{\epsilon=\epsilon_b} \right. \\
&\quad \left. - \int d\epsilon \frac{d}{d\epsilon} \big(\psi(\epsilon) G(\epsilon) \big) \frac{d}{d\epsilon} \mathbb{E}_s[h_{n,\gamma}(t_*, s)] \right)^{1/2} \\
&= \frac{n^{2\delta-\frac{1}{2}}}{2p} C_3 \int_0^\gamma ds \left(\int d\epsilon \psi(\epsilon) F(\epsilon) - \int d\epsilon \frac{d}{d\epsilon} \big(\psi(\epsilon) G(\epsilon) \big) \frac{d}{d\epsilon} \mathbb{E}_s[h_{n,\gamma}(t_*, s)] \right)^{1/2}.
\end{aligned}$$

where (a) follows from the integration by parts formula and the last equality follows because the first derivative of the average conditional entropy is bounded uniformly in n and s (see Appendix 2.E) by a constant $k(\epsilon)$ that has at most a power singularity at $\epsilon = 0$, and again this is not a problem as we can take the test function to be vanishing sufficiently fast. Note that from the bounds in Appendix 2.C, $F(\epsilon)$, $G(\epsilon)$ and $G'(\epsilon)$ are integrable except possibly at the edge of the ϵ range defined by H. This is not a problem because we can take the support of $\psi(\epsilon)$ away from such points or alternatively take a $\psi(\epsilon)$ which vanishes sufficiently fast at these points. Thus by choosing $0 < \delta < \frac{1}{4}$ we obtain

$$\lim_{n \to +\infty} \mathcal{Q} = 0.$$

Applying Lebesgue's dominated convergence theorem to convergent subsequences (of the integrand of $\int d\epsilon \psi(\epsilon)$ in (2.71)) we deduce that

$$\int d\epsilon \psi(\epsilon) \lim_{n_k \to +\infty} \int_0^\gamma ds \mathbb{P}_s \left[\left| P(Q_{2p}) - P\big(\langle Q_{2p} \rangle_s\big) \right| \geq \frac{2p}{n_k^\delta} \right] = 0,$$

which implies that along any convergent subsequences, for almost all ϵ

$$\lim_{n_k \to +\infty} \int_0^\gamma ds \mathbb{P}_s \left[\left| P(Q_{2p}) - P\big(\langle Q_{2p} \rangle_s\big) \right| \geq \frac{2p}{n_k^\delta} \right] = 0, \tag{2.72}$$

as long as $\delta < \frac{1}{4}$. Now we apply this last statement to two subsequences that attain the lim inf and the lim sup (on the intersection of the two measure one ϵ sets). This proves that the $\lim_{n \to +\infty}$ exists and vanishes. □

2.7 Conclusion

The main new tool introduced in this chapter are relationships between the second derivative of the conditional entropy and correlation functions or mutual information between code bits. This allowed us to estimate the overlap fluctuations in order to get a better handle on the remainder. Some aspects of our analysis bear some similarity with techniques introduced by Talagrand [87] but is independent. One difference is that we use specific symmetry properties of the communications problem.

We expect that the technique developed here can be extended to remove the restriction to high noise (condition H). Indeed the only place in the analysis where we need this restriction is Lemma 2.4. For the BEC and BIAWGNC the lemma is trivially satisfied for any noise level (with appropriate constants). Another issue that would be worthwhile investigating is whether the related inequalities of paragraph 2.1.3 and the converse of Lemma 2.4 can be derived irrespective of the noise level.

The next obvious problem is to prove the converse of the variational bound (Theorem 2.1). For this one should show that the remainder vanishes when d_V is replaced by the maximizing distribution of $h_{RS}[d_V; \Lambda, P]$. This program has been carried out explicitly in the case of the BEC and the Poisson-LDPC ensemble in the next chapter. It would be desirable to extend this to more general ensembles and channels but the problem becomes quite hard.

2.A Useful Tools and The Interpolation Method

2.A.1 Griffiths-Kelly-Sherman (GKS) Correlation Inequalities

Consider a spin system given by the Hamiltonian $\mathcal{H} = -\sum_a l_a x_{\partial a}$ with $l_a \geq 0$ for all a. Such a system in statistical mechanics is known as a *ferromagnetic* system. For a ferromagnetic system the GKS inequalities state that,

$$\begin{aligned}&\langle x_A \rangle \geq 0,\\ &\langle x_A x_B \rangle - \langle x_A \rangle \langle x_B \rangle \geq 0,\end{aligned}$$

where A, B are any two finite support subsets of the set of spins and $\langle - \rangle$ is the Gibbs average associated with the Hamiltonian $\mathcal{H}$.

The case of transmission using LDGM codes over the BEC is clearly an instance of a ferromagnetic system, as seen by (1.25). Therefore the GKS inequalities apply directly. For the case of LDPC codes, the contribution of any check node a can be written as

$$\frac{1}{2}(1 + x_{\partial a}) = \lim_{l_a \to +\infty} e^{l_a(x_{\partial a} - 1)}.$$

Thus we can rewrite the Gibbs measure (2.11) as

$$\mu(\underline{x}) = \lim_{\substack{l_a \to +\infty \\ \text{for all } a}} \frac{1}{Z'} \sum_{\underline{x}} \prod_{a=1}^{m} e^{l_a x_{\partial a}} \prod_{i=1}^{n} e^{l_i x_i},$$

where $Z' = \sum_{\underline{x}} \prod_{a=1}^{m} e^{l_a x_{\partial a}} \prod_{i=1}^{n} e^{l_i x_i}$ with $l_a \geq 0$ for all $1 \leq a \leq m$ and $l_i \geq 0$ for all $1 \leq i \leq n$. As a result of this reformulation, the case of LDPC codes over the BEC also represents a ferromagnetic system and the GKS inequalities apply.

In general the first GKS inequality is not true if the interactions or the LLR can take both positive and negative values. But when transmission takes place over a BMS channel, then the first inequality is satisfied on average, when the average is w.r.t. the noise realizations. We say more about this in the next section on Nishimori identities. The second inequality may not be true even on average in the case of general BMS channels.

Let us give a proof of the first GKS inequality. For the proof of the second inequality we refer the reader to the literature [93], [94], [95]. We consider the Gibbs measure given by

$$\mu(\underline{x}) = \frac{1}{Z} \sum_{\underline{x}} \prod_{a} e^{l_a x_{\partial a}},$$

where all the notations adhere to the previous ones, $l_a \geq 0$, and the total number of spins is n.

We have $\langle x_A \rangle \geq 0$ if and only if $Z\langle x_A \rangle \geq 0$. Using the identity $e^{l_a x_{\partial a}} = \cosh l_a (1 + x_{\partial a} \tanh l_a)$, we immediately obtain

$$\begin{aligned} Z\langle x_A \rangle &= \prod_a \cosh l_a \sum_{\underline{x}} x_{\partial a} \prod_a (1 + x_{\partial a} t_a) \\ &= \sum_{\underline{x}} \sum_{a_1, a_2, \ldots, a_k} x_{\partial a} x_{\partial a_1} x_{\partial a_2} \cdots x_{\partial a_k} t_{a_1} t_{a_2} \cdots t_{a_k}, \end{aligned}$$

where the sum over $a_1, a_2, \ldots, a_k$ is over all possible subsets of interactions and $t = \tanh l$. Since the interaction terms l_a are independent of the spin values, we have

$$Z\langle x_A \rangle = \sum_{a_1, a_2, \ldots, a_k} t_{a_1} t_{a_2} \cdots t_{a_k} \sum_{\underline{x}} x_{\partial a} x_{\partial a_1} x_{\partial a_2} \cdots x_{\partial a_k}.$$

Now we obtain the first GKS inequality by realizing that all the $t_a \geq 0$ and that for any set $a_1, a_2, \ldots, a_k$, $\sum_{\underline{x}} x_{\partial a} x_{\partial a_1} x_{\partial a_2} \cdots x_{\partial a_k} \in \{0, 2^n\}$.

2.A.2 Nishimori Identities

The symmetry of the channel mentioned in Fact 1.1 implies remarkable Nishimori identities. Consider transmission of a fixed linear code (need not be low-density) over a general BMS channel. Consider the random Gibbs measure, under the assumption of the all-one codeword transmission, given by (2.11). Then for any collection of subsets $A_1, A_2, \ldots, A_m$ of the spins and any positive integers $b_1, b_2, \ldots, b_m$ we have the Nishimori identity

$$\mathbb{E}_{\underline{l}}\Big(\langle x_{A_1} \rangle^{b_1} \langle x_{A_2} \rangle^{b_2} \cdots \langle x_{A_m} \rangle^{b_m} \Big) = \mathbb{E}_{\underline{l}}\Big(\langle x_{A_1}^{b_1} x_{A_2}^{b_2} \cdots x_{A_m}^{b_m} \rangle \langle x_{A_1} \rangle^{b_1} \langle x_{A_2} \rangle^{b_2} \cdots \langle x_{A_m} \rangle^{b_m} \Big).$$

Note that $x_{A_j}^{b_j} = 1$ if b_j is even and $x_{A_j}^{b_j} = x_{A_j}$ if b_j is odd. We do not provide the proof of this identity here, but refer the reader to [84].

An immediate consequence of the Nishimori identity is that

$$\mathbb{E}_{\underline{l}}\big[\langle x_A \rangle \big] = \mathbb{E}_{\underline{l}}\big[\langle x_A \rangle^2 \big] \geq 0.$$

We have the following lemma for the BEC which is crucial for providing an expression for the second derivative of the conditional entropy in terms of spin-spin correlation and also for simplifying the remainder term of the interpolation method as shown in Chapter 3.

Lemma 2.5. *For transmission over the BEC using a fixed linear code we have that for every noise realization, $\langle x_A \rangle$ either equals zero or one, where A is any set of code-bits.*

Proof. From the Nishimori identity, we have

$$\mathbb{E}_{\underline{l}}\big[\langle x_A \rangle \big] = \mathbb{E}_{\underline{l}}\big[\langle x_A \rangle^2 \big]. \tag{2.73}$$

Re-writing the above we get

$$\sum_{\underline{l}\in\{0,+\infty\}^n} \mathbb{P}(\underline{l})\Big(\langle x_A\rangle(1-\langle x_A\rangle)\Big) = 0, \tag{2.74}$$

where $\mathbb{P}(\underline{l})$ denotes the probability of the noise realization $\underline{l}$, which belongs to $\{0,+\infty\}^n$. Since we are transmitting over the BEC we use the first GKS inequality to get

$$\langle x_A\rangle \geq 0, \tag{2.75}$$

which implies that $\big(\langle x_A\rangle(1-\langle x_A\rangle)\big) \geq 0$. As a result from (2.74) we get, for every noise realization,

$$\langle x_A\rangle(1-\langle x_A\rangle) = 0, \tag{2.76}$$

and hence the lemma. □

2.A.3 Gaussian Integration

Consider transmission over a BIAWGNC(ϵ) using any fixed linear code of blocklength n. Recall that $c_L(l)$ denotes the distribution of the LLR l. Let $f(\underline{l},\underline{x})$ be any function of the code-bits and noise realization vector $\underline{l}$, such that $c_L(l_i)f(\underline{l},\underline{x}) = 0$ and $\frac{dc_L(l_i)}{dl_i}f(\underline{l},\underline{x}) = 0$ for $l_i = +\infty$ or $l_i = -\infty$. For example, $f(\underline{l},\underline{x}) = \ln Z$ (Z is the partition function for the Gibbs measure introduced in 2.11 and is bounded by n) or $f(\underline{l},\underline{x}) = \langle x_i\rangle$ (bounded by 1), then we have

Fact 2.1 (Gaussian Intergration by Parts)**.**

$$\frac{d}{d\epsilon}\mathbb{E}_{\underline{l}}[f(\underline{l},\underline{x})] = -4\epsilon^{-3}\sum_{i=1}^{n}\mathbb{E}_{\underline{l}}\Big[\Big(\frac{\partial}{\partial l_i}+\frac{\partial^2}{\partial l_i^2}\Big)f(\underline{l},\underline{x})\Big]. \tag{2.77}$$

Proof. For convenience assume that each code-bit is transmitted through independent channels with noise variance ϵ_i^2 (finally we will put $\epsilon_i = \epsilon$). From the smoothness of the channel we get

$$\frac{d}{d\epsilon}\mathbb{E}_{\underline{l}}[f(\underline{l},\underline{x})] = \sum_{i=1}^{n}\mathbb{E}_{\underline{l}\setminus l_i}\int dl_i \frac{\partial c_L(l_i)}{\partial\epsilon_i}f(\underline{l},\underline{x}).$$

For the BIAWGNC, recall that we have, for all i,

$$c_L(l_i) = \frac{1}{\sqrt{8\pi\epsilon_i^{-2}}}e^{-\frac{(l_i-2\epsilon_i^{-2})^2}{8\epsilon_i^{-2}}}.$$

It is easy to check that

$$\frac{\partial c_L(l_i)}{\partial\epsilon_i} = -4\epsilon_i^{-3}\Big(-\frac{\partial c_L(l_i)}{\partial l_i}+\frac{\partial^2 c_L(l_i)}{\partial l_i^2}\Big).$$

Thus we have

$$\sum_{i=1}^{n} \mathbb{E}_{\underline{l}\setminus l_i} \int dl_i \frac{\partial c_L(l_i)}{\partial \epsilon_i} f(\underline{l},\underline{x}) = -4\epsilon^{-3} \sum_{i=1}^{n} \mathbb{E}_{\underline{l}\setminus l_i} \int dl_i \Big(-\frac{d}{dl_i} + \frac{d^2}{dl_i^2} \Big) c_L(l_i) f(\underline{l},\underline{x}).$$

Note that in the above formula we have used total derivatives instead of partial derivatives, because all LLR are independent of each other. Using integration by parts we have

$$\begin{aligned} \int dl_i \frac{dc_L(l_i)}{dl_i} f(\underline{l},\underline{x}) &= c_L(l_i) f(\underline{l},\underline{x}) \Big|_{l_i=-\infty}^{l_i=+\infty} - \int dl_i c_L(l_i) \frac{df(\underline{l},\underline{x})}{dl_i} \\ &= -\int dl_i c_L(l_i) \frac{df(\underline{l},\underline{x})}{dl_i}. \end{aligned}$$

Similarly, using integration by parts twice we obtain

$$\begin{aligned} \int dl_i \frac{d^2 c_L(l_i)}{dl_i^2} f(\underline{l},\underline{x}) &= \frac{dc_L(l_i)}{dl_i} f(\underline{l},\underline{x}) \Big|_{l_i=-\infty}^{l_i=+\infty} - \int dl_i \frac{dc_L(l_i)}{dl_i} \frac{df(\underline{l},\underline{x})}{dl_i} \\ &= \int dl_i c_L(l_i) \frac{d^2 f(\underline{l},\underline{x})}{dl_i^2}. \end{aligned}$$

Substituting above we get the lemma. □

2.A.4 Interpolation Method for the Poisson-LDPC ensemble

In this section we demonstrate the interpolation method, in our notation, in the context of Poisson-LDPC code ensemble and transmission over any general BMS channel.

The interpolation method for the more complicated Multi-Poisson ensemble will not be shown here, but the basic idea remains the same. Nevertheless, we will give a little explanation about it a bit later. This will also shed some light on the complicated overlap parameters (see (2.49)).

Consider the interpolated partition function for the Poisson-LDPC $(n, (1-\mathtt{R})s, P(x))$

$$Z(s) = \sum_{\underline{x}} \prod_{a=1}^{m_s} \frac{(1+x_{\partial a})}{2} \prod_{i=1}^{n} e^{(\frac{l_i}{2} + \sum_{j=1}^{e_i} \frac{u_j}{2}) x_i}, \tag{2.78}$$

where recall that $\tanh \frac{u_j}{2} = \prod_{t=1}^{k-1} \tanh \frac{v_{jt}}{2}$ and v_{jt} are i.i.d. random variables distributed as d_V and m_s is the total number of check nodes which is a Poisson random variable with mean $n(1-R)s$ and e_i is a Poisson random variable with mean $P'(1)(1-\mathtt{R})(1-s)$ for every i. We emphasize that the randomness in the interpolated ensemble arises in the number of check nodes m_s present and in the number of "auxiliary" observations e_i, added to each variable node. We denote the expectation with respect to this randomness by $\mathbb{E}_s$. From

(1.24) we can write $\mathbb{E}_s[h_n(s)] = \frac{1}{n}\mathbb{E}_{s,\underline{l}}[\ln Z(s)] - \mathbb{E}_l[\frac{l}{2}]$, where $h_n(s)$ is the per-bit conditional entropy of the interpolated ensemble. Using the fundamental theorem of calculus we immediately obtain

$$\mathbb{E}_{\mathcal{P}}[h_n] = \mathbb{E}[h_n(s=0)] + \int_{s=0}^{1} ds \frac{d\mathbb{E}_s[h_n[s]]}{ds}, \tag{2.79}$$

where $\mathbb{E}_{\mathcal{P}}$ denotes the expectation with respect to the original Poisson-LDPC ensemble corresponding to $s = 1$. Since the term $\mathbb{E}_l[\frac{l}{2}]$ is independent of s we get

$$\frac{d\mathbb{E}_s[h_n(s)]}{ds} = \frac{1}{n}\frac{d\mathbb{E}_{s,\underline{l}}[\ln Z(s)]}{ds}.$$

The dependence on s occurs only in the mean of the Poisson random variables, m_s and e_i. For any function $f(X)$ of a Poisson random variable X, with mean ν, we have the following remarkable identity

$$\frac{d\mathbb{E}[f(X)]}{d\nu} = \mathbb{E}[f(X+1)] - \mathbb{E}[f(X)], \tag{2.80}$$

Using the above, we get

$$\begin{aligned}\frac{1}{n}\frac{d\mathbb{E}_{s,\underline{l}}[\ln Z(s)]}{ds} &= (1-\mathtt{R})\Big(\mathbb{E}_{s,\underline{l}}[\ln Z(m_s+1)] - \mathbb{E}_{s,\underline{l}}[\ln Z(m_s)]\Big) \\ &\quad - \frac{P'(1)(1-\mathtt{R})}{n}\sum_{i=1}^{n}\Big(\mathbb{E}_{s,\underline{l}}[\ln Z(e_i+1)] - \mathbb{E}_{s,\underline{l}}[\ln Z(e_i)]\Big).\end{aligned} \tag{2.81}$$

where $Z(m_s+1)$ corresponds to the partition function obtained from adding, u.a.r., one more check node to the partition function, $Z(m_s)$ (we abuse the notation to denote $Z(s)$ in (2.78) by $Z(m_s)$), with m_s number of check nodes already present. This extra check node is called as "dummy" check node in [29]. Similarly, $Z(e_i+1)$ corresponds to the partition function got by adding an extra "auxiliary" observation to the variable node i in the partition function $Z(e_i)$. Since $Z(m_s+1)$ contains an extra check node, it is not hard to see that

$$\begin{aligned}Z(m_s+1) &= \sum_{\underline{x}} \frac{(1+x_{i_1}x_{i_2}\cdots x_{i_k})}{2}\mu_s(\underline{x})Z(m_s) \\ &= Z(m_s)\left\langle \frac{1+x_{i_1}x_{i_2}\cdots x_{i_k}}{2}\right\rangle_s,\end{aligned}$$

where $(1+x_{i_1}x_{i_2}\cdots x_{i_k})/2$ is the parity-check function associated to the extra check node. Since the extra check node is added u.a.r. to any k variable nodes

we have,

$$\begin{aligned}\mathbb{E}_{s,\underline{l}}[\ln Z(m_s+1)] - \mathbb{E}_{s,\underline{l}}[\ln Z(m_s)] &= \mathbb{E}_{s,\underline{l}}\Big[\ln\Big(\frac{Z(m_s+1)}{Z(m_s)}\Big)\Big] \\ &= \sum_k P_k \frac{1}{n^k} \sum_{i_1,i_2,\ldots,i_k} \mathbb{E}\Big[\ln\Big(\frac{1+\langle x_{i_1}x_{i_2}\cdots x_{i_k}\rangle_s}{2}\Big)\Big] \\ &= \sum_k P_k \mathbb{E}\Big[\sum_{p=1}\frac{(-1)^{p+1}}{p}\frac{1}{n^k}\sum_{i_1,i_2,\ldots,i_k}\Big(\langle x_{i_1}x_{i_2}\cdots x_{i_k}\rangle_s\Big)^p\Big] - \ln 2, \qquad (2.82)\end{aligned}$$

where the variable nodes $x_{i_1},\ldots,x_{i_k}$ represent the extra check node which is attached uniformly to any k variable nodes (of the system represented by the partition function $Z(m_s)$). To get the second equality we expand the logarithm present in the previous equality. Also for convenience we abuse the notation to denote $\mathbb{E}$ to be the average with respect to any randomness. As usual $\langle - \rangle_s$ denotes the average w.r.t. to $Z(m_s)$. Similarly,

$$\begin{aligned}\frac{1}{n}\sum_i \mathbb{E}_{s,\underline{l}}[\ln Z(e_i+1)] - \mathbb{E}_{s,\underline{l}}[\ln Z(e_i)] &= \frac{1}{n}\sum_i \mathbb{E}\Big[\ln\Big(\langle e^{x_i\frac{u_i}{2}}\rangle_s\Big)\Big] \\ &= \frac{1}{n}\sum_i\Big(\mathbb{E}\Big[\ln\Big(1+\tanh\frac{u_i}{2}\langle x_i\rangle_s\Big)\Big] + \mathbb{E}\Big[\ln\cosh\frac{u_i}{2}\Big]\Big) \\ &= \frac{1}{n}\sum_i \mathbb{E}\Big(\sum_{p=1}\frac{(-1)^{p+1}}{p}\Big(\tanh\frac{u_i}{2}\Big)^p\langle x_i\rangle_s^p\Big) + \mathbb{E}\Big[\ln\cosh\frac{u}{2}\Big], \\ &= \mathbb{E}\Big[\sum_{p=1}\frac{(-1)^{p+1}}{p}\prod_{j=1}^{k-1}\Big(\tanh\frac{v_j}{2}\Big)^p\Big(\frac{1}{n}\sum_i\langle x_i\rangle_s^p\Big)\Big] + \mathbb{E}\Big[\ln\cosh\frac{u}{2}\Big],\end{aligned}$$

where we use $e^{x_i\frac{u_i}{2}} = \cosh\frac{u_i}{2}(1+x_i\tanh\frac{u_i}{2})$ to get the second equality and we replaced $\tanh\frac{u_i}{2}$ with $\prod_{j=1}^{k-1}\tanh\frac{v_j}{2}$ to obtain the last equality. Since the random variable u_i does not appear in $\langle - \rangle_s$ (recall that u_i was the extra "auxiliary" observation added independently to the original partition function), and since v_j are i.i.d., we find that the last expression above is equal to

$$\sum_{p=1}\frac{(-1)^{p+1}}{p}\sum_k \rho_k\Big(\mathbb{E}\Big(\tanh\frac{v}{2}\Big)^p\Big)^{k-1}\mathbb{E}\Big[\Big(\frac{1}{n}\sum_i\langle x_i\rangle_s^p\Big)\Big] + \mathbb{E}\Big[\ln\cosh\frac{u}{2}\Big]$$

From the definition of the overlap parameters in (2.45) we have

$$\begin{aligned}\langle (Q_p)^k\rangle_s &= \frac{1}{n^k}\sum_{i_1,i_2,\ldots,i_k}\big\langle (x_{i_1}^{(1)}x_{i_2}^{(1)}\cdots x_{i_k}^{(1)})(x_{i_1}^{(2)}x_{i_2}^{(2)}\cdots x_{i_k}^{(2)})\cdots(x_{i_1}^{(p)}x_{i_2}^{(p)}\cdots x_{i_k}^{(p)})\big\rangle_s \\ &= \frac{1}{n^k}\sum_{i_1,i_2,\ldots,i_k}\langle x_{i_1}x_{i_2}\cdots x_{i_k}\rangle_s^p.\end{aligned}$$

Using $\rho_k = kP_k/P'(1)$ and combining all the above we get

$$\frac{1}{n}\frac{d\mathbb{E}_{s,\underline{l}}[\ln Z(s)]}{ds} = (1-\mathtt{R})\Big(\sum_{p=1}\frac{(-1)^{p+1}}{p}\sum_k P_k\mathbb{E}\langle Q_p^k\rangle_s\Big) - (1-\mathtt{R})\ln 2$$
$$-(1-\mathtt{R})\Big(\sum_{p=1}\frac{(-1)^{p+1}}{p}\sum_k P_k k\Big(\mathbb{E}\Big(\tanh\frac{v}{2}\Big)^p\Big)^{k-1}\mathbb{E}\langle Q_p\rangle_s\Big)$$
$$-P'(1)(1-\mathtt{R})\mathbb{E}\Big[\ln\cosh\frac{u}{2}\Big]. \tag{2.83}$$

Adding and subtracting the term $(1-\mathtt{R})\sum_{p=1}\frac{(-1)^{p+1}}{p}\sum_k P_k(k-1)\big(\mathbb{E}[\tanh^p\frac{v}{2}]\big)^k$ which can be equivalently written as $(1-\mathtt{R})\sum_k P_k(k-1)\mathbb{E}\ln\big(1+\prod_{i=1}^k\tanh\frac{v_i}{2}\big)$ we get that the derivative above is equal to

$$(1-\mathtt{R})\sum_{p=1}\frac{(-1)^{p+1}}{p}\mathbb{E}\Big[\big\langle P(Q_{2p}) - P'(q_{2p})(Q_{2p}-q_{2p}) - P(q_{2p})\big\rangle_{2p,s}\Big]$$
$$-P'(1)(1-\mathtt{R})\mathbb{E}\Big[\ln\cosh\frac{u}{2}\Big] - (1-\mathtt{R})\sum_k P_k(k-1)\mathbb{E}\ln\big(1+\prod_{i=1}^k\tanh\frac{v_i}{2}\big). \tag{2.84}$$

We compute the conditional entropy at $s=0$ as follows,

$$\mathbb{E}[h_n(s=0)] = \mathbb{E}\Big[\ln\Big(e^{\frac{l}{2}}e^{\sum_j u_j} + e^{-\frac{l}{2}}e^{-\sum_j u_j}\Big)\Big]$$
$$= \mathbb{E}\Big[\ln\Big(e^{\frac{l}{2}}\prod_{j=1}^l(1+\tanh\frac{u_j}{2}) + e^{-\frac{l}{2}}\prod_{j=1}^\ell(1-\tanh\frac{u_j}{2})\Big)\Big] + \mathbb{E}\Big[\sum_{j=1}^\ell\ln\cosh\frac{u_j}{2}\Big]$$
$$\overset{\ell\sim\Lambda_\ell,\mathbb{E}[\ell]=\Lambda'(1)=P'(1)(1-\mathtt{R})}{=} \mathbb{E}\ln\Big(e^{\frac{l}{2}}\prod_{j=1}^l(1+\tanh\frac{u_j}{2}) + e^{-\frac{l}{2}}\prod_{j=1}^\ell(1-\tanh\frac{u_j}{2})\Big)$$
$$+P'(1)(1-\mathtt{R})\mathbb{E}\Big[\ln\cosh\frac{u}{2}\Big], \tag{2.85}$$

where to obtain the last equality we use that u_j are i.i.d. Combining (2.79), (2.84) and (2.85) we obtain the sum-rule of Montanari.

To get the remainder in the required for as in (2.44) we do the following. First we use the Nishimori identities to find

$$\mathbb{E}\langle(Q_{2p})^k\rangle_s = \frac{1}{n^k}\sum_{i_1,i_2,\ldots,i_k}\mathbb{E}\langle x_{i_1}x_{i_2}\cdots x_{i_k}\rangle_s^{2p}$$
$$= \frac{1}{n^k}\sum_{i_1,i_2,\ldots,i_k}\mathbb{E}\langle x_{i_1}x_{i_2}\cdots x_{i_k}\rangle_s^{2p-1} = \mathbb{E}\langle(Q_{2p-1})^k\rangle_s,$$

and $\mathbb{E}[\tanh^{2p}\frac{v}{2}] = \mathbb{E}[\tanh^{2p-1}\frac{v}{2}]$. This allows us then to combine odd and even terms in the sum over p in (2.83) and we get the required form for the remainder.

For the interpolated Multi-Poisson ensemble, at some t_*, when we take the derivative w.r.t. s, we again have term corresponding to an extra check node and an extra auxiliary observation (see (2.81)). But here, the extra check node is present between the variable nodes $i_1, \ldots, i_k$ with probability governed by the free sockets at that time, i.e., by $w_{i_1}(t_*) \cdots w_{i_k}(t_*)$. The presence of this "dummy" check node changes the distribution of $w_i(t)$ for all $t > t_*$ and hence we cannot combine the two partition functions as easily as we did in (2.82). To overcome this, the random variables X_i are introduced which "tilt" the old distribution of $w_i(t)$ to the new ones in order to facilitate the combining of the partition functions in the analysis. As a result we have a more complicated overlap parameter.

2.B Proof of Identities (2.15), (2.16), (2.17).

We prove the identities (2.15), (2.16), (2.17). By definition

$$\langle x_i e^{-\frac{l_i}{2}x_i}\rangle = \frac{1}{Z}\sum_{\underline{x}} x_i \prod_c \frac{1}{2}(1+x_{\partial c}) \prod_{j\neq i} e^{\frac{l_i}{2}x_i},$$

and

$$\langle e^{-\frac{l_i}{2}x_i}\rangle = \frac{1}{Z}\sum_{\underline{x}} \prod_c \frac{1}{2}(1+x_{\partial c}) \prod_{j\neq i} e^{\frac{l_i}{2}x_i}.$$

Thus

$$\langle x_i\rangle_{\sim i} = \frac{\langle x_i e^{-\frac{l_i}{2}x_i}\rangle}{\langle e^{-\frac{l_i}{2}x_i}\rangle},$$

and plugging the identity

$$e^{-\frac{l_i}{2}x_i} = e^{-\frac{l_i}{2}}\frac{1-x_i t_i}{1-t_i},$$

in the brackets immediately leads to (2.15). For the second and third identities we proceed similarly. Namely,

$$\langle x_i\rangle_{\sim ij} = \frac{\langle x_i e^{-\frac{l_i}{2}x_i} e^{-\frac{l_j}{2}x_j}\rangle}{\langle e^{-\frac{l_i}{2}x_i} e^{-\frac{l_j}{2}x_j}\rangle},$$

and

$$\langle x_i x_j\rangle_{\sim ij} = \frac{\langle x_i x_j e^{-\frac{l_i}{2}x_i} e^{-\frac{l_j}{2}x_j}\rangle}{\langle e^{-\frac{l_i}{2}x_i} e^{-\frac{l_j}{2}x_j}\rangle}.$$

Plugging

$$e^{-\frac{l_i}{2}x_i} e^{-\frac{l_j}{2}x_j} = e^{\frac{l_i+l_j}{2}}\frac{1-x_i t_i - x_j t_j + x_i x_j t_i t_j}{1-t_i-t_j+t_i t_j},$$

in the brackets, leads immediately to (2.16) and (2.17).

2.C High Noise Expansion

We indicate the main steps of the derivation of the full high noise expansion for

$$\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j}H(\underline{X}\mid\underline{Y})=\delta_{ij}S_1+(1-\delta_{ij})S_2.$$

The expansion for S_1 is given by (2.86) and that for S_2 by (2.91). They are derived in a form that is suitable to prove Lemma 2.4 of Section 2.6 (see Appendix 2.D). For this later proof we need to extract a square correlation at each order as in (2.91). This is achieved here through the use of appropriate remarkable Nishimori identities, and in order to use these we take the extrinsic forms (2.34) and (2.38) of S_1 and S_2.

Let us start with S_1 which is simple. Using the power series expansion of $\ln(1+x)$ we have

$$\ln\left(\frac{1+t_i\langle x_i\rangle_{\sim i}}{1+t_i}\right)=\sum_{p=1}^{+\infty}\frac{(-1)^{p+1}}{p}t_1^p(\langle x_i\rangle_{\sim i}^p-1).$$

This yields an infinite series for S_1 which we will now simplify. Because of the Nishimori identities

$$\mathbb{E}[t_i^{2p-1}]=\mathbb{E}[t_i^{2p}],\qquad \mathbb{E}_{\underline{t}^{\sim i}}[\langle x_i\rangle_{\sim i}^{2p-1}]=\mathbb{E}_{\underline{t}^{\sim i}}[\langle x_{\sim i}\rangle_1^{2p}],$$

we can combine odd and even terms and using (2.1) we get

$$S_1=\sum_{p=1}^{+\infty}\frac{m_2^{(2p)}}{2p(2p-1)}\big(\mathbb{E}_{\underline{t}^{\sim i}}[\langle x_i\rangle_{\sim i}^{2p}]-1\big).\tag{2.86}$$

This series is absolutely convergent as long as

$$\sum_{p=1}^{+\infty}\frac{m_2^{(2p)}}{2p(2p-1)}<+\infty,$$

which is true for channels satisfying H.

In the rest of the appendix we deal with S_2 which is considerably more complicated. However the general idea is the same as above. First we use the expansion of $\ln(1+x)$ to get

$$\ln\left(\frac{1+\langle x_i\rangle_{\sim ij}t_i+\langle x_j\rangle_{\sim ij}t_j+\langle x_ix_j\rangle_{\sim ij}t_it_j}{1+\langle x_i\rangle_{\sim ij}t_i+\langle x_j\rangle_{\sim ij}t_j+\langle x_i\rangle_{\sim ij}\langle x_j\rangle_{\sim ij}t_it_j}\right)=\mathrm{I}-\mathrm{II}-\mathrm{III},\tag{2.87}$$

where

$$\mathrm{I}=\sum_{p=1}^{\infty}\frac{(-1)^{p+1}}{p}\left(\langle x_i\rangle_{\sim ij}t_i+\langle x_j\rangle_{\sim ij}t_j+\langle x_ix_j\rangle_{\sim ij}t_it_j\right)^p,$$

$$\mathrm{II} = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} t_i^p \langle x_i \rangle_{\sim ij}^p, \qquad \mathrm{III} = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} t_j^p \langle x_j \rangle_{\sim ij}^p.$$

We expand the multinomial in I

$$\sum_{\substack{k_a,k_b,k_c \\ k_a+k_b+k_c=p}} \frac{p!}{k_a!k_b!k_c!} t_i^{k_a+k_c} t_j^{k_b+k_c} \langle x_i \rangle_{\sim ij}^{k_a} \langle x_j \rangle_{\sim ij}^{k_b} \langle x_i x_j \rangle_{\sim ij}^{k_c}, \tag{2.88}$$

and subtract the terms II and III. Then only terms that have powers of the form $t_i^k t_j^l$ with both $k, l \geq 1$ will survive in (2.87). Moreover because of the identities $\mathbb{E}[t_i^{2k-1}] = \mathbb{E}[t_i^{2k}]$ and $\mathbb{E}[t_j^{2l-1}] = \mathbb{E}[t_j^{2l}]$ we find for S_2

$$\begin{aligned} S_2 = &\sum_{k \geq l \geq 1}^{+\infty} m_1^{(2k)} m_1^{(2l)} \big(T_{00} + T_{01} + T_{10} + T_{11} \big) \\ &+ \sum_{l > k \geq 1}^{+\infty} m_1^{(2k)} m_1^{(2l)} \big(T'_{00} + T'_{01} + T'_{10} + T'_{11} \big), \end{aligned} \tag{2.89}$$

with (we abuse notation by not indicating the (kl) and (ij) dependence in the T and T' factors)

$$\begin{aligned} T_{\kappa\lambda} = \sum_{p=2k-\kappa}^{2k-\kappa+2l-\lambda} &\frac{(-1)^{p+1}}{p} \frac{p!}{(p-(2l-\lambda))!(p-(2k-\kappa))!(2k-\kappa+2l-\lambda-p)!} \\ &\times \mathbb{E}_{\underline{t}^{\sim ij}} \Big[\langle x_i \rangle_{\sim ij}^{p-(2l-\lambda)} \langle x_j \rangle_{\sim ij}^{p-(2k-\kappa)} \langle x_i x_j \rangle_{\sim ij}^{2k-\kappa+2l-\lambda-p} \Big], \end{aligned}$$

and

$$T'_{\kappa\lambda} = \text{exchange k, l and } \kappa, \lambda \text{ and i, j.}$$

Let us show the derivation of one of the terms T_{11}. This term corresponds to all the terms in the multinomial expansion (2.88) which contribute to the coefficient of $t_i^{2k-1} t_j^{2j-1}$. Clearly $p < 2k-1$ cannot contribute to $t_i^{2k-1} t_j^{2l-1}$. Thus we consider only $p \geq 2k-1$. We also have

$$k_a + k_c = 2k-1, \quad k_b + k_c = 2l-1, \quad k_a + k_b + k_c = p.$$

Solving these constraints, we get $k_a = p - 2l + 1$, $k_b = p - 2k + 1$ and $k_c = 2k + 2l - 2 - p$ and since we must have $k_c \geq 0$, we get $p \leq 2k + 2l - 2$. Thus the term T_{11} equals

$$\begin{aligned} \sum_{p=2k-1}^{2k+2l-2} &\frac{(-1)^{p+1}}{p} \frac{p!}{(p-(2l-1))!(p-(2k-1))!(2k+2l-2-p)!} \\ &\times \mathbb{E}_{\underline{t}^{\sim ij}} \Big[\langle x_i \rangle_{\sim ij}^{p-(2l-1)} \langle x_j \rangle_{\sim ij}^{p-(2k-1)} \langle x_i x_j \rangle_{\sim ij}^{2k+2l-2-p} \Big]. \end{aligned}$$

We obtain the other terms in a similar manner. Note that the T' terms correspond to $k < l$ are got simply by exchanging all the labels in T terms.

The next simplification step occurs by using the Nishimori identity for the expectation in the above formula

$$\mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i\rangle_{\sim ij}^{m_1}\langle x_j\rangle_{\sim ij}^{m_2}\langle x_ix_j\rangle_{\sim ij}^{m_3}\right] = \mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i^{m_1}x_j^{m_2}(x_ix_j)^{m_3}\rangle_{\sim ij}\langle x_i\rangle_{\sim}^{m_1}\langle x_j\rangle_{\sim ij}^{m_2}\langle x_ix_j\rangle_{\sim ij}^{m_3}\right],$$

and using $x_i \in \{\pm 1\}$, to "linearize" the terms $(x_ix_j)^{m_1}x_i^{m_2}x_j^{m_3}$. As κ, λ vary in $\{0,1\}$, p varies from $2k-1$ to $2k+2l$. For each p, we will simplify the contributions from all $T_{\kappa\lambda}$ using the Nishimori identities. For $p = 2k-1$, $T_{10}+T_{11}$ contributes to (T_{01}, T_{00} do not contribute)

$$\frac{1}{2k-1}\frac{(2k-1)!}{(2k-2l-1)!(2l)!}\mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i\rangle_{\sim ij}^{2k-2l-1}\langle x_ix_j\rangle_{\sim ij}^{2l}\right] + \frac{1}{2k-1}\frac{(2k-1)!}{(2k-2l)!(2l-1)!}\mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i\rangle_{\sim ij}^{2k-2l}\langle x_ix_j\rangle_{\sim ij}^{2l-1}\right].$$

Now applying the Nishimori identities we get the that the contribution to $p = 2k-1$ is

$$\Big(\frac{(2k)!}{(2k-1)((2k-2l)!)(2l)!}\Big)\mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i\rangle_{\sim ij}^{2k-2l}\langle x_ix_j\rangle_{\sim ij}^{2l}\right].$$

Now consider any $2k-1 < p \le 2k+2l-2$. Such p has contribution from all the terms, $T_{00}, T_{01}, T_{10}, T_{11}$. Thus applying Nishimori identities we get the contribution to be

$$\sum_{\kappa,\gamma}\frac{(-1)^{p+1}}{p}\frac{p!}{(p-(2l-\lambda))!(p-(2k-\kappa))!(2k-\kappa+2l-\lambda-p)!} \times \mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i^{2k-\kappa}x_j^{2l-\lambda}\rangle_{\sim ij}\langle x_i\rangle_{\sim ij}^{p-(2l-\lambda)}\langle x_j\rangle_{\sim ij}^{p-(2k-\kappa)}\langle x_ix_j\rangle_{\sim ij}^{2k-\kappa+2l-\lambda-p}\right].$$

Notice that for $(\kappa,\lambda) = (0,1),(1,0),(1,1)$ we have

$$\mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i^{2k-\kappa}x_j^{2l-\lambda}\rangle_{\sim ij}\langle x_i\rangle_{\sim ij}^{p-(2l-\lambda)}\langle x_j\rangle_{\sim ij}^{p-(2k-\kappa)}\langle x_ix_j\rangle_{\sim ij}^{2k-\kappa+2l-\lambda-p}\right] = \mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i\rangle_{\sim ij}^{p-2l+1}\langle x_j\rangle_{\sim ij}^{p-2k+1}\langle x_ix_j\rangle_{\sim ij}^{2k+2l-1-p}\right].$$

The combinatorial terms for $(\kappa,\lambda) = (0,1),(1,0),(1,1)$ add upto $(p+1)!/\Big((p-2l+1)!(p-2k+1)!(2k+2l-1-p)!\Big)$. Thus the final contribution for any $2k-1 < p \le 2k+2l-2$ is,

$$\frac{(-1)^{p+1}}{p}\frac{(p+1)!}{(p-2l+1))!(p-2k+1))!(2k+2l-1-p)!}\mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i\rangle_{\sim ij}^{p-2l+1}\langle x_j\rangle_{\sim ij}^{p-2k+1}\langle x_ix_j\rangle_{\sim ij}^{2k+2l-1-p}\right] + \frac{(-1)^{p+1}}{p}\frac{p!}{(p-2l))!(p-2k))!(2k+2l-p)!}\mathbb{E}_{\underline{t}^{\sim ij}}\left[\langle x_i\rangle_{\sim ij}^{p-2l}\langle x_j\rangle_{\sim ij}^{p-2k}\langle x_ix_j\rangle_{\sim ij}^{2k+2l-p}\right].$$

The last term in the above sum corresponds to $(\kappa, \lambda) = (0, 0)$. In above, forgetting the factor $(-1)^{p+1}/p$, the first term is got from the second by replacing p with $p+1$. Hence for two consecutive values of say $p', p'+1$ we can add up the $p'+1$ contribution of p' and $p'+1$ contribution of $p'+1$ to finally get

$$\sum_{\kappa,\lambda} T_{\kappa,\lambda} = \sum_{p=2k-1}^{2k+2l-1} \frac{(-1)^{p+1}}{p(p+1)} \frac{(p+1)!}{(p+1-2k)!(p+1-2l)!(2k+2l-p-1)!} \times \mathbb{E}_{\underline{t}\sim ij}\left[\langle x_i \rangle_{\sim ij}^{p-2l+1} \langle x_j \rangle_{\sim ij}^{p-2k+1} (\langle x_i x_j \rangle_{\sim ij}^{2k+2l-1-p} \right].$$

We have skipped the derivation of the contribution for $p = 2k + 2l - 1$ and $p = 2k + 2l$ from T_{01}, T_{10}, T_{00}, because it is done on very similar lines as above. A similar formula obtained by exchanging k, l and i, j holds for $\sum_{\kappa\lambda} T'_{\kappa\lambda}$. Replacing these sums in (2.89) yields a high noise expansion for S_2.

However this is not yet practical for us because we need to extract a general square correlation factor $\big(\langle x_i x_j \rangle_{\sim ij} - \langle x_i \rangle_{\sim ij} \langle x_j \rangle_{\sim ij}\big)^2$. The fact that this is possible is a "miracle" that comes out of the Nishimori identities that were used. Setting

$$X = \langle x_i \rangle_{\sim ij} \langle x_j \rangle_{\sim ij}, \qquad Y = \langle x_i x_j \rangle_{\sim ij},$$

and using the change of variables $m = p - 2k + 1$, the last expression becomes (for $k \geq l$)

$$\mathbb{E}_{\underline{t}\sim ij}\left[\frac{\langle x_i \rangle_{\sim ij}^{2k-2l}}{(2l)!} \sum_{m=0}^{2l} (-1)^m \binom{2l}{m} X^m Y^{2l-m} (m+2k-2) \cdots (m+2k-(2l-1)) \right].$$

One can check that this is equal to

$$\frac{\langle x_i \rangle_{\sim ij}^{2k-2l}}{(2l)!} X^{2l-2k} \frac{\partial^{2l-2}}{\partial X^{2l-2}} \left(X^{2k-2} (X-Y)^{2l} \right),$$

by expanding $(X-Y)^{2l}$ first and then differentiating. On the other hand one can use the Leibnitz product rule

$$\frac{\partial^{2l-2}}{\partial X^{2l-2}} \left(X^{2k-2} (X-Y)^{2l} \right) = \sum_{r=0}^{2l-2} \binom{2l-2}{r} \frac{\partial^{2l-2-r}}{\partial X^{2l-2-r}} X^{2k-2} \frac{\partial^r}{\partial X^r} (X-Y)^{2l},$$

to find that the last expectation above is equal to

$$\mathbb{E}_{\underline{t}\sim ij}\left[(X-Y)^2 \langle x_i \rangle_{\sim ij}^{2k-2l} \sum_{r=0}^{2l-2} A_{rlk} X^r (X-Y)^{2l-r-2} \right],$$

where

$$A_{rlk} = \frac{1}{(2l)!} \tbinom{2l-2}{r} [2l]_r [2k-2]_{2l-2-r}, \qquad [m]_r = m(m-1)\cdots(m-r+1).$$

We define $A_{011} = \frac{1}{2}$. We proceed similarly for the terms with $k < l$. Finally one finds

$$S_2 = \sum_{k \geq l \geq 1} m_1^{(2k)} m_1^{(2l)} \mathbb{E}_{\underline{t}^{\sim ij}} \Bigg[\Bigg(\langle x_i x_j \rangle_{\sim ij} - \langle x_i \rangle_{\sim ij} \langle x_i \rangle_{\sim ij} \Bigg)^2 \langle x_i \rangle_{\sim ij}^{2k-2l} \times \sum_{r=0}^{2l-2} A_{rlk} \langle x_i \rangle_{\sim ij}^r \langle x_j \rangle_{\sim ij}^r \Bigg(\langle x_i \rangle_{\sim ij} \langle x_j \rangle_{\sim ij} - \langle x_i x_j \rangle_{\sim ij} \Bigg)^{2l-2-r} \Bigg] + \sum_{l > k \geq 1} \text{identical expression with k, l and i, j exchanged.} \tag{2.90}$$

Let us now briefly justify that the series is absolutely convergent for channels satisfying H. We note the following facts: the absolute value of the spin averages is less than 1 and the value of the correlation in absolute terms is less than 2, $A_{rlk} \leq \binom{2l-2}{r} 2^{2k-3}$ and $2^{2k-2} 3^{2l-2} \leq (\frac{5}{2})^{2k+2l-4}$ for $k \geq l$ (together with the version with k, l exchanged). It easily follows that

$$|S_2| \leq \frac{8}{625} \mathbb{E}_{\underline{t}^{\sim ij}} \Bigg[\Bigg(\langle x_i x_j \rangle_{\sim ij} - \langle x_i \rangle_{\sim ij} \langle x_i \rangle_{\sim ij} \Bigg)^2 \Bigg] \sum_{k,l \geq 1} (\frac{5}{2})^{2k+2l} |m_1^{(2k)} m_1^{(2l)}|. \tag{2.91}$$

Thus the series for S_2 is absolutely convergent as long as

$$\sum_{p=1}^{+\infty} (\frac{5}{2})^{2p} |m_1^{(2p)}| < +\infty.$$

Note that we have not attempted to optimize the above estimates.

2.D Proof of Lemma 2.4

We prove Lemma 2.4 for highly noisy general BMS channels. For this we use the high noise expansion derived in Appendix 2.C. There it was derived for a general linear code ensemble, and this is also the framework of the proof below. Of course the result applies to the interpolating ensemble of Lemma 2.4 because the interpolating random variable U_a does not depend apriori on the noise value ϵ and depends only on the distribution of independent random variable V. Note that the the final constants $F(\epsilon)$ and $G(\epsilon)$ do not depend on the code ensemble but only on the channel.

Consider equation (2.7) for $\frac{d^2}{d\epsilon^2} \mathbb{E}_{\mathcal{C},\underline{t}}[h_n]$. By the same estimates than those for S_1 in Appendix 2.C, the first term on the right hand side is certainly greater than

$$-\sum_{p=1}^{+\infty} \frac{|m_2^{(2p)}|}{2p(2p-1)} = -A.$$

To get a lower bound for the second term we consider the series expansion given by that for S_2 in (2.91). In that series we keep the first term corresponding to $k = l = 1$, namely

$$\frac{1}{2}(m_1^{(2)})^2 \sum_{j\neq 1} \mathbb{E}_{\mathcal{C},\underline{t}^{\sim 1j}} \left[\left(\langle x_1 x_j \rangle_{\sim 1j} - \langle x_1 \rangle_{\sim 1j} \langle x_j \rangle_{\sim 1j} \right)^2 \right] = B,$$

and lower bound the rest of the series $(k, l) \neq (1, 1)$ by using estimates of Appendix 2.C. More precisely, this part is lower bounded by

$$-\left(\frac{8}{625} \left(\sum_{p=1}^{+\infty} \left(\frac{5}{2}\right)^{2p} |m_1^{(2p)}| \right)^2 - \frac{1}{2}(m_1^{(2)})^2 \right) \times \sum_{j\neq 1} \mathbb{E}_{\mathcal{C},\underline{t}^{\sim 1j}} \left[\left(\langle x_1 x_j \rangle_{\sim 1j} - \langle x_1 \rangle_{\sim 1j} \langle x_j \rangle_{\sim 1j} \right)^2 \right] = -C.$$

Putting these three estimates together we get

$$\frac{d^2}{d\epsilon^2} \mathbb{E}_{\mathcal{C},\underline{t}}[h_n] \geq -A + B - C. \tag{2.92}$$

As long as the noise level is high enough so that (see H)

$$\sum_{p=2}^{+\infty} \left(\frac{5}{2}\right)^{2p} |m_1^{(2p)}| < (\sqrt{2} - 1) \left(\frac{5}{2}\right)^2 |m_1^{(2)}|,$$

the inequality (2.92) implies

$$\sum_{j\neq 1} \mathbb{E}_{\mathcal{C},\underline{t}^{\sim 1j}} \left[\left(\langle x_1 x_j \rangle_{\sim 1j} - \langle x_1 \rangle_{\sim 1j} \langle x_j \rangle_{\sim 1j} \right)^2 \right] \leq \tilde{F}(\epsilon) + \tilde{G}(\epsilon) \frac{d^2}{d\epsilon^2} \mathbb{E}_{\mathcal{C},\underline{t}}[h_n], \tag{2.93}$$

for two noise dependent positive finite constants $\tilde{F}(\epsilon)$, $\tilde{G}(\epsilon)$.

The final step of the proof consists in passing from the extrinsic average $\langle - \rangle_{\sim 1j}$ in the correlation to the ordinary one $\langle - \rangle_{1j}$. This is achieved as follows. From the formulas (2.16) and (2.17) we deduce that

$$\langle x_j x_i \rangle - \langle x_j \rangle \langle x_i \rangle = \left(\langle x_j x_i \rangle_{\sim ij} - \langle x_j \rangle_{\sim ij} \langle x_i \rangle_{\sim ij} \right) R_{ij},$$

with

$$R_{ij} = \frac{\left(1 - \langle x_i \rangle t_i - \langle x_j \rangle t_j + \langle x_i x_j \rangle t_i t_j\right)^2}{(1 - t_i^2)(1 - t_j^2)} \leq \frac{4}{(1 - t_i^2)(1 - t_j^2)},$$

a function that depends on all log-likelihood variables.

Thus we have

$$\begin{aligned} (\langle x_j x_i \rangle - \langle x_j \rangle \langle x_i \rangle)^2 &= \left(\langle x_j x_i \rangle_{\sim ij} - \langle x_j \rangle_{\sim ij} \langle x_i \rangle_{\sim ij} \right)^2 R_{ij}^2 \\ &\leq \left(\langle x_j x_i \rangle_{\sim ij} - \langle x_j \rangle_{\sim ij} \langle x_i \rangle_{\sim ij} \right)^2 \frac{16}{(1 - t_i^2)^2 (1 - t_j^2)^2} \end{aligned}$$

Taking now the expectation $\mathbb{E}_{\mathcal{C},\underline{t}}$ we get

$$\mathbb{E}_{\mathcal{C},\underline{t}}\left[\left(\langle x_j x_i\rangle - \langle x_j\rangle\langle x_i\rangle\right)^2\right] \leq \mathbb{E}_{\mathcal{C},\underline{t}^{\sim ij}}\left[\left(\langle x_j x_i\rangle_{\sim ij} - \langle x_j\rangle_{\sim ij}\langle x_i\rangle_{\sim ij}\right)^2\right] \times \mathbb{E}_{t_i,t_j}\left[\frac{16}{(1-t_i^2)^2(1-t_j^2)^2}\right].$$

Since t_i, t_j are independent we get

$$\mathbb{E}_{t_i,t_j}\left[\frac{16}{(1-t_i^2)^2(1-t_j^2)^2}\right] = 16\left(\mathbb{E}\left[\frac{1}{(1-t^2)^2}\right]\right)^2 = 16\left(\mathbb{E}\left[\sum_{p\geq 0}(p+1)t^{2p}\right]\right)^2 = 16\left(\left[\sum_{p\geq 0}(p+1)m_0^{(2p)}\right]\right)^2, \tag{2.94}$$

which converges for highly noisy channels satisfying H. The result of the lemma follows by combining (2.93) and (2.94). The constants $F(\epsilon)$ and $G(\epsilon)$ are equal to $\tilde{F}(\epsilon)$ and $\tilde{G}(\epsilon)$ divided by the expression on the right hand side of the last inequality.

2.E Boundedness of GEXIT

We prove the boundedness and positivity of $\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)]$ which is needed in the proof of Lemma 2.2.

Lemma 2.6. *For the BEC and BIAWGNC with any noise level, and any BMS satisfying H, there exists a constant $k(\epsilon)$ independent of n, γ, t_* and s such that*

$$0 \leq \frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)] \leq k(\epsilon). \tag{2.95}$$

For the BEC we can take $k(\epsilon) = \frac{\ln 2}{\epsilon}$ and for the BIAWGNC $k(\epsilon) = \frac{2}{\epsilon^{-3}}$. For general BMS channels satisfying H the constant remains bounded as a function of ϵ (i.e. in the high noise regime).

Here we have stated the lemma for the Multi-Poisson interpolating ensemble which is our specific need. However as the proof below shows it is independent of the specific code ensemble and the bound depends only on the channel.

Proof. We will use the GEXIT formulas of Section 2.3. Since the lemma applies for any linear code it also applies for the interpolating ensemble of interest here. In the case of the BEC and BIAWGNC we have (see (2.28), (2.31)

$$\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)] = \frac{\ln 2}{\epsilon}(1 - \mathbb{E}_s[\langle x_1\rangle_s]),$$

and

$$\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*,s)] = \frac{2}{\epsilon^3}(1-\mathbb{E}_s[\langle x_1\rangle_s]).$$

The bounds of the lemma follow immediately since $0 \leq \mathbb{E}_s[\langle x_1\rangle_s \leq 1$ (Nishimori identity implies $\mathbb{E}_s[\langle x_1\rangle_s] = \mathbb{E}_s[\langle x_1\rangle_s^2]$).

For highly noisy BMS channels we proceed by expansions. For this reason we have to use the "extrinsic form" of the GEXIT formula (analogous to (2.34))

$$\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*,s)] = \int_{-1}^{+1} dt_1 \frac{\partial c_D(t_1)}{\partial \epsilon}\mathbb{E}_{s,\sim t_1}\left[\ln\left(\frac{1+t_1\langle x_1\rangle_{s,\sim 1}}{1+t_1}\right)\right].$$

Expanding the logarithm and using Nishimori identities (as in the expansion of S_1 in Appendix 2.B we obtain

$$\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*,s)] = \sum_{p=1}^{\infty}\frac{m_1^{(2p)}}{2p(2p-1)}\mathbb{E}_{s,\sim 1}[\langle x_1\rangle_{s,\sim 1}^{2p}-1].$$

The positivity follows from $m_1^{(2p)} \leq 0$ (see [23], Section VI in [91]) and $-1 \leq x_1 \leq 1$. The upper bound (and absolute convergence) follow from condition H. In particular we get

$$k(\epsilon) = \sum_{k=1}^{+\infty}\frac{|m_1^{(2p)}|}{2p(2p-1)},$$

which is independent of n, γ, t_* and s. □

Binary Erasure Channel: Second Interpolation Method

3

3.1 Introduction

In the previous chapter we showed, using the interpolation method, that the replica solution is a lower bound on the average per-bit conditional entropy $\mathbb{E}_{\mathcal{C}}[h_n]$. Let us call this application of the interpolation method the *first interpolation method*.

The natural next step is to prove the equality of the replica solution by showing a matching upper bound. For the case of transmission over the BEC, this equality has been rigorously established for a class of LDPC code ensembles, which includes regular LDPC code ensembles ($\Lambda(x) = x^{\mathtt{l}}$, $P(x) = x^{\mathtt{r}}$) [23], [24]. There, the methods used are combinatorial and, as such, do not extend to the case of transmission over general BMS channels.

In this chapter we consider transmission over the BEC using the *Poisson-LDPC (LDGM)* code ensemble (c.f. Section 2.5). We use the *second interpolation method*, developed by Guerra and Toninelli [31], [30] in the context of the SK model [71], [87] to prove that the replica solution is correct.

In the case of the SK model, the first interpolation shows that the replica solution[1] is a lower bound on the free energy. In [31], Guerra and Toninelli use the second interpolation to demonstrate that the replica solution is also an *upper bound* on the free energy. They are able to show this in the high-temperature regime. [2]

[1] More precisely, we refer here to the *replica symmetric solution*.

[2] In fact, for the SK model, the simple replica symmetric solution (as is the case for the coding problem) is not correct for all temperatures. A long standing open question was whether the Parisi formula, which corresponds to *full replica symmetric breaking solution*, is the correct solution [71]. Recently, Talagrand [75] used the second interpolation method as a crucial ingredient to prove the celebrated Parisi formula for the SK model.

In the case of LDPC ensembles the second interpolation has not yet been developed. This is in essence what we do in this chapter for the simplest case of Poisson-LDPC ensembles and the BEC. Part of the mathematical analysis involved in the second interpolation is reminiscent of the one developed recently for the (simpler) case of a "gauge symmetric p-spin model" [96].

We believe it should be possible to extend these results to any LDPC code ensemble using the multi-Poisson ensemble approximation of Montanari [29]. Currently, we are unable to extend our proof to case of transmission over general channels. However, since our techniques are non-combinatorial in nature, we hope that it will be possible to extend our proof to the general case.

For the ease of exposition we restrict ourselves to the Poisson-LDPC code ensemble.

3.2 Organization of the Chapter

In the next section we state our main result concerning the equality of the replica solution for the case of transmission over the BEC using Poisson-LDPC code ensemble. We also describe our proof strategy there. In Section 3.4 we simplify the statistical physics formulation to the case of transmission over the BEC. We then show that the remainder term in the first interpolation simplifies to the evaluation of the "magnetization". Before estimating the remainder, we discuss the BP decoder for the "interpolated codes". In Section 3.5, we apply the second interpolation method of Guerra and Toninelli to provide an upper bound on the remainder of the first interpolation. As a consequence, we will infer the exactness of the replica solution. We conclude the chapter in Section 3.6 by discussing the case of transmission over general channels and the relevance of the magnetization.

3.3 Main Result and Strategy of Proof

Consider communication through a BEC with transition probability $p_{Y|X}(y|x)$ and erasure probability given by ϵ. We briefly recall the Poisson-LDPC code ensemble here, see Section 2.5 for details. The design rate of the code ensemble is equal to $\mathtt{R}$. The number of check nodes is a Poisson number with mean $n(1-\mathtt{R})$. The check node degree distribution is given by $P(x)$. Each check node of degree k is attached to k variable nodes u.a.r. As the block-length goes to infinity, one can show that the variable node degree distribution converges to a Poisson distribution. More precisely, the variable node degree distribution is given by,

$$\Lambda(x) = e^{\gamma(x-1)}.$$

For ease of notation we use $\gamma = P'(1)(1-\mathtt{R})$ to denote the average left degree.

For Poisson-LDPC code ensemble it can be shown that

$$\lambda(x) = \Lambda(x).$$

In words, the edge and variable node degree distribution are the same. In this chapter, we will denote the Poisson-LDPC ensemble by LDPC(n, γ, P).

Recall that the replica solution consists of optimizing the function $h_{RS}[d_V; \gamma, P]$ over the space of symmetric probability distributions $d_V(v)$. For the case of transmission over the BEC, code-bits are either known perfectly ($\mathsf{LLR} = +\infty$) or are completely erased ($\mathsf{LLR} = 0$). Thus, without loss of generality, we assume that the $d_V(v) = p\delta_0(v) + (1-p)\delta_\infty(v)$ with $0 \leq p \leq 1$. With this, we can simplify the replica solution functional $h_{RS}[d_V; \gamma, P]$ to

$$h_{RS}[p] = \ln 2\Big(\Lambda'(1)p\rho(1-p) + \epsilon\Lambda(1-\rho(1-p)) - \frac{\Lambda'(1)}{P'(1)}(1-P(1-p))\Big). \tag{3.1}$$

Since the replica functional is now only a function of one parameter p, we denote it by $h_{RS}[p]$. Also, the variational problem (optimization of $h_{RS}[p]$) can now be explicitly solved. Note that the replica solution given in, (3.1), was called *trial entropy* in [24], [23].

Differentiating $h_{RS}[p]$ with respect to p (for a fixed ϵ) we get

$$\begin{aligned}
\frac{dh_{RS}[p]}{dp} &= \ln 2\Bigg(\Lambda'(1)\rho(1-p) - \Lambda'(1)p\rho'(1-p) + \epsilon\Lambda'(1-\rho(1-p))\rho'(1-p) \\
&\qquad - \Lambda'(1)\frac{P'(1-p)}{P'(1)}\Bigg) \\
&= \ln 2\Big(\Lambda'(1)\rho'(1-p)\Big)\Big(-p + \epsilon\frac{\Lambda'(1-\rho(1-p))}{\Lambda'(1)}\Big) \\
&= \ln 2\Big(\Lambda'(1)\rho'(1-p)\Big)\Big(-p + \epsilon\lambda(1-\rho(1-p))\Big).
\end{aligned} \tag{3.2}$$

To get the second equality we used $\rho(x) = P'(x)/P'(1)$ to cancel out $\Lambda'(1)\rho(1-p)$ and $\Lambda'(1)\frac{P'(1-p)}{P'(1)}$. We used the relation $\lambda(x) = \Lambda'(x)/\Lambda'(1)$ to obtain the last equality.

Thus, from (3.2) we deduce that the maximizer of $h_{RS}[p]$ is a solution of the stationary point equation

$$p = \epsilon\lambda(1-\rho(1-p)). \tag{3.3}$$

We denote the unique maximizer of $h_{RS}[p]$ by p_{RS}. We remark that p_{RS} is a function of ϵ, but we suppress this dependence to lighten the notation.

The density evolution analysis of the BP decoder leads to the same stationary point equation (3.3). The equation is solved iteratively starting from the initial condition $p = \epsilon$ (channel observation) to obtain the BP decoder

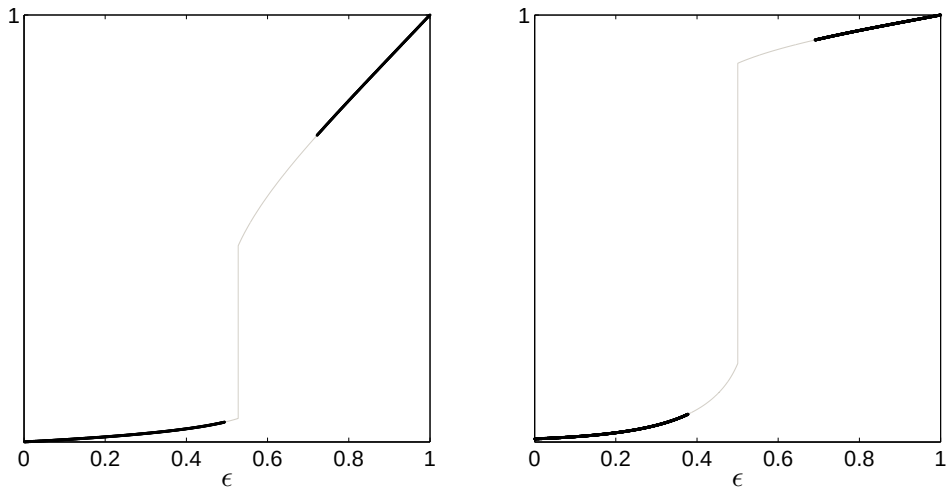

Figure 3.1: In this figure we demonstrate our result. We plot p_{RS} as a function of ϵ for two class of code ensembles. We then mark out in dark, the regime of noise where our result holds. Left figure: The thin line is the plot of p_{RS} as a function of ϵ (obtained from the replica solution) for Poisson-LDPC code with $\gamma = 3, P(x) = x^6$. The dark line corresponds to the regime of channel erasure fractions, where we prove the exactness of the replica solution. Right figure: The thin line is plot of p_{RS} as a function of ϵ obtained from the replica solution for Poisson-LDGM code with $\gamma = 5, P(x) = 0.1x + 0.9x^5$. The dark line is the region of channel erasure fraction where we prove the exactness of the replica solution.

estimate. The average (over the code ensemble) erasure probability of BP decoder is the largest fixed-point obtained from this initial condition and we will call it p_{BP}. Our main result states that, asymptotically, in appropriate noise regimes, p_{BP} gives the average error probability of the MAP decoder.

In order to state our theorem below we define an auxiliary function

$$\begin{aligned} f(z) =& \frac{\gamma}{P'(1)}\Big(P(z) - zP'(z)\Big) \\ &+ (1-p_{RS})\frac{\gamma}{P'(1)}\Big(P'(z) - P'(1-p_{RS})\Big) \\ &+ p_{RS}\ln\cosh\Big[\frac{\gamma}{P'(1)}\Big(P'(z) - P'(1-p_{RS})\Big)\Big]. \end{aligned} \tag{3.4}$$

Note that the stationary points of f are the solutions of

$$z = p_{RS}\tanh\Big(\frac{\gamma}{P'(1)}(P'(z) - P'(\bar{p}_{RS}))\Big) + \bar{p}_{RS},$$

where we use the notation $\bar{p}_{RS}$ to denote $1 - p_{RS}$. It is clear that at $z = \bar{p}_{RS}$, the stationary point equation is satisfied. Hence $1 - p_{RS}$ is a critical point.

Theorem 3.1 (Exactness of Replica Solution for BEC)**.** *Assume communication using a Poisson-LDPC(n, γ, P) code ensemble, through a BEC with erasure probability ϵ. Assume that the function f, defined in* (3.4)*, has a unique maximizer $\hat{z}$. Recall that p_{BP} denotes the fixed-point of density evolution equation* (3.3) *obtained via BP. Then for all ϵ, such that (i) $p_{RS} = p_{BP}$ and (ii) $\epsilon \in C_{P,\gamma} = \{\epsilon \in [0,1] \mid \hat{z} = 1 - p_{RS}\}$ we have*

$$\lim_{n \to +\infty} \mathbb{E}_{\mathcal{C}}[h_n] = h_{RS}[p_{RS}].$$

Example 3.1 (Replica Solution for Poisson-LDPC$(n, \gamma = 3, P(x) = x^6)$)**.** *Consider transmission over the BEC using Poisson-LDPC code ensemble LDPC(n, γ, P) with $\gamma = 3$ and $P(x) = x^6$. We illustrate our main theorem in the Figure 3.3. In the figure, the thin line corresponds to p_{RS} as a function of the channel erasure probability ϵ. The dark or thick line corresponds to the regime of noise where we prove that the replica solution is correct.*

For the case of Poisson-LDPC codes we have

$$\begin{aligned} \frac{dh_{RS}[p_{RS}]}{d\epsilon} &= \Lambda(1 - \rho(1 - p_{RS})) \\ &= \lambda(1 - \rho(1 - p_{RS})) \\ &= \frac{p_{RS}}{\epsilon}. \end{aligned}$$

To get the second inequality we use the fact that for a Poisson-LDPC code ensemble we have $\Lambda(x) = \lambda(x)$. As a consequence of the above calculations, our result also implies that, for the channel erasure fractions depicted by the thick line in Figure 3.3,

$$\lim_{n \to +\infty} \frac{d\mathbb{E}_{\mathcal{C}}[h_n]}{d\epsilon} = \frac{p_{RS}}{\epsilon}.$$

I.e. the asymptotic average MAP-GEXIT function is given by p_{RS}/ϵ. From the various characterizations of the MAP-GEXIT curve for the case of transmission over the BEC, as shown in [23], we have that the MAP error probability is given by $\frac{\epsilon}{2}\frac{dh_n}{d\epsilon}$. Thus, a consequence of the theorem is that (at least over the range of ϵ depicted by the dark region) the MAP error probability is given by p_{RS} (which is computable from the replica formulas).

Remark 3.1. *The condition over the range of ϵ is not optimal and comes from the second interpolation that we use. In [24] there is also a condition which is different from ours. In [24], Measson et al. show that the asymptotic average MAP-GEXIT and BP-GEXIT curve are equal under the condition that the rate of the residual graph (graph obtained after the termination of BP decoding) sharply concentrates on the design rate of the residual graph ensemble. In fact, the trial entropy or the replica solution corresponds to the uncertainty in determining the transmitted codeword at the end of BP decoding. Thus, it is*

equal to the design rate of the residual graph ensemble (see proof of Theorem 10 in [24]). Both, our result and the result in [24] hold only when $p_{RS} = p_{BP}$. It is an open question, how to prove the exactness of the replica solution when $p_{RS} \neq p_{BP}$.

3.3.1 Strategy of Proof

We know from the previous chapter that the interpolation method allows us to write the per-bit conditional entropy as a sum of the replica solution and a remainder term. We have already proved in the previous chapter that the remainder (c.f. (2.44)) is non-negative in the limit of large block-length. A way to prove the equality of the replica solution would be to show that the remainder term is exactly zero.

To do this we use the second interpolation method of Guerra and Toninelli. In the case of SK model, the remainder term of the first interpolation method consists of only one term, Q_2 (c.f. (2.45)). The second interpolation method of Guerra and Toninelli is the following. They first "couple" the quadratic term, Q_2, with the partition function of the original system, to create a new partition function. Then using the interpolation method on this new partition function they prove that the remainder term is less than or equal to zero. Coupled with the non-negativity of the remainder term, they conclude that it is exactly zero in the limit $n \to +\infty$. This immediately provides the exactness of the replica solution. We will provide in Section 3.5.1 an intuitive explanation of the second interpolation.

In the case of LDPC spin-glass, the remainder term is much more complicated. More precisely, the remainder is a sum of infinite terms, each containing an overlap parameter (see (2.44) in Section 2.5). As a result, it would be hard to appropriately "couple the remainder" with the partition function, as done by Guerra and Toninelli. Fortunately, as we will see in the next section, for the special case of transmission over the BEC, the remainder simplifies drastically. Thus, we are able to proceed along such methods.

3.4 Statistical Mechanical Formulation and First Interpolation

In this section we recall the statistical physics formulation. We also specialize the formulation to the case of the BEC. We then apply the interpolation method to compute the per-bit conditional entropy. We will then see that the result of the interpolation method simplifies for the case of the BEC.

As usual the Tanner graph has variable nodes, denoted by i that are connected to check nodes denoted by a. We will work in terms of the LLR l (see Definition 1.3) and their distribution $c_L(l)$ under the assumption of the all-one codeword transmission.

In the case of the BEC, the LLR value can be $+\infty$ with positive probability. As a result, to have a well defined Gibbs (MAP) measure (see Section 1.7), $p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y})$, we re-write it as,

$$p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{1}{Z}\prod_a \frac{1}{2}(1+x_{\partial a})\prod_{i=1}^n \frac{e^{\frac{l_i}{2}x_i}}{2\cosh\frac{l_i}{2}} \triangleq \mu(\underline{x}).$$

Recall that $x_{\partial c} = \prod_{i\in c} x_i$ and Z is the normalization factor or partition function given by

$$Z \triangleq \sum_{\underline{x}}\prod_a \frac{1}{2}(1+x_{\partial a})\prod_{i=1}^n \frac{e^{\frac{l_i}{2}x_i}}{2\cosh\frac{l_i}{2}}.$$

Also, recall that the expectations with respect to the Gibbs measure for a fixed graph and a fixed channel output are denoted by the bracket $\langle -\rangle$. More precisely for any $A \subset \{1,...,n\}$, $\langle x_A\rangle = \sum_{\underline{x}} x_A\mu(\underline{x})$ where $x_A = \prod_{i\in A} x_i$. As usual, expectations with respect to the code ensemble and the channel outputs will be denoted by $\mathbb{E}_{\mathcal{C},\underline{l}}[-]$. Note that for the BEC, we have a positive mass at $+\infty$ in the distribution $c_L(l)$. Note that this re-definition does not affect the Gibbs measure, $\mu(\underline{x})$. Since the Gibbs distribution is unchanged, all the averages also remain unaffected.

With this re-formulation, a little algebra, on the same lines as (1.24), gives the following formula for the per-bit conditional entropy

$$\mathbb{E}_{\mathcal{C}}[h_n] = \frac{1}{n}\mathbb{E}_{\mathcal{C},\underline{l}}[\ln Z] + \epsilon\ln 2.$$

Since the free energy, $n^{-1}\mathbb{E}_{\mathcal{C},\underline{l}}[\ln Z]$, differs from the average conditional entropy only by a constant, we will focus on the evaluation of the free energy.

3.4.1 First Interpolation

We briefly recall the interpolation method here (see Section 2.5 and Appendix 2.A.4 for details). Recall that the main idea behind the interpolation technique is to recursively remove the check node constraints from the code. Then to compensate the increase in rate, extra "observations", given by $U_a = 2\tanh^{-1}\left[\prod_{i=1}^{k-1}\tanh\frac{V_i}{2}\right]$, are added to each variable node i. Note that here k is an random number whose distribution is given by $P(x)$. Also, recall that V is a random variable with a symmetric density $d_V(v)$. In the case of BEC we have $d_V(v) = p\delta_0(v) + (1-p)\delta_\infty(v)$. Here, $0 \le p \le 1$ is an arbitrary parameter. Notice that the equation for U_a mimics a check node message in the belief propagation decoding algorithm.

Let $t \in [0,1]$ be an interpolating parameter and consider the Tanner graphs C_t from the ensemble LDPC$(n, \gamma t, P)$. Thus, at "time" t, the number of check nodes is a Poisson r.v. with mean $n\gamma t$. As said before, the loss of check nodes is compensated by "extra observations". Variable nodes i receive e_i i.i.d. copies

$\{U_a^i\}$, $a = 1, ..., e_i$ of the r.v. U_a. Here e_i are i.i.d. copies of a random Poisson number with mean $n\gamma(1-t)$. The interpolating Gibbs measure is then given by,

$$\mu_t(\underline{x}) = \frac{1}{Z(t)} \prod_{a \in C_t} \frac{1}{2}(1 + x_{\partial a}) \prod_{i=1}^{n} \frac{e^{(\frac{l_i}{2} + \sum_{a=1}^{e_i} \frac{u_a^i}{2})x_i}}{2\cosh\frac{l_i}{2} \prod_{a=1}^{e_i} 2\cosh\frac{u_a^i}{2}}$$

where C_t denotes the code at "time" t. These will be also called as "interpolated codes" in the sequel. Expectations with respect to this measure will be denoted by $\langle - \rangle_t$. The average free energy, corresponding to the interpolated code at time t is given by

$$\alpha_n(t) = \frac{1}{n}\mathbb{E}_{\mathcal{C}_t, \underline{l}, \{u_a^i\}}[\ln Z(t)].$$

At $t = 1$ one recovers the original free energy

$$\mathbb{E}_{\mathcal{C}}[h_n] - \epsilon \ln 2 = \alpha_n(1) = \frac{1}{n}\mathbb{E}_{\mathcal{C}, \underline{l}}[\ln Z].$$

While at $t = 0$, we have a simple product measure which is tailored to yield $h_{RS}[p]$.

In order to lighten the formulas, *from now on we consider the right-regular case* $P(x) = x^r$. It is straightforward to extend the arguments to general polynomials. Also, we use the simplified notation $\mathbb{E}_{\mathcal{C}_t, \underline{l}, \{u_a^i\}} = \mathbb{E}_t$ and $\bar{p} = 1 - p$.

From [29] and as shown in the Appendix 2.A.4, we have the following sum-rule,

$$\alpha_n(1) = h_{RS}[p] + \int_0^1 dt \mathcal{R}(t),$$

$$\mathcal{R}(t) = \frac{\gamma}{r} \sum_{l \geq 1} \frac{(-1)^{l+1}}{l} \mathbb{E}_t\big[\langle R(\bar{p}, Q_l)\rangle_{l,t}\big]. \tag{3.5}$$

Recall that $Q_l = n^{-1}\sum_i x_i^1 x_i^2 \cdots x_i^l$ are the overlap parameters and $\langle - \rangle_{l,t}$ is the replicated Gibbs average (c.f. (2.45) and (2.46)). Also, recall that the polynomial $R(a, b)$ is equal to $(r-1)a^r - ra^{r-1}b + b^r$.

For the case of BEC, there are two major simplifications that occur for the remainder term $\mathcal{R}(t)$. Firstly, it was shown in the Section 2.6 in the Chapter 2, that for almost every ϵ,

$$\mathcal{R}(t) = \frac{\gamma}{r} \sum_{l \geq 1} \frac{(-1)^{l+1}}{l} \mathbb{E}_t\big[R(\bar{p}, \langle Q_l\rangle_{l,t})\big] + o_n(1). \tag{3.6}$$

We stress that the Gibbs bracket $\langle - \rangle_{l,t}$ is now present inside the polynomial rather than outside as in (3.5). Furthermore, when transmitting over the BEC, we receive a bit either perfectly or it is erased. Thus, we have

Lemma 3.1. *For the BEC the random variables* $\langle x_i \rangle_t$ *take values* 0 *or* 1.

Proof. The proof for the interpolated code ensemble at any "time" t, follows on the same lines as the proof of Lemma 2.5 in the Appendix 2.A.2 in Chapter 2. □

The lemma implies that

$$\begin{aligned}\langle Q_l \rangle_{l,t} &= \frac{1}{n}\sum_{i=1}^{n} \left\langle x_i^{(1)} x_i^{(2)} \cdots x_i^{(l)} \right\rangle_{l,t} \\ &= \frac{1}{n}\sum_{i=1}^{n} \langle x_i \rangle_t^l = \frac{1}{n}\sum_{i=1}^{n} \langle x_i \rangle_t = \langle m \rangle_t. \end{aligned} \tag{3.7}$$

Here $m = \frac{1}{n}\sum_{i=1}^{n} x_i$ is called as the "magnetization". From (3.7) we see that the remainder simplifies to

$$\begin{aligned}\frac{\gamma}{r}\sum_{l\geq 1}\frac{(-1)^{l+1}}{l}\mathbb{E}_t\big[R(\bar{p}, \langle Q_l \rangle_{l,t})\big] &= \frac{\gamma}{r}\sum_{l\geq 1}\frac{(-1)^{l+1}}{l}\mathbb{E}_t\big[R(\bar{p}, \langle m \rangle_t)\big] \\ &= \frac{\gamma}{r}\mathbb{E}_t\big[R(\bar{p}, \langle m \rangle_t)\big]\Big(\sum_{l\geq 1}\frac{(-1)^{l+1}}{l}\Big) \\ &= (\ln 2)\frac{\gamma}{r}\mathbb{E}_t\big[R(\bar{p}, \langle m \rangle_t)\big].\end{aligned}$$

Thus, the sum-rule simplifies to

$$\alpha_n(1) = h_{RS}[p] + \ln 2 \frac{\gamma}{r}\int_0^1 dt \mathbb{E}_t\big[R(\bar{p}, \langle m \rangle_t)\big] + o_n(1), \tag{3.8}$$

By abuse of notation we will also denote the remainder term $\frac{\gamma}{r}\mathbb{E}_t\big[R(\bar{p}, \langle m \rangle_t)\big]$ also by $\mathcal{R}(t)$. Above we replaced $1 - \mathtt{R}$ by $\frac{\gamma}{r}$. Note that the polynomial $R(a, b)$ is positive for all $a \geq 0$, $b \geq 0$ and for all r. Since $\langle m \rangle_t \geq 0$ (because of the first GKS inequality, c.f. Appendix 2.A.1), we have that the remainder is positive. As a consequence we re-derive the lower bound

$$\mathbb{E}_{\mathcal{C}}[h_n] \geq h_{RS}[p_{RS}] + o_n(1). \tag{3.9}$$

The simplification of the remainder term gives us a handle on the computations we will do later. It also shows how the magnetization enters the interpolation method analysis. We saw in Chapter 2, the MAP-GEXIT function, in the case of transmission over general channels, can be written in terms of the average (over the Gibbs measure) magnetization. Moreover, for the case of BEC, it can be shown that the average (over the Gibbs measure and noise realizations) magnetization is proportional to the bit-MAP error rate. We will discuss more about the magnetization in the concluding section of this chapter.

Before we prove the exactness of the replica solution, we digress a bit and discuss properties of the BP decoding of the interpolated codes. This will provide us with important technical lemmas that will help us control the remainder term $\mathcal{R}(t)$.

3.4.2 Belief Propagation for the Interpolated Codes

The aim of this section is to study the BP decoder for the interpolated codes. The interpolating system at time t can be thought of as a communication system, where codewords from $\mathcal{C}_t \in \text{LDPC}(n, \gamma t, P)$ are sent via the BEC. The receiver also collects "extra observations" U, distributed as $d_U(u) = \big(1 - \rho(\bar{p})\big)\delta_0(u) + \rho(\bar{p})\delta_\infty(u)$, where $0 < p < 1$ is a free parameter.

Alternatively one can view the system as follows. Codewords from $\mathcal{C}_t$ are transmitted through a BEC with effective erasure probability given by $\epsilon\lambda_{1-t}(1 - \rho(\bar{p}))$. Indeed, the channel is in erasure with probability ϵ and all the "extra observations" are in erasure with probability $\lambda_{1-t}(1 - \rho(\bar{p}))$.

Following standard density evolution analysis, we have that the BP decoder estimate after ℓ iterations is given by

$$x_{\ell+1,t} = \epsilon\lambda_{1-t}\big(1 - \rho(\bar{p})\big)\lambda_t\big(1 - \rho(1 - x_{\ell,t})\big), \tag{3.10}$$

where $\lambda_t(x) = e^{\gamma t(x-1)}$. Above, $x_{\ell+1,t}$ corresponds to the probability that a variable-node-to-check-node message is in erasure.

Recall that, p_{BP} denotes the BP fixed-point of the density evolution equation for the original Poisson-LDPC ensemble $\text{LDPC}(n, \gamma, P)$. In other words, p_{BP} is BP fixed-point of the interpolated code at time $t = 1$. We remark here that later we will set the free parameter p equal to $p_{BP} + \delta$ for some $\delta > 0$. For the purpose of our analysis, it will be important to study the interpolated codes, at time $t < 1$, for this value of p. Let us provide an example illustrating the fixed-point behavior of the interpolated codes at $p = p_{BP} + \delta$.

Example 3.2 (Density Evolution for the Interpolated Codes)**.** *Consider an interpolated Poisson-LDPC$(n, \gamma t = 3t, P(x) = x^6)$ code ensemble. From* (3.10)*, the density evolution equation, for $p = p_{BP} + \delta$, is given by*

$$x = \epsilon e^{-3(1-t)(1-p_{BP}-\delta)^5} e^{-3t(1-x)^5}.$$

In Figure 3.2, we plot the density evolution curves for $\epsilon = 0.55$, $\delta = 0.01$ and for $t \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The BP decoder gets stuck at the largest fixed-point. In the case of $t = 1$, the BP decoder fixed-point is equal to $p_{BP} \approx 0.5015$. The thick dashed curve corresponds to $t = 0$ and the thick continuous curve corresponds to $t = 1$. Notice that for $x > p_{BP} + \delta$, the curves are ordered with $t = 1$ curve at the top followed by $t = 0.8$ and so on with the curve at $t = 0$ at the bottom. This is illustrated in the zoomed-in figure on the right in Figure 3.2. This previous observation is crucial to prove that the fixed-point of the BP decoder for the interpolated code at time $t < 1$ is not "far away" from the fixed-point at $t = 1$ when δ is small.

As we mentioned before, we want to show that if we take $p = p_{BP} + \delta$, for some small $\delta > 0$, then the BP decoder estimate, for an interpolated code at time $t < 1$ is not too worse off the BP fixed-point at $t = 1$. Let us now formally describe this. Consider the BP decoder estimate of the spin (or bit)

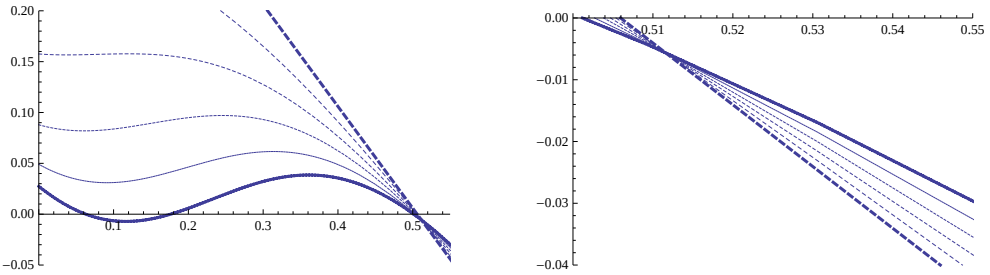

Figure 3.2: Left figure: Density evolution curves for the interpolated Poisson-LDPC$(n, \gamma t = 3t, P(x) = x^6)$ code ensemble for $\epsilon = 0.55$, $\delta = 0.01$ and $t \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1\}$. The BP fixed-point for $t = 1.0$ is equal to $p_{BP} = 0.5015$. The thick continuous curve corresponds to $t = 1$. The thick dashed curve corresponds to $t = 0$. Right figure: A zoomed-in picture around the point $p_{BP} + \delta = 0.5115$, where all the curves intersect. The thick continuous curve corresponds to $t = 1$. The thick dashed curve corresponds to $t = 0$.

x_i after ℓ iterations for the interpolated code C_t. It is possible to regard the BP estimate as a statistical mechanical average on a *computation tree* $T_i(\ell)$ [23] of depth ℓ for node i. One simply considers the Gibbs measure with appropriate check node constraints and observations associated to all nodes appearing in the labeled tree graph $T_i(\ell)$. We denote this average by $\langle x_i \rangle_{t,T_i(\ell)}$. Then we have

Lemma 3.2 (BP Estimate for the Interpolated Code at Time t)**.** *For any t and any $\delta > 0$, one can find a positive number $\ell(\delta)$ (which is equal to the number of iterations of the BP decoder) and a block-length $n(\delta)$, both independent of t, such that for all $n \geq n(\delta)$ and $\ell \geq \ell(\delta)$, if $p = p_{BP} + \delta$, we have*

$$\mathbb{E}_t \left[\langle x_i \rangle_{t,T_i(\ell)} \right] \geq \bar{p}_{BP} - g(\delta),$$

where $0 < g(\delta) < \delta$.

We furnish the proof of the above lemma in the Appendix 3.A. Finally we will need a concentration property for the BP estimate of the interpolated codes.

Lemma 3.3 (Concentration Lemma)**.** *One can find a numerical constant $\beta > 0$, such that for any δ small enough, any fixed ℓ, for n large enough*

$$\mathbb{P}_t \left[\left| \sum_{i=1}^{n} \langle x_i \rangle_{t,T_i(\ell)} - \sum_{i=1}^{n} \mathbb{E}_t \langle x_i \rangle_{t,T_i(\ell)} \right| \geq n\delta \right] \leq e^{-\frac{n\beta\delta^2}{(\ln n)^{2\ell}}} + n^{1-\frac{1}{2}\ln(\ln n)}.$$

The proof of this is also provided in Appendix 3.A.

3.5 Upper Bound on the Remainder Term $\mathcal{R}(t)$

To prove the exactness of RS solution it remains to show that (c.f. (3.8))

$$\int_0^1 dt\mathbb{E}_t\left[R(\bar{p}_{RS},\langle m\rangle_t)\right] \leq o_n(1).$$

Loosely speaking, the philosophy behind the interpolation method is that for $p = p_{RS}$ the removal of the check nodes is "perfectly compensated" by the addition of the extra observations. This implies that $\mathbb{E}_t\left[R(\bar{p}_{RS},\langle m\rangle_t)\right]$ does not change with t. It is straightforward to verify that for $t = 0$ we have $\mathbb{E}_{t=0}\left[R(\bar{p}_{RS},\langle m\rangle_{t=0})\right] = 0$. Hence, the remainder term $\mathbb{E}_t\left[R(\bar{p}_{RS},\langle m\rangle_t)\right]$ is 0 for all t.

We are unable to show directly this perfect compensation. But we will show that for any $\delta > 0$ and $p = p_\delta = p_{RS} + \delta$ the compensation is "almost" perfect. I.e.,

$$\int_0^1 dt\mathbb{E}\left[R(\bar{p}_\delta,\langle m\rangle_t)\right] \leq o_n(1) + O(\delta).$$

Therefore,

$$h_{RS}[p_{RS}] \leq \lim_{n\to\infty}\mathbb{E}_{\mathcal{C}}[h_n] \leq h_{RS}[p_{RS}+\delta] + O(\delta). \tag{3.11}$$

where the lower bound follows from (3.9). Since δ can be made as small as desired, using the continuity of $h_{RS}[p]$ we prove the theorem.

We remark here that we can prove the Lemma 3.2 only when $p = p_{BP} + \delta$ and not when $p = p_{RS} + \delta$. Thus, we see that for the present proof to work, ϵ has to be in the range such that $p_{RS} = p_{BP}$. This is the first condition appearing in Theorem 3.1.

3.5.1 Second Interpolation

From the previous chapter, since the fluctuation of $R(\bar{p}_\delta, m)$ around $R(\bar{p}_\delta,\langle m\rangle_t)$ (w.r.t. to the randomness in noise and code realizations) is small, it is enough to show that $\mathbb{E}\left[\langle R(\bar{p}_\delta, m)\rangle_t\right] \leq o_n(1)+O(\delta)$. We will use a second interpolation in a similar fashion than in [31], [32], [96].

The basic idea is to test the "stability"of the replica solution by coupling the remainder term with the partition function of the interpolated code. More precisely, consider the following partition function at time t, parametrized by a non-negative real number ν,

$$Z(t,\nu,\bar{p}) = \sum_{\underline{x}} e^{\nu n R(\bar{p},m)} \prod_{a\in\mathcal{C}_t}\left(\frac{1+x_{\partial a}}{2}\right)\prod_{i=1}^{n}\frac{e^{\left(\frac{l_i}{2}+\sum_{a=1}^{e_i}\frac{u_a^i}{2}\right)x_i}}{2\cosh\left(\frac{l_i}{2}\right)\prod_{a=1}^{e_i}2\cosh\left(\frac{u_a^i}{2}\right)}. \tag{3.12}$$

We denote by $\langle\cdot\rangle_{t,\nu}$ the average associated to this partition function. We stress here that whenever we write $\langle -\rangle_t$, it denotes the Gibbs average w.r.t. the

partition function with $\nu = 0$ in (3.12). Also, the associated free energy is given by

$$\alpha_n(t, \nu, \bar{p}) = \frac{1}{n} \ln Z(t, \nu, \bar{p}),$$

The intuition behind modifying the partition function as in (3.12) is as follows. Whenever the term $R(\bar{p}, m) > 0$, the corresponding configuration, $\underline{x}$, has large weight. The idea of the second interpolation method of Guerra [31] is that, if the parameter ν is chosen carefully, then it can be shown that the average free energy ($\nu > 0$) is very close to the average of $\alpha_n(t, 0, \bar{p})$ ($\nu = 0$) (in the thermodynamic limit ($n \to +\infty$)). This implies that most configurations have $R(\bar{p}, m) = 0$ and one concludes that $\mathbb{E}\left[\langle R(\bar{p}, m)\rangle_t\right]$ is small.

Let us now make this idea more precise. First, we need to relate the free energy to the remainder term. To do this, we use the convexity of $\alpha_n(t, \nu, \bar{p})$ with respect to ν. This is the content of the following lemma,

Lemma 3.4 (Convexity of Free Energy)**.** *The free energy $\alpha_n(t, \nu, \bar{p})$ is convex with respect to ν for any $\nu \geq 0$.*

Proof. Taking the derivative of the free energy we get

$$\frac{d}{d\nu}\alpha_n(t, \nu, \bar{p}) = \langle R(\bar{p}, m)\rangle_{t,\nu}.$$

Taking the derivative of $\langle R(\bar{p}, m)\rangle_{t,\nu}$ we have

$$\frac{d}{d\nu}\langle R(\bar{p}, m)\rangle_{t,\nu} = \langle R^2(\bar{p}, m)\rangle_{t,\nu} - \langle R(\bar{p}, m)\rangle_{t,\nu}^2 \geq 0.$$

Since the second derivative of the free energy is non-negative, we have that the free energy is convex. □

Now using the convexity of the free energy we relate the remainder term to the difference of free energies as follows,

$$\begin{aligned}\mathbb{E}_t[\langle R(\bar{p}, m)\rangle_t] &= \frac{\partial}{\partial\nu}\mathbb{E}_t[\alpha_n(t, \nu, \bar{p})]\Big|_{\nu=0} \\ &\leq \frac{1}{\nu}\mathbb{E}_t[\alpha_n(t, \nu, \bar{p}) - \alpha_n(t, 0, \bar{p})]. \end{aligned} \tag{3.13}$$

The inequality above follows from the convexity of $\alpha_n(t, \nu, \bar{p})$ w.r.t. ν.

So now the next task would be to show that the free energy at $\nu \neq 0$ is "very close" to $\nu = 0$. However, the presence of the code structure in the partition function makes it difficult to estimate the free energy $\alpha_n(t, -, -)$, for any $\nu \geq 0$ and $t \neq 0$.

To circumvent this problem, we compute $\alpha_n(0, -, -)$ and estimate $\alpha_n(t, -, -)$ by using again the interpolation method. The "miracle" that occurs now is that, by carefully choosing ν to depend on t, the remainder of the second interpolation is negative. Thus, if the erasure fraction is such that the estimate of

$\alpha_n(0,-,-)$ is small, then the remainder term is small. Note that the absence of code structure allows us to estimate $\alpha(0,-,-)$.

It turns out that the good choice for ν is $\nu(t) = \frac{\gamma}{r}(1-t)$. Applying the interpolation method to (3.13) we get,

Lemma 3.5 (Second Interpolation)**.**

$$\mathbb{E}_t\big[\alpha_n(t,\nu(t),\bar{p}) - \alpha_n(t,0,\bar{p})\big] \leq \Delta_n\big(\frac{\gamma}{r},\bar{p}\big) + \frac{\gamma}{r}\int_0^t \mathfrak{R}(\bar{p},s)ds + o_n(1), \quad (3.14)$$

where

$$\Delta_n\big(\frac{\gamma}{r},\bar{p}\big) = \mathbb{E}\Big[\alpha_n(0,\frac{\gamma}{r},\bar{p}) - \alpha_n(0,0,\bar{p})\Big], \quad (3.15)$$

$$\mathfrak{R}(\bar{p},s) = \sum_{l\geq 2}\frac{(-1)^{l+1}}{l}\mathbb{E}\langle R(\bar{p},Q_l)\rangle_{l,s,\nu(s)}. \quad (3.16)$$

Where Q_l is, as before, the overlap parameter (c.f. (2.45)) consisting of l copies of identical and independent spins $\underline{x}^{(\alpha)}$ for $\alpha = 1,2,\ldots,l$. Above, $\langle - \rangle_{l,s,\nu(s)}$ denotes the *replicated* Gibbs average. More precisely, let $\mu_{s,\nu(s)}(\underline{x})$ denotes the Gibbs measure associated to the partition function in (3.12). Then the replicated Gibbs measure is given by

$$\mu_{l,s,\nu(s)}(\underline{x}^{(1)},\underline{x}^{(2)},\ldots,\underline{x}^{(l)}) \triangleq \prod_{\alpha=1}^{l}\mu_{s,\nu(s)}(\underline{x}^{(\alpha)}). \quad (3.17)$$

To lighten the notation we denote the replicated Gibbs measure by $\mu_{l,s,\nu(s)}$ and the un-replicated by $\mu_{s,\nu(s)}$. Also, we stress that at this point the parameter $\bar{p}$ is still free. As we will see shortly that, setting it to $p_{BP}+\delta$ allows us to estimate the r.h.s. (3.14).

We remark (and we will show this) that the judicious choice of $\nu(t)$ allows us to cancel the $l=1$ term in (3.16). The proof of the Lemma 3.5 is provided in the Appendix 3.B.

Note that with the choice of $\nu(t) = \frac{\gamma}{r}(1-t)$, the upper bound provided by (3.13) on the remainder term $\mathbb{E}_t[\langle R(\bar{p},m)\rangle_t]$, diverges. But we circumvent this problem as follows. We split the integral over t in to the two intervals $[0,1-\delta]$ and $[1-\delta,1]$. Thus, we get

$$\begin{aligned}\int_0^1 dt\mathbb{E}\langle R(\bar{p},m)\rangle_t &= \int_0^{1-\delta} dt\mathbb{E}\langle R(\bar{p},m)\rangle_t + \int_{1-\delta}^1 dt\mathbb{E}\langle R(\bar{p},m)\rangle_t \\ &\leq \int_0^{1-\delta} dt\mathbb{E}\langle R(\bar{p},m)\rangle_t + 2r\delta \\ &\leq 2r\delta + \frac{r}{\gamma}\Delta_n(\frac{\gamma}{r},\bar{p})\ln\frac{1}{\delta} + \frac{\gamma}{r}\int_0^{1-\delta} dt\frac{1}{\nu(t)}\int_0^t \mathfrak{R}(\bar{p},s)ds. \end{aligned} \quad (3.18)$$

We use the fact that $|R(\bar{p},m)| \leq 2r$ for check node of degree r to get the first inequality. To obtain the last inequality we combined (3.13) and (3.14). Note

that in obtaining the term $\left(\frac{r}{\gamma}\Delta_n(\frac{\gamma}{r},\bar{p})\ln\frac{1}{\delta}\right)$, we have used the independence of $\Delta_n(\frac{\gamma}{r},\bar{p})$ w.r.t. t.

Note that the r.h.s. of (3.18) has three terms. In the remaining of the chapter we show that the last term corresponding to the remainder $\mathfrak{R}(\bar{p},s)$ is non-positive when we put $\bar{p}=\bar{p}_\delta = 1-p_{RS}-\delta$. We will also show that the second term in (3.18) yields a contribution of the form $o_n(1)+O(\delta\ln\delta)$, for the choice of $p=p_\delta=p_{RS}+\delta$. As a consequence, we obtain (3.11) for any $\delta>0$. Hence we prove the main theorem on the equality of the replica solution.

3.5.2 Estimate of (3.15)

Recall that we are assuming $P(x)=x^r$, i.e. each check node has degree r. One can explicitly compute (3.15). Indeed, the free energy $\alpha_n(0,0,\bar{p})$ corresponds to a Gibbs measure with a product form (without the code structure and without the term $e^{\nu nR(\bar{p},m)}$). For the other free energy $\alpha_n(0,\frac{\gamma}{r},\bar{p})$ the situation is more complicated. Nevertheless, the code $\mathcal{C}_t$ is absent and the problem is similar to the computation of a free energy of a non-random complete p-spin model [96]. This can be computed by the saddle point methods similar to [96]. Recall that $p_\delta = p_{RS}+\delta$. The net result of this long calculation, which we show briefly in Appendix 3.E, is

$$\Delta_n\Big(\frac{\gamma}{r},\bar{p}_\delta\Big)=\frac{\gamma}{r}(r-1)\bar{p}_{RS}^r+\max_z f(z)+O(\delta)+o_n(1)$$

where $f(z)$ was defined in the Section 3.3. For $z=\bar{p}_{RS}$, we have $f(z)=\frac{\gamma}{r}\bar{p}_{RS}^r\Big(1-r\Big)$. Thus, for ϵ such that the maximizer of $f(z)$ is at $\hat{z}=\bar{p}_{RS}$, the contribution of $\Delta_n\big(\frac{\gamma}{r},\bar{p}_\delta\big)$ vanishes as $n\to+\infty$.

3.5.3 Estimate of Second Interpolation Remainder Term $\mathfrak{R}(\bar{p}_\delta,s)$

Recall that we are assuming $P(x)=x^r$, i.e. each check node has degree r. In this section we will show that the remainder of the second interpolation, $\mathfrak{R}(\bar{p}_\delta,s)$, is non-positive.

Note that when $\nu=0$, we could combine the even and odd terms in the remainder (c.f. (2.44)) using channel symmetry. This then helped to determine the sign of the remainder term. For example, when $P(x)$ is convex, then the remainder was positive.

It is difficult to estimate the sign of $\mathfrak{R}(\bar{p}_\delta,s)$, because unlike the case $\nu=0$ we cannot use the channel symmetry to combine odd and even terms (which could potentially help us to estimate the sign). The term $e^{\nu nR(\bar{p},m)}$ is said to *break the symmetry* in the partition function (3.12). More precisely, tools such as the GKS inequality or Nishimori identities (channel symmetry) fail to apply for $\nu>0$. As a result, we are also unable to reduce all the overlap terms to magnetization, as we did in (3.7).

To estimate the sign of $\mathfrak{R}(\bar{p}_\delta, s)$ we first establish a relation between the $\nu \neq 0$ system and $\nu = 0$ system. Recall that $\langle - \rangle_s$ denotes the Gibbs average w.r.t. to the partition function (3.12) with $\nu = 0$. We have the following two crucial technical lemmas.

Lemma 3.6. *For any $0 \leq s \leq 1$ and $\nu \geq 0$, if $\langle x_i \rangle_s = 1$, then $\langle x_i \rangle_{s,\nu} = 1$.*

Lemma 3.7. *For any $0 \leq s \leq 1$ and $\nu \geq 0$, we have $\langle x_i \rangle_{s,\nu} \geq -1 + e^{-4n\nu r}$.*

The proofs of the above two lemmas are relegated to the Appendix 3.C.

Before we go on to estimate the sign of the remainder, we first relate the overlap parameters appearing in $\mathfrak{R}(\bar{p}_\delta, s)$ to the magnetization of the original interpolated code. More precisely, define the random variable $q = \langle m \rangle_s$. Also define the random set of variable nodes $S = \{i : \langle x_i \rangle_s = 0\}$. We stress here that $\langle - \rangle_s$ is the Gibbs average associated to the partition function in (3.12) with $\nu = 0$. In words, this is the MAP measure for the original interpolated code at time s. The randomness of the above defined quantities arises because of randomness in channel noise and code realizations (at time s).

Recall that $\mu_{s,\nu(s)}$ denotes the Gibbs measure associated to $\langle - \rangle_{s,\nu(s)}$. Also recall that $\mu_{l,s,\nu(s)}$ denotes the Gibbs measure associated to $\langle - \rangle_{l,s,\nu(s)}$. We have

Lemma 3.8. *We have*

$$\begin{aligned}
\left\langle Q_l \right\rangle_{l,s,\nu(s)} &= q + \frac{1}{n} \sum_{i \in S} \left\langle x_i \right\rangle^l_{s,\nu(s)}, \\
\left\langle Q_l^r \right\rangle_{l,s,\nu(s)} &= \sum_{j=0}^{r} \binom{r}{j} q^{r-j} \frac{1}{n^j} \sum_{i_1, \ldots i_j \in S} \langle x_{i_1} \ldots x_{i_j} \rangle^l_{s,\nu(s)}.
\end{aligned}$$

Proof. From the definition of q we have

$$q = \langle m \rangle_s = \frac{1}{n} \sum_{i=1}^{n} \langle x_i \rangle_s = \frac{1}{n} \sum_{i \notin S} \langle x_i \rangle_s = \frac{1}{n} \sum_{i \notin S} 1. \tag{3.19}$$

From the Lemma 3.6, for $i \notin S$ we have $\langle x_i \rangle_s = 1$ which implies $\langle x_i \rangle_{s,\nu} = 1$. Since $-1 \leq x_i \leq 1$ and $\langle - \rangle_{s,\nu}$ is an average w.r.t. a probability measure, we immediately obtain that $x_i = +1$ with probability 1 under the Gibbs measure given by $\mu_{s,\nu(s)}$.

Since in Q_l we couple l *identical* systems, associated with the measure $\mu_{s,\nu(s)}$, we must have that $x_i^{(1)} = x_i^{(2)} = \cdots = x_i^{(l)} = 1$ for all $i \notin S$ with probability 1. Thus, we have

$$\begin{aligned}
\langle Q_l \rangle_{l,s,\nu(s)} &= \left\langle \frac{1}{n} \sum_{i \notin S} x_i^{(1)} \ldots x_i^{(l)} \right\rangle_{l,s,\nu(s)} + \left\langle \frac{1}{n} \sum_{i \in S} x_i^{(1)} \ldots x_i^{(l)} \right\rangle_{l,s,\nu(s)}, \\
&= \frac{1}{n} \sum_{i \notin S} 1 + \frac{1}{n} \sum_{i \in S} \left\langle x_i^{(1)} \ldots x_i^{(l)} \right\rangle_{l,s,\nu(s)}, \\
&= q + \frac{1}{n} \sum_{i \in S} \langle x_i \rangle^l_{s,\nu(s)}.
\end{aligned}$$

To get the second equality above, we used that for $i \notin S$, $x_i^{(1)} = x_i^{(2)} = \cdots = x_i^{(l)} = +1$ under the Gibbs measure given by $\mu_{l,s,\nu(s)}$. The last equality is got by using (3.19).

Let us now prove the second expression in the statement of the lemma. Usual computations give us

$$\langle Q_l^r \rangle_{l,s,\nu(s)} = \Big\langle \frac{1}{n^r} \sum_{i_1,i_2,\ldots,i_r} \prod_{\alpha=1}^{l} x_{i_1}^{(\alpha)} x_{i_2}^{(\alpha)} \ldots x_{i_r}^{(\alpha)} \Big\rangle_{l,s,\nu(s)} \tag{3.20}$$

$$= \frac{1}{n^r} \sum_{i_1,i_2,\ldots,i_r} \Big\langle \prod_{\alpha=1}^{l} x_{i_1}^{(\alpha)} x_{i_2}^{(\alpha)} \ldots x_{i_r}^{(\alpha)} \Big\rangle_{l,s,\nu(s)} \tag{3.21}$$

$$= \frac{1}{n^r} \sum_{i_1,i_2,\ldots,i_r} \Big\langle x_{i_1} x_{i_2} \ldots x_{i_r} \Big\rangle_{s,\nu(s)}^{l}. \tag{3.22}$$

Suppose that $i_1, i_2, \ldots, i_j \in S$. The remaining indices are such that the corresponding variable nodes $\notin S$. Under the Gibbs measure given by $\mu_{s,\nu(s)}$ we know that $x_{i_{j+1}} = \cdots = x_{i_r} = 1$. This would imply that

$$\sum_{i_1,i_2,\ldots,i_r} \Big\langle x_{i_1} x_{i_2} \ldots x_{i_r} \Big\rangle_{s,\nu(s)}^{l} = \sum_{i_1,i_2,\ldots,i_j} \Big\langle x_{i_1} x_{i_2} \ldots x_{i_j} \Big\rangle_{s,\nu(s)}^{l} \Big(\sum_{i_{j+1} \notin S, \ldots, i_r \notin S} 1 \Big) \tag{3.23}$$

$$= \sum_{i_1,i_2,\ldots,i_j} \Big\langle x_{i_1} x_{i_2} \ldots x_{i_j} \Big\rangle_{s,\nu(s)}^{l} \Big(\sum_{i_{j+1} \notin S} 1 \ldots \sum_{i_r \notin S} 1 \Big) \tag{3.24}$$

$$= q^{r-j} n^{r-j} \sum_{i_1,i_2,\ldots,i_j} \Big\langle x_{i_1} x_{i_2} \ldots x_{i_j} \Big\rangle_{s,\nu(s)}^{l}. \tag{3.25}$$

We get the last equality above by using (3.19). We can choose j number of indices, belonging to S, in $\binom{r}{j}$ ways. Combining (3.20) and (3.23) along with the previous fact, we get the second expression of the lemma. □

Thanks to the lemmas above we can prove the main result,

Lemma 3.9 (Sign of the Remainder Term)**.** *Let the channel erasure probability, ϵ, be such that $p_{RS} = p_{BP}$, i.e. the replica solution is maximized by the fixed-point of density evolution obtained via BP decoding. Then for any $\delta > 0$, there exists $n(\delta)$ such that for all $n \geq n(\delta)$,*

$$\mathfrak{R}(\bar{p}_\delta, s) = \sum_{l \geq 2} \frac{(-1)^{l+1}}{l} \mathbb{E} \langle R(\bar{p}_\delta, Q_l) \rangle_{l,s,\nu(s)} \leq 0.$$

Proof. Since $R(a,b) = (r-1)a^r - ra^{r-1}b + b^r$, using Lemma 3.8 we get

$$\langle R(\bar{p}_\delta, Q_l) \rangle_{s,\nu(s)} = \sum_{j=0}^{r} \binom{r}{j} q^{r-j} \frac{1}{n^j} \sum_{i_1,\ldots i_j \in S} \langle x_{i_1} \ldots x_{i_j} \rangle_{s,\nu(s)}^{l} - r\bar{p}_\delta^{r-1} q$$

$$- r\bar{p}_\delta^{r-1} \frac{1}{n} \sum_{i \in S} \langle x_i \rangle_{s,\nu(s)}^{l} + (r-1)\bar{p}_\delta^{r}. \tag{3.26}$$

Let us write the previous expression in a convenient form. We separate out the $j = 0, 1$ term in the sum in (3.26) to find

$$\begin{aligned}\langle R(\bar{p}_\delta, Q_l)\rangle_{s,\nu(s)} = &\sum_{j=2}^{r} \binom{r}{j} q^{r-j} \frac{1}{n^j} \sum_{i_1,\dots i_j \in S} \langle x_{i_1}\dots x_{i_j}\rangle^l_{s,\nu(s)} \\ &+ \left(rq^{r-1} - r\bar{p}_\delta^{r-1}\right)\frac{1}{n}\sum_{i\in S}\langle x_i\rangle^l_{s,\nu(s)} \\ &+ (r-1)\bar{p}_\delta^r - r\bar{p}_\delta^{r-1}q + q^r. \end{aligned} \tag{3.27}$$

We recognize that the last term above is equal to $R(\bar{p}_\delta, q)$.

Using $\langle x_{i_1}\dots x_{i_j}\rangle^{2k}_{s,\nu(s)} \geq \langle x_{i_1}\dots x_{i_j}\rangle^{2k+1}_{s,\nu(s)}$ and $\langle x_{i_1}\dots x_{i_j}\rangle^{2k}_{s,\nu(s)} \geq 0$, we have

$$\begin{aligned}&\sum_{j\geq 2}^{r} \binom{r}{j} q^{r-j} \frac{1}{n^j}\left(\frac{-1}{2k}\sum_{i_1,\dots i_j\in S}\langle x_{i_1}\dots x_{i_j}\rangle^{2k}_{s,\nu(s)} + \frac{1}{2k+1}\sum_{i_1,\dots i_j\in S}\langle x_{i_1}\dots x_{i_j}\rangle^{2k+1}_{s,\nu(s)}\right) \\ &\qquad \leq \frac{-1}{2k+1}\sum_{j\geq 2}^{r}\binom{r}{j}q^{r-j}\frac{1}{n^j}\sum_{i_1,\dots i_j\in S}\left(\langle x_{i_1}\dots x_{i_j}\rangle^{2k}_{s,\nu(s)} - \langle x_{i_1}\dots x_{i_j}\rangle^{2k+1}_{s,\nu(s)}\right) \\ &\qquad \leq 0. \end{aligned} \tag{3.28}$$

Using (3.28) and combining $j = 0$ and $j = 1$ term, the addition of the odd and even terms $2k$ and $2k + 1$ in the remainder (3.16) gives us,

$$-\frac{1}{2k}\mathbb{E}\langle R(\bar{p}_\delta, Q_{2k})\rangle_{s,\nu(s)} + \frac{1}{2k+1}\mathbb{E}\langle R(\bar{p}_\delta, Q_{2k+1})\rangle_{s,\nu(s)} \tag{3.29}$$

$$\leq -\frac{1}{2k(2k+1)}\mathbb{E}_s[R(\bar{p}_\delta, q)] - \frac{1}{2k}\mathbb{E}_s[\mathcal{T}_{2k}] + \frac{1}{2k+1}\mathbb{E}_s[\mathcal{T}_{2k+1}]. \tag{3.30}$$

We use the notation,

$$\mathcal{T}_l = r(q^{r-1} - \bar{p}_\delta^{r-1})\frac{1}{n}\sum_{i\in S}\langle x_i\rangle^l_{s,\nu(s)}.$$

Let $\mathcal{A}_s$ be the set of channel noise realizations and code realizations (at time s) such that $\{q \geq \bar{p}_{BP} - \frac{\delta}{2} - \frac{g(\delta)}{2}\}$. Here $g(\delta)$ is the same as in Lemma 3.2. Since $g(\delta) < \delta$, we must have $q > \bar{p}_\delta$ in the set $\mathcal{A}_s$. As a consequence, in the set $\mathcal{A}_s$, we have $-\frac{\mathcal{T}_{2k}}{2k} + \frac{\mathcal{T}_{2k+1}}{2k+1} \leq 0$. Thus, we can upper bound the r.h.s. in (3.29) by

$$-\frac{1}{2k(2k+1)}\mathbb{E}_s[R(\bar{p}_\delta, q)] - \frac{1}{2k}\mathbb{E}_{\mathcal{A}_s^c}[\mathcal{T}_{2k}] + \frac{1}{2k+1}\mathbb{E}_s[\mathcal{T}_{2k+1}]. \tag{3.31}$$

Furthermore, since $|q| \leq 1$ and $|\bar{p}_\delta| \leq 1$ we have $|R(\bar{p}_\delta, q)| \leq 2r$. For the set $\mathcal{A}_s$, a quick calculation shows that $R(\bar{p}_\delta, q) \geq r(r-1)(\delta - g(\delta))^2\bar{p}_{BP}^{r-2}$. Thus,

we can bound

$$-\frac{1}{2k(2k+1)}\mathbb{E}_s[R(\bar{p}_\delta,q)] = -\frac{1}{2k(2k+1)}\mathbb{E}_{\mathcal{A}_s}[R(\bar{p}_\delta,q)] - \frac{1}{2k(2k+1)}\mathbb{E}_{\mathcal{A}_s^c}[R(\bar{p}_\delta,q)]$$
$$\leq -\frac{1}{2k(2k+1)}r(r-1)(\delta-g(\delta))^2\bar{p}_{BP}^{r-2}\mathbb{P}[\mathcal{A}_s] + \frac{2r}{2k(2k+1)}\mathbb{P}[\mathcal{A}_s^c]. \tag{3.32}$$

Thus, combining (3.29), (3.31) and (3.32) we get

$$-\frac{1}{2k}\mathbb{E}\langle R(\bar{p}_\delta,Q_{2k})\rangle_{s,\nu(s)} + \frac{1}{2k+1}\mathbb{E}\langle R(\bar{p}_\delta,Q_{2k+1})\rangle_{s,\nu(s)}$$
$$\leq \mathbb{E}_{\mathcal{A}_s^c}\left[\frac{-\mathcal{T}_{2k}}{2k}+\frac{\mathcal{T}_{2k+1}}{2k+1}\right] - \frac{r(r-1)(\delta-g(\delta))^2}{2k(2k+1)}\bar{p}_{BP}^{r-2}\mathbb{P}[\mathcal{A}_s] + \frac{2r}{2k(2k+1)}\mathbb{P}[\mathcal{A}_s^c].$$

Summing over $k\geq 1$ and using the fact that $\sum_{k\geq 1}\frac{1}{2k(2k+1)}$ converges to a constant κ, we get

$$\mathfrak{R}(\bar{p}_\delta,s) \leq \sum_{k\geq 1}\mathbb{E}_{\mathcal{A}_s^c}\left[\frac{-\mathcal{T}_{2k}}{2k}+\frac{\mathcal{T}_{2k+1}}{2k+1}\right] - \kappa r(r-1)(\delta-g(\delta))^2\bar{p}_{BP}^{r-2}\mathbb{P}[\mathcal{A}_s]$$
$$+ 2r\kappa\mathbb{P}[\mathcal{A}_s^c].$$

Let us estimate the first term above. Writing explicitly the terms $\mathcal{T}_{2k}$ and $\mathcal{T}_{2k+1}$ we have

$$\sum_{k\geq 1}\mathbb{E}_{\mathcal{A}_s^c}\left[\frac{-\mathcal{T}_{2k}}{2k}+\frac{\mathcal{T}_{2k+1}}{2k+1}\right] = r\mathbb{E}_{\mathcal{A}_s^c}\left[(q^{r-1}-\bar{p}_\delta^{r-1})\frac{1}{n}\sum_{i\in S}\sum_{k\geq 1}\left(\frac{-\langle x_i\rangle_{s,\nu(s)}^{2k}}{2k}+\frac{\langle x_i\rangle_{s,\nu(s)}^{2k+1}}{2k+1}\right)\right]$$
$$= r\mathbb{E}_{\mathcal{A}_s^c}\left[(q^{r-1}-\bar{p}_\delta^{r-1})\frac{1}{n}\sum_{i\in S}\sum_{k\geq 2}\frac{(-1)^{k+1}}{k}\langle x_i\rangle_{s,\nu(s)}^{k}\right]$$
$$= r\mathbb{E}_{\mathcal{A}_s^c}\left[(q^{r-1}-\bar{p}_\delta^{r-1})\frac{1}{n}\sum_{i\in S}\Big(\ln(1+\langle x_i\rangle_{s,\nu(s)})-\langle x_i\rangle_{s,\nu(s)}\Big)\right]$$

Using the Lemma 3.7,

$$r\mathbb{E}_{\mathcal{A}_s^c}\Big[(q^{r-1}-\bar{p}_\delta^{r-1})\frac{1}{n}\sum_{i\in S}\ln(1+\langle x_i\rangle_{s,\nu(s)}^{l})\Big] \leq 2r\mathbb{E}_{\mathcal{A}_s^c}\Big[\frac{1}{n}\sum_{i\in S}4nv(s)r\Big]$$
$$\leq \frac{8nr^2\nu(s)}{\gamma}\mathbb{P}[\mathcal{A}_s^c].$$

Putting everything together, we get

$$\mathfrak{R}(\bar{p}_\delta,s) \leq \Big(\frac{8nr^2\nu(s)}{\gamma}+2r(\kappa+1)\Big)\mathbb{P}[\mathcal{A}_s^c] - \kappa r(r-1)(\delta-g(\delta))^2\bar{p}_{BP}^{r-2}\mathbb{P}[\mathcal{A}_s].$$

This yields a negative contribution for any fixed δ and n large enough provided that we show $\mathbb{P}[\mathcal{A}_s^c]\to 0$ faster than n. To bound the probability of

the event $\mathcal{A}_s^c$ we note that for any realization of the randomness (channel noise and code) $\langle x_i \rangle_s \geq \langle x_i \rangle_{s,T_i(\ell)}$. Note that this is w.r.t. to the original interpolated code at time s. Indeed, if the iterative decoder succeeds and the BP estimate is 1 then the MAP estimate is also necessarily 1. On the other hand if the iterative decoder fails then the BP estimate is 0 and is surely less than the MAP estimate which is 0 or 1. As a result, when $\frac{1}{n}\sum_i \langle x_i \rangle_s < \bar{p}_{BP} - \frac{\delta}{2} - \frac{g(\delta)}{2}$, we must have $\frac{1}{n}\sum_i \langle x_i \rangle_{s,T_i(\ell)} < \bar{p}_{BP} - \frac{\delta}{2} - \frac{g(\delta)}{2}$. Combining the above with Lemma 3.2, implies that there is a $n(\delta)$ such that for $n \geq n(\delta)$,

$$\begin{aligned}
\mathbb{P}[\mathcal{A}_s^c] &= \mathbb{P}[\frac{1}{n}\sum_i \langle x_i \rangle_s < \bar{p}_{BP} - \frac{\delta}{2} - \frac{g(\delta)}{2}] \\
&\leq \mathbb{P}[\frac{1}{n}\sum_i \langle x_i \rangle_{s,T_i(\ell)} < \bar{p}_{BP} - \frac{\delta}{2} - \frac{g(\delta)}{2}] \\
&\leq \mathbb{P}\Big[\sum_i \langle x_i \rangle_{s,T_i(l)} \leq \sum_i \mathbb{E}\big[\langle x_i \rangle_{s,T_i(l)}\big] - \frac{n}{2}(\delta - g(\delta))\Big],
\end{aligned}$$

and the result follows from Lemma 3.3. □

3.6 Road (Block) Ahead

In this chapter we showed that the replica solution provides the correct solution to the per-bit conditional entropy, in the asymptotic limit, when transmitting over the BEC. Similar results can be obtained for the Poisson LDGM ensembles. The extension of the present methods to a standard irregular LDPC ensembles and general BMS channels is still open[3]. Because of the non-combinatorial nature of the proof we are hopeful that such an extension is possible.

One of the highlights of the current approach in the case of BEC is that the remainder term of the first interpolation simplified radically. The remainder consisted only of the magnetization. As a result, the coupled partition function is kept simple (c.f. (3.12)). However, for the case of transmission over general BMS channels, we are unable to simplify the remainder term. Thus, the coupled partition function may involve a large number of replicas. Consequently, controlling the remainder term of the second interpolation might be very difficult. Fortunately, as the next lemma shows, if the magnetization (or the first term in the remainder) can be "controlled", then the entire remainder term can be shown to be small.

The lemma holds under the assumption that all the overlap parameters concentrate on their average. More precisely, we conjecture that the remainder, $\mathcal{R}(t)$, in (3.5) can be written as follows.

[3]Although, in the next two chapters we will attack the problem for the case of transmission over general BMS, by another method.

Conjecture 3.1 (Concentration of the Overlap Parameters)**.** *Consider transmission over a BMS channel using a standard LDPC*(Λ, P) *code ensemble. Then the remainder term of the first interpolation method is given by*

$$\mathcal{R}(t) = \frac{\gamma}{r} \sum_{l \geq 1} \frac{(-1)^{l+1}}{l} R(q_l, \mathbb{E}_t[\langle Q_l \rangle_t]) + o_n(1).$$

Notice that the average w.r.t. all randomness is present inside the polynomial $R(-,-)$. We remark that we have already proven one concentration in the form of equation (3.6).

We state and prove the following lemma only for Poisson-LDPC code ensembles. We hope it would be possible to extend it to any standard LDPC ensemble via the multi-Poisson ensemble approximation of Montanari [29].

Lemma 3.10 (Magnetization as the Order Parameter)**.** *Consider transmission over any general BMS*(ϵ) *channel using Poisson-LDPC*(n, γ, r) *code ensemble with* $r \geq 2$*. Let* C_{BP} *be the set of channel noise values, such that the maximizer of the replica solution is the BP fixed-point of density evolution. Further assume Conjecture 3.1 holds. If, for a given channel noise value* $\epsilon \in C_{BP}$ *and for any* δ *such that* $0 < \delta \ll q_1$*, we have,*

$$\lim_{n \to +\infty} \int_0^1 dt R(q_1, \mathbb{E}_t[\langle x_i \rangle_t]) < \delta, \tag{3.33}$$

then we have

$$\lim_{n \to +\infty} \int_0^1 dt \mathcal{R}(t) < \frac{4\gamma}{(q_1 - \delta)^{\frac{r-2}{2}}} \delta^{1/4} + \frac{2\gamma}{r} \delta^{3/4}. \tag{3.34}$$

Proof. We provide the proof of this lemma in the Appendix 3.D. □

The above lemma allows us to consider a partition function coupled with only the magnetization. Even with the above simplification, proving Lemma 3.9 (controlling the remainder term of the second interpolation) for channels other than the BEC is still open.

Before we conclude the chapter, we provide a concentration result for the magnetization and demonstrate some of its implications.

3.6.1 Role of Magnetization

As we saw in the above section, controlling the magnetization would allow us to prove that the replica solution is correct. Magnetization also appears in a straightforward way in the MAP-GEXIT formulas for the BEC and BIAWGNC (see Section 2.3).

In most relevant models in statistical mechanics or communications, the magnetization is believed to concentrate. But proving this fact is not trivial because it involves a bound on the second derivative of a free energy. This

quantity is not uniformly bounded (with respect to system size) at phase transition thresholds. In this section we provide a concentration theorem for the magnetization for any linear block-length code. We show various implications of the concentration result. One of the implications is an *informal justification of the exactness of the replica solution for the communication problem.*

The concentration result (Theorem 3.2) is valid for general fixed linear codes C of length n. This theorem states that, for a given linear code, for almost all values of the noise parameter, the magnetization concentrates on its average over the channel output realizations. By minor modifications in the proofs one can also obtain the corresponding statements for the standard LDPC and LDGM code ensembles.

Recall that the magnetization is given by $m = \frac{1}{n}\sum_{i=1}^{n} x_i$. With all the standard notations we have

Theorem 3.2. *Consider communication over the BEC(ϵ), where ϵ is the erasure probability or over the BIAWGNC(ϵ) where ϵ^2 is the noise variance. There exists a finite positive constant c possibly depending on a, b and independent of n such that for any linear block code of length n,*

$$\int_a^b d\epsilon \mathbb{E}_{\underline{l}}\big[\langle |m - \mathbb{E}_{\underline{l}}[\langle m\rangle]|\rangle\big] \leq \frac{c}{n^{\frac{1}{8}}},$$

where $[a, b] \in (0, 1)$ for BEC and $[a, b] \in (0, \infty)$ for BIAWGNC.

Proof. For a proof, please see [97]. □

Remark 3.2. *For standard LDPC ensembles and LDGM ensembles, we can show the same statement with $\mathbb{E}_{\underline{l}}$ replaced by $\mathbb{E}_{\mathcal{C},\underline{l}}$. In particular, by dominated convergence it follows that for Lebesgue almost every $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{C},\underline{l}}\langle |m - \mathbb{E}_{\mathcal{C},\underline{l}}\langle m\rangle|\rangle = 0, \tag{3.35}$$

The statement holds for almost every ϵ because it cannot be valid at phase transitions. In fact, one expects that the phase transitions occur at isolated points. Thus, the statement of the above theorem should hold for all ϵ away from these points.

3.6.2 Implications of Concentration of Magnetization

Fraction of bits in error. Consider the BEC and let $P_e(\underline{l})$ denote the fraction of bits in error for a given channel output realization $\underline{l}$,

$$P_e(\underline{l}) = \frac{1}{2}(1 - \langle m\rangle).$$

This formula is valid for the BEC because a bit x_i is either decoded correctly or not decoded, which implies $\langle x_i\rangle \in \{0, 1\}$ (for more general channels the

right hand side has to be replaced by $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(1-\mathrm{sgn}\langle x_i\rangle)$ with $\mathrm{sgn}(0)=0$). The theorem implies concentration of the fraction of errors over its average which is nothing else than the bit MAP error probability,

$$P_{e,\mathrm{MAP}} = \frac{1}{2}(1-\mathbb{E}_{\underline{l}}\langle m\rangle).$$

This follows by bounding

$$\mathbb{E}_{\underline{l}}\big[|\langle m\rangle - \mathbb{E}_{\underline{l}}\langle m\rangle|\big] \le \mathbb{E}_{\underline{l}}\langle |m-\mathbb{E}_{\underline{l}}\langle m\rangle|\rangle.$$

From Remark 3.2, the bit error probability of MAP decoding further concentrates on its average over the code ensemble $\mathbb{E}_{\mathcal{C}}[P_e]$.

Concentration of MAP-GEXIT function. For BEC(ϵ) and BIAWGNC(ϵ), recall from Section 2.3 that we have the following formulas for the MAP-GEXIT function,

$$\begin{aligned}\text{BEC:}\quad h_n'(\epsilon) &= \frac{dh_n}{d\epsilon} = \frac{\ln 2}{\epsilon}(1-\mathbb{E}_{\underline{l}}[\langle m\rangle]),\\ \text{BIAWGNC:}\quad h_n'(\epsilon) &= \frac{dh_n}{d\epsilon} = \epsilon^{-3}(1-\mathbb{E}_{\underline{l}}[\langle m\rangle]).\end{aligned}$$

Theorem 3.2 and (3.35) imply $\lim_{n\to\infty}\mathbb{E}_{\mathcal{C}}|h_n' - \mathbb{E}_{\mathcal{C}}[h_n']| = 0$.

Absence of replica symmetry breaking. Some random spin systems are believed to exhibit the phenomenon of *replica symmetry breaking* [71], [74]. In the present context this would mean that in the large block-length limit the posterior measure becomes a convex combination of large number (exponential in the block-length) of measures. A criterion that signals such a phenomenon can be formulated thanks to the overlap parameter

$$q = \frac{1}{n}\sum_{i=1}^{n} x_i^{(1)}x_i^{(2)}, \tag{3.36}$$

where the superscript is the usual replica index and means that we take two independent samples $(x_1^{(a)},...,x_n^{(a)})$ from $p(\underline{x}\mid\underline{y})$. When the replica symmetry is broken the overlap parameter has a non trivial distribution [71]. Let $\mathbb{P}_m(x) = \mathbb{E}_{\mathcal{C},\underline{l}}\langle \mathbb{1}(m=x)\rangle$, $\mathbb{P}_q(x) = \mathbb{E}_{\mathcal{C},\underline{l}}\langle \mathbb{1}(q=x)\rangle_{1,2}$, where $\langle\cdot\rangle_{1,2}$ denotes the product measure $p(\underline{x}^{(1)}\mid\underline{y})\cdot p(\underline{x}^{(2)}\mid\underline{y})$ for the same output realization of $\underline{y}$. For any BMS channel and any linear block code, we have the Nishimori identity [65]

$$\mathbb{P}_m(x) = \mathbb{P}_q(x). \tag{3.37}$$

Thus, concentration of m is equivalent to concentration of q so that the distribution of the overlap parameter remains a trivial delta function. This argument further strengthens the belief that the replica solution for communication over BMS channels using sparse graph codes is correct[4].

[4]Note the replica solution, given in Section 1.7.4, is really the *replica symmetric solution* in the language of statistical physics.

3.A Proof of Lemma 3.2 and Lemma 3.3

We begin with the proof of Lemma 3.2.

Recall that p_{BP} is the BP fixed-point of the density evolution equation for the interpolated code at $t = 1$ (the original Poisson-LDPC code ensemble). I.e., p_{BP} is the largest fixed-point of

$$x = \epsilon\lambda(1 - \rho(1 - x)). \tag{3.38}$$

For the proof we will require the following. Let $\delta' = \epsilon\lambda(1 - \rho(1 - p_{BP} - \delta)) - p_{BP}$. Since p_{BP} is the largest fixed point of the equation we must have $\epsilon\lambda(1 - \rho(1 - p_{BP} - \delta)) < p_{BP} + \delta$. As a result,[5],

$$\delta' < \delta. \tag{3.39}$$

In the sequel, x and $x_{\ell,t}$ will denote erasure probabilities. It should not be confused with the spin/code-bit variable x_i.

Proof. We first claim that for any x such that $p_{BP} + \delta \leq x \leq 1$, we have

$$\frac{d}{dt}\Big[\epsilon\lambda_{1-t}(1 - \rho(1 - p_{BP} - \delta))\lambda_t(1 - \rho(1 - x))\Big] \geq 0. \tag{3.40}$$

Let us compute the derivative (3.40) explicitly. We have

$$\begin{aligned}
&\frac{d}{dt}\Big[\epsilon\lambda_{1-t}(1 - \rho(1 - p_{BP} - \delta))\lambda_t(1 - \rho(1 - x))\Big] \\
&\qquad = \frac{d}{dt}\Big[\epsilon e^{-\gamma(1-t)\rho(1-p_{BP}-\delta)} e^{-\gamma t\rho(1-x)}\Big] \\
&\qquad = \Big[\epsilon e^{-\gamma(1-t)\rho(1-p_{BP}-\delta)} e^{-\gamma t\rho(1-x)} \gamma\Big(\rho(1 - p_{BP} - \delta) - \rho(1 - x)\Big)\Big].
\end{aligned}$$

Since $\rho(\cdot)$ is monotonic in its argument, for $x \geq p_{BP} + \delta$ we have $\rho(1 - x) \leq \rho(1 - p_{BP} - \delta)$, from which we immediately prove the claim. As a consequence of the above claim we have

$$\begin{aligned}
\epsilon\lambda_{1-t}(1 - \rho(1 - p_{BP} - \delta))\lambda_t(1 - \rho(1 - x)) \leq \epsilon\lambda(1 - \rho(1 - x)) \\
\forall\, x \in [p_{BP} + \delta, 1].
\end{aligned} \tag{3.41}$$

Recall that we denote the interpolated code at time t by C_t. Our aim is to find a positive integer ℓ such that, the average erasure estimate of a random code-bit after ℓ iterations of the BP decoder is less than $p_{BP} + \delta$.

From the density evolution equations at time t we have,

$$x_{\ell+1,t} = \epsilon\lambda_{1-t}(1 - \rho(1 - p))\lambda_t(1 - \rho(1 - x_{\ell,t})).$$

Above, $x_{\ell+1}$ denotes the density evolution estimate of the probability that the variable-node-to-check-node message is in erasure at iteration ℓ.

[5]Note that δ is a small number. Thus, $p_{BP} + \delta < \epsilon$.

Let ℓ (which depends only on δ) be the number of iterations such that $x_{\ell,1} \geq p_{BP} + \delta$ and $x_{\ell+1,1} < p_{BP} + \delta$. We claim that $x_{\ell+1,t} < p_{BP} + \delta$ for all t. Indeed, without loss of generality, consider only those values of $t < 1$, for which $x_{\ell,t} \geq p_{BP} + \delta$. We claim that for such t we have $x_{\ell,t} \leq x_{\ell,1}$. Let us prove this sub-claim immediately. Indeed, we have that $x_{0,t} = \epsilon$ for all t. Thus, at $\ell = 0$, we have $x_{0,t} \leq x_{0,1}$. Now suppose that for some $0 < \ell' < \ell$ we have $x_{\ell',t} \leq x_{\ell',1}$. Since $\ell' < \ell$, we must have $x_{\ell',t} \geq x_{\ell,t}$ (density evolution estimate of the erasure fraction decreases with increasing iterations). Thus $x_{\ell',t} \geq p_{BP} + \delta$. As a result, using (3.41) we have,

$$\begin{aligned} x_{\ell'+1,t} &= \epsilon\lambda_{1-t}(1-\rho(1-p_{BP}-\delta))\lambda_t(1-\rho(1-x_{\ell',t})) \\ &\overset{(3.41)}{\leq} \epsilon\lambda(1-\rho(1-x_{\ell',t})) \end{aligned} \tag{3.42}$$

Since by assumption $x_{\ell',t} \leq x_{\ell',1}$, we get

$$\epsilon\lambda(1-\rho(1-x_{\ell',t})) \leq \epsilon\lambda(1-\rho(1-x_{\ell',1})) = x_{\ell'+1,1}. \tag{3.43}$$

Combining above, we obtain $x_{\ell'+1,t} \leq x_{\ell'+1,1}$. Thus, we deduce that, starting from $\ell = 0$, as long as $x_{\ell',t} \geq p_{BP} + \delta$, we must have $x_{\ell',t} \leq x_{\ell,1}$. Thus, we prove that $x_{\ell,t} \leq x_{\ell,1}$.

Now since $x_{\ell,t} \leq x_{\ell,1}$, we obtain

$$\begin{aligned} x_{\ell+1,t} &= \epsilon\lambda_{1-t}(1-\rho(1-p_{BP}-\delta))\lambda_t(1-\rho(1-x_{\ell,t})) \\ &\leq \epsilon\lambda(1-\rho(1-x_{\ell,t})) \leq \epsilon\lambda(1-\rho(1-x_{\ell,1})) \\ &= x_{\ell+1,1} < p_{BP} + \delta. \end{aligned}$$

Where the first inequality again follows from (3.41), since $x_{\ell,t} \geq p_{BP} + \delta$. The second inequality follows since $x_{\ell,t} \leq x_{\ell,1}$. Hence we have $x_{\ell+1,t} < p_{BP} + \delta$.

Since $x_{\ell+1,t} < p_{BP} + \delta$, from the density evolution equation at time t, we have

$$\begin{aligned} x_{\ell+2,t} &= \epsilon\lambda_{1-t}(1-\rho(1-p_{BP}-\delta))\lambda_t(1-\rho(1-x_{\ell+1,t})) \\ &\leq \epsilon\lambda_{1-t}(1-\rho(1-p_{BP}-\delta))\lambda_t(1-\rho(1-p_{BP}-\delta)) \\ &= \epsilon\lambda(1-\rho(1-p_{BP}-\delta)) = p_{BP} + \delta'. \end{aligned}$$

Recall that we denote the neighborhood of depth ℓ of a random code-bit (spin) x_i in C_t by $N_\ell(i)$. Let us compute $\mathbb{E}[\langle x_i\rangle_{t,T_{\ell+2}(i)} | N_{\ell+2}(i) \text{ is tree}]$. Recall that $T_{\ell+2}(i)$ is the computation tree of depth $\ell + 2$. In this case the computation tree coincides with $N_{\ell+2}(i)$. Thus, $\langle x_i\rangle_{t,T_{\ell+2}} = \langle x_i\rangle_{t,N_{\ell+2}(i)}$. Recall that $\langle x_i\rangle_{t,N_{\ell+2}(i)} \in \{0,1\}$ (c.f. 2.5). This implies that $\mathbb{E}[\langle x_i\rangle_{t,T_{\ell+2}(i)} | N_{\ell+2}(i) \text{ is tree}]$ equals the probability of not-erasure. From density evolution, we know that this equals $1 - \epsilon\lambda_{1-t}(1-\rho(1-p_{BP}-\delta))\Lambda_t(1-\rho(1-x_{\ell+2,t}))$. Note that, since the interpolated code is also a Poisson-LDPC code, we can replace $\Lambda_t(1-\rho(1-x_{\ell+2,t}))$ with $\lambda_t(1-\rho(1-x_{\ell+2,t}))$ in the previous expression.

Thus, using $\epsilon\lambda_{1-t}(1-\rho(1-p_{BP}-\delta))\lambda_t(1-\rho(1-x_{\ell+2,t})) \leq x_{\ell+2,t}$ we get

$$\mathbb{E}[\langle x_i\rangle_{t,T_{\ell+2}(i)}] \geq \mathbb{E}[\langle x_i\rangle_{t,T_{\ell+2}(i)}|N_{\ell+2}(i) \text{ is tree}]\Pr(N_{\ell+2}(i)\text{is tree})$$
$$\geq \bar{x}_{\ell+2,t}\left(1-\frac{(c\ln n)^{4(\ell+2)}}{n}\right) \geq \bar{p}_{BP}-\delta'-\frac{\delta-\delta'}{3}$$

where in the second inequality we use the fact that for a Poisson-LDPC code, $N_{\ell+2}(i)$ is a tree with probability greater than $1-\frac{(c\ln n)^{4(\ell+2)}}{n}$. The last inequality follows by choosing n large enough. We get the lemma by putting $g(\delta) = \frac{2\delta'}{3}+\frac{\delta}{3} < \delta$. □

We now turn to the proof of Lemma 3.3.

Proof. The sub exponential term comes from removing the codes with maximum degree greater than $\ln n$. The exponential term follows by adapting the proof of [9] for maximum degree of $\ln n$. □

3.B Second Interpolation

In this appendix we prove Lemma 3.5 which uses the interpolation method. Notice that with the modified partition function, $Z(t,\nu,\bar{p})$, there is no more channel symmetry (Nishimori symmetry). This makes the determination of the sign of the remainder of the interpolation term a bit more complicated. However by choosing $\nu(t) = \frac{\gamma}{r}(1-t)$, we can easily determine the sign and we obtain the lemma. Using the fundamental theorem of calculus we have

$$\mathbb{E}[\alpha_n(t,\nu(t),\bar{p})-\alpha_n(t,0,\bar{p})] = \mathbb{E}[\alpha_n(0,\frac{\gamma}{r},\bar{p})-\alpha_n(0,0,\bar{p})] + \int_0^t \frac{d\mathbb{E}[\alpha_n(s,\nu(s),\bar{p})-\alpha_n(s,0,\bar{p})]}{ds}ds.$$

Since $\nu(s)$ also depends on s, the total derivative w.r.t. s is given by

$$\frac{d\mathbb{E}[\alpha_n(s,\nu(s),\bar{p})]}{ds} = \frac{\partial\mathbb{E}[\alpha_n(s,\nu,\bar{p})]}{\partial s} - \frac{\gamma}{r}\frac{\partial\mathbb{E}[\alpha_n(s,\nu,\bar{p})]}{\partial\nu}. \tag{3.44}$$

Let us compute the partial derivative with respect to s first. This computation is similar to the one in Appendix 2.A.4. We have

$$\begin{aligned}\frac{\partial\mathbb{E}[\alpha_n(s,\nu,\bar{p})]}{\partial s} &= \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{(-1)^{l+1}}{l}\left(\langle Q_l^r\rangle_{s,\nu} - r\big(\mathbb{E}[\tanh^l(v)]\big)^{r-1}\langle Q_l\rangle_{s,\nu}\right)\\ &\quad+\gamma\ln 2\Big(1-\frac{1}{r}\Big)\\ &= \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{(-1)^{l+1}}{l}\langle R(\bar{p},Q_l)\rangle_{s,\nu} - \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{(-1)^{l+1}}{l}(r-1)\bar{p}^r\\ &\quad+\gamma\ln 2\Big(1-\frac{1}{r}\Big).\end{aligned} \tag{3.45}$$

Above, we used $d_V(v) = p\delta_0(v) + \bar{p}\delta_\infty(v)$ to deduce $\big(\mathbb{E}[\tanh^l(v)]\big)^r = \bar{p}^r$ and $\big(\mathbb{E}[\tanh^l(v)]\big)^{r-1} = \bar{p}^{r-1}$. The partial derivative with respect to ν is easily found out to be,

$$\frac{\partial \mathbb{E}[\alpha_n(s,\nu,\bar{p})]}{\partial \nu} = \mathbb{E}[\langle R(\bar{p},m)\rangle_{s,\nu}]. \tag{3.46}$$

Thus, combining the two partial derivatives we get that the total derivative is equal to

$$\frac{d\mathbb{E}[\alpha_n(s,\nu(s),\bar{p})]}{ds} = \frac{\gamma}{r}\mathbb{E}\sum_{l=2}\frac{(-1)^{l+1}}{l}\langle R(\bar{p},Q_l)\rangle_{s,\nu} - \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{(-1)^{l+1}}{l}(r-1)\bar{p}^r + \gamma\ln 2\Big(1-\frac{1}{r}\Big). \tag{3.47}$$

We remark that the first sum over l in (3.47), starts from $l = 2$ instead of $l = 1$. This is because $\mathbb{E}[\langle R(\bar{p},m)\rangle_{s,\nu}]$ from (3.46) gets canceled with $\mathbb{E}[\langle R(\bar{p},Q_1)\rangle_{s,\nu}] = \mathbb{E}[\langle R(\bar{p},m)\rangle_{s,\nu}]$ in (3.45). Performing similar calculations for $\frac{d\mathbb{E}[\alpha_n(s,0,\bar{p})]}{ds}$ we get,

$$\frac{d\mathbb{E}[\alpha_n(s,0,\bar{p})]}{ds} = \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{(-1)^{l+1}}{l}\langle R(\bar{p},Q_l)\rangle_{s} - \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{(-1)^{l+1}}{l}(r-1)\bar{p}^r + \gamma\ln 2\Big(1-\frac{1}{r}\Big). \tag{3.48}$$

Recall that in $\langle - \rangle_s$, the term ν is set to zero. Thus, combining we get,

$$\mathbb{E}[\alpha_n(t,\nu(t),\bar{p})] - \mathbb{E}[\alpha_n(t,0,\bar{p})] = \mathbb{E}[\alpha_n(0,\frac{\gamma}{r},\bar{p}) - \mathbb{E}[\alpha_n(0,0,\bar{p})] + \int_0^t \frac{\gamma}{r}\mathbb{E}\sum_{l=2}\frac{(-1)^{l+1}}{l}\langle R(\bar{p},Q_l)\rangle_{s,\nu}ds - \int_0^t \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{(-1)^{l+1}}{l}\langle R(\bar{p},Q_l)\rangle_{s}. \tag{3.49}$$

Since in the last term above, the parameter ν is set to zero, we use the channel symmetry (Nishimori identities) to get (c.f. (2.44))

$$\int_0^t \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{(-1)^{l+1}}{l}\langle R(\bar{p},Q_l)\rangle_{s} = \int_0^t \frac{\gamma}{r}\mathbb{E}\sum_{l=1}\frac{1}{2l(2l-1)}\langle R(\bar{p},Q_{2l})\rangle_{s}$$

If r is even, then the above expression is positive from convexity arguments [29]. If r is odd then, recall that in the previous chapter, we showed that this term is positive in the limit $n \to +\infty$. Note that the $o_n(1)$ term in (3.14) comes when r is odd. Hence the lemma.

3.C Proof of Lemma 3.6 and Lemma 3.7

In this appendix we prove Lemma 3.6 and Lemma 3.7. Let us denote the Gibbs measure of $\nu = 0$ system by $\mu(\underline{x})$ and the Gibbs measure of the $\nu \neq 0$ system by $\mu(\nu,\underline{x})$.

Proof of the Lemma 3.6. Since $\langle x_i \rangle_s = 1$ we have

$$\sum_{\underline{x}^{\sim i}} \mu(x_i = -1, \underline{x}^{\sim i}) = 0,$$

which implies that for any value of $\underline{x}^{\sim i}$ we find $\mu(x_i = -1, \underline{x}^{\sim i}) = 0$. We observe that we can write the Gibbs measure of the $\nu \neq 0$ spin system as $\mu(\nu, \underline{x}) = Z(s)\mu(\underline{x})e^{n\nu R(\bar{p},m)}/Z(s,\nu,\bar{p})$. Thus

$$\mu(\nu, x_i = -1, \underline{x}^{\sim i}) = Z(s)\mu(x_i = -1, \underline{x}^{\sim i})e^{n\nu R(\bar{p},m)}/Z(s,\nu,\bar{p}) = 0,$$

for any value of $\underline{x}^{\sim i}$. This immediately implies the lemma. □

Proof of the Lemma 3.7. Let $\langle x_i \rangle_{s,\nu} = -1 + \zeta$. Thus

$$\sum_{\underline{x}^{\sim i}} \mu(\nu, x_i = 1, \underline{x}^{\sim i}) - \sum_{\underline{x}^{\sim i}} \mu(\nu, x_i = -1, \underline{x}^{\sim i}) = -1 + \zeta.$$

Since $\sum_{\underline{x}^{\sim i}} \mu(\nu, x_i = 1, \underline{x}^{\sim i}) = 1 - \sum_{\underline{x}^{\sim i}} \mu(\nu, x_i = -1, \underline{x}^{\sim i})$ and $\mu(\nu, \underline{x}) = Z(s)\mu(\underline{x})e^{n\nu R(\bar{p},m)}/Z(s,\nu,\bar{p})$, we have

$$\sum_{\underline{x}^{\sim i}} \mu(x_i = 1, \underline{x}^{\sim i})e^{\nu n R(\bar{p},m)} = \frac{Z(s,\nu,\bar{p})}{Z(s)}\frac{\zeta}{2}$$

It is easy to see that

$$\frac{Z(s,\nu,\bar{p})}{Z(s)} = \left\langle e^{n\nu R(\bar{p},m)} \right\rangle_s \leq e^{2n\nu r}$$

since $|R(\bar{p},m)| \leq 2r$. Thus, we have

$$\sum_{\underline{x}^{\sim i}} \mu(x_i = 1, \underline{x}^{\sim i})e^{-\nu n|R(\bar{p},m)|} \leq \sum_{\underline{x}^{\sim i}} \mu(x_i = 1, \underline{x}^{\sim i})e^{-\nu n|R(\bar{p},m)|} \leq e^{\nu n 2r}\frac{\zeta}{2}.$$

As a result, we have $\sum_{\underline{x}^{\sim i}} \mu(x_i = 1, \underline{x}^{\sim i}) \leq e^{\nu n 4r}\frac{\zeta}{2}$. This implies $\langle x_i \rangle_s = 2\sum_{\underline{x}^{\sim i}} \mu(x_i = 1, \underline{x}^{\sim i}) - 1 \leq e^{\nu n 4r}\zeta - 1$. From GKS inequality for the BEC (and any code) we know that $\langle x_i \rangle_s \geq 0$, which implies $\zeta \geq e^{-\nu n 4r}$. □

3.D Proof of Lemma 3.10

We prove here Lemma 3.10. The Conjecture 3.1 implies that

$$\begin{aligned}\mathcal{R}(t) &= \frac{\gamma}{r}\sum_{l\geq 1}\frac{(-1)^{l+1}}{l}R(q_l, \mathbb{E}_t[\langle Q_l \rangle_t]) + o_n(1),\\ &= \frac{\gamma}{r}\sum_{l\geq 1}\frac{(-1)^{l+1}}{l}R(q_l, \mathbb{E}_t[\langle x_i \rangle_t^l]) + o_n(1).\end{aligned}$$

Proof. Choose a positive integer, $L = \left\lceil \frac{1}{2}\left(1 + \frac{1}{\delta^{\frac{1}{4}}}\right) \right\rceil$ for any $0 < \delta \ll 1$[6]. Since $|R(q_l, \mathbb{E}[\langle x_i \rangle_t^{2l}])| \leq 2r$, we have

$$\begin{aligned}\mathcal{R}(t) &\leq \frac{\gamma}{r} \sum_{l=1}^{L-1} \frac{1}{2l(2l-1)} R(q_{2l-1}, \mathbb{E}[\langle x_i \rangle_t^{2l-1}]) + \frac{\gamma}{r} \sum_{l \geq L} \frac{2r}{2l(2l-1)} \\ &\leq \frac{\gamma}{r} \sum_{l=1}^{L-1} \frac{1}{2l(2l-1)} R(q_{2l-1}, \mathbb{E}[\langle x_i \rangle_t^{2l-1}]) + 2\gamma\delta^{\frac{1}{4}}. \end{aligned} \tag{3.50}$$

where the last inequality follows from

$$\sum_{l \geq L} \frac{1}{2l(2l-1)} \leq \frac{1}{2L-1} \leq \delta^{\frac{1}{4}}. \tag{3.51}$$

Note that we use $\mathbb{E}$ to denote the expectation w.r.t. to randomness in both code and noise realizations.

We now proceed to control the sum till $L-1$. To do this we first develop a *sum-rule* to relate $\mathbb{E}[\langle x_i \rangle_t^{2l-1}]$ and q_{2l-1}. Remarkably, thanks to the channel symmetry, we will see that the difference between the two terms is "almost" the same for $l \geq 1$. Consider a fixed code C. Consider the code-bit x_i and consider a neighborhood $N_d(i)$, of depth d, around the spin. We "soften" any hard parity-check node a belonging to the boundary $\mathring{N}_d(i)$ of the neighborhood as follows. Associate a random variable $l_a \sim \mathcal{N}(t_a, t_a)$ (symmetric gaussian) to the parity-check node a, then we can write,

$$\frac{1 + x_a}{2} = \lim_{t_a \to +\infty} e^{l_a(x_a - 1)}. \tag{3.52}$$

Consider the modified Hamiltonian with every check node, $a \in \mathring{N}_d(1)$ represented by $e^{l_a(x_a-1)}$. We remark that if $t_a = 0$ for any check node a, then the "interaction" between the code-bits present in the check node a vanishes. This process of setting $t_a = 0$ we call as "erasing" the check node from the graph. With slight abuse of notation, $\underline{l}$ now denotes the random realization of the channel noise as well as the random variables l_a associated to each check node.

We now provide a sum rule for estimating $\mathbb{E}_{\underline{l}}[\langle x_i^m \rangle_t]$ for any positive integer m. Order the check nodes in the boundary, $\mathring{N}_d(i)$, of the neighborhood, in a given arbitrary way and call $\langle - \rangle_{t;\leq k}$ the Gibbs measure with $t_k = 0$ for the first k check nodes of $\mathring{N}_d(i)$. With this notation, $\langle - \rangle_{t,\leq 0}$ implies that no check node has been erased. For the first check node (call it 1) we use the fundamental theorem of calculus to find,

$$\mathbb{E}_{\underline{l}}[\langle x_i \rangle_t^m] = \mathbb{E}_{\underline{l}}[\langle x_i \rangle_{t;\leq 1}^m] + \int_{t_1=0}^{\infty} dt_1 \frac{d\mathbb{E}_{\underline{l},l_1}[\langle x_i \rangle_t^m]}{dt_1}. \tag{3.53}$$

[6]Here δ is a very small real number.

We can now take the second check node (call it 2) and erase it from the first term in the r.h.s of the above equation to get,

$$\mathbb{E}_{\underline{l}}[\langle x_i\rangle_t^m] = \mathbb{E}_{\underline{l}}[\langle x_i\rangle_{t;\leq 2}^m] + \sum_{j=1}^{2}\int_{t_j=0}^{+\infty} dt_j \frac{d\mathbb{E}_{\underline{l}}[\langle x_i\rangle_{t;\leq j-1}^m]}{dt_j}. \tag{3.54}$$

We repeat this procedure for all check nodes in $\mathring{N}_d(i)$ to get,

$$\mathbb{E}_{\underline{l}}[\langle x_i\rangle_t^m] = \mathbb{E}_{\underline{l}}[\langle x_i\rangle_{t,N_d(i)}^m] + \sum_{j\in\mathring{N}_d(i)}\int_{t_j=0}^{\infty} dt_j \frac{d\mathbb{E}_{\underline{l}}[\langle x_i\rangle_{t;\leq j-1}^m]}{dt_j}. \tag{3.55}$$

The bracket, $\langle -\rangle_{t,N_d(i)}$ denotes the Gibbs measure w.r.t. the code with all the check nodes in the boundary $\mathring{N}_d(i)$ erased, or in other words, it corresponds to an average w.r.t. the neighborhood $N_d(i)$.

Consider any code with one check node converted to a soft check node. More precisely consider the Hamiltonian with the constraint $e^{l_c(x_c-1)}$. Because of the Gaussian identity and integration by parts formula (see 2.A.3) we have

$$\frac{d\mathbb{E}_{\underline{l}}[\langle x_i\rangle_t^m]}{dt_c} = \mathbb{E}_{\underline{l}}\left[\Big(\frac{d}{dl_c} + \frac{1}{2}\frac{d^2}{dl_c}\Big)\langle x_i\rangle_t^m\right]. \tag{3.56}$$

The first derivative equals

$$\frac{d\mathbb{E}_{\underline{l}}\langle x_i\rangle_t^m}{dl_c} = m\mathbb{E}_{\underline{l}}\Big[\langle x_i\rangle_t^{m-1}\big(\langle x_i x_c\rangle_t - \langle x_i\rangle_t\langle x_c\rangle_t\big)\Big]. \tag{3.57}$$

Taking now the derivative of the above expression we obtain the second derivative

$$\begin{aligned}\frac{1}{2}\left[\frac{d^2\mathbb{E}_{\underline{l}}\langle x_i\rangle_t^m}{dl_c}\right] &= \frac{1}{2}m(m-1)\mathbb{E}_{\underline{l}}\Big[\langle x_i\rangle_t^{m-2}\big(\langle x_i x_c\rangle_t - \langle x_i\rangle_t\langle x_c\rangle_t\big)^2\Big] \\ &\quad + \frac{1}{2}m\mathbb{E}_{\underline{l}}\Big[\langle x_i\rangle_t^{m-1}\big(-2\langle x_i x_c\rangle_t\langle x_c\rangle_t + 2\langle x_i\rangle_t\langle x_c\rangle_t^2\big)\Big].\end{aligned} \tag{3.58}$$

Combining the above two we get the derivative w.r.t. t_c to be

$$\begin{aligned}\frac{d\mathbb{E}_{\underline{l}}[\langle x_i\rangle_t^m]}{dt_c} &= m\mathbb{E}_{\underline{l}}\Big[\langle x_i\rangle_t^{m-1}\big(\langle x_i x_c\rangle_t - \langle x_i\rangle_t\langle x_c\rangle_t\big)(1-\langle x_c\rangle_t)\Big] \\ &\quad + \frac{1}{2}m(m-1)\mathbb{E}_{\underline{l}}\Big[\langle x_i\rangle_t^{m-2}\big(\langle x_i x_c\rangle_t - \langle x_i\rangle_t\langle x_c\rangle_t\big)^2\Big].\end{aligned} \tag{3.59}$$

We apply Nishimori identities (see 2.A.2) to all the terms above and for the case when $m = 2l-1$ we get

$$\frac{d\mathbb{E}_{\underline{l}}[\langle x_i\rangle_t^{2l-1}]}{dt_c} = \frac{1}{2}2l(2l-1)\mathbb{E}_{\underline{l}}\Big[\langle x_i\rangle_t^{2l-2}\big(\langle x_i x_c\rangle_t - \langle x_i\rangle_t\langle x_c\rangle_t\big)^2\Big]. \tag{3.60}$$

Using this in our sum rule for $m = 2l-1$ we get

$$\mathbb{E}_{\underline{l}}[\langle x_i\rangle_t^{2l-1}] = \mathbb{E}_{\underline{l}}[\langle x_i\rangle_{t,N_d(i)}^{2l-1}] + \frac{1}{2}2l(2l-1)\sum_{j\in\mathring{N}_d(i)}\int_{t_j=0}^{\infty} dt_j \times \mathbb{E}_{\underline{l},\le j-1}\Big[\langle x_i\rangle_{t,\le j-1}^{2l-2}\big(\langle x_ix_j\rangle_{t,\le j-1} - \langle x_i\rangle_{t,\le j-1}\langle x_j\rangle_{t,\le j-1}\big)^2\Big]. \quad (3.61)$$

Thus, taking expectation over the code ensemble we get

$$\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_t^{2l-1}] = \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_{t,N_d(i)}^{2l-1}] + \frac{1}{2}2l(2l-1)\mathsf{Corr}(l), \quad (3.62)$$

where

$$\mathsf{Corr}(l) = \mathbb{E}_{\mathcal{C}}\sum_{j\in\mathring{N}_d(i)}\int_{t_j=0}^{\infty} dt_j \times \mathbb{E}_{\underline{l},\le j-1}\Big[\langle x_i\rangle_{t,\le j-1}^{2l-2}\big(\langle x_ix_j\rangle_{t,\le j-1} - \langle x_i\rangle_{t,\le j-1}\langle x_j\rangle_{t,\le j-1}\big)^2\Big]. \quad (3.63)$$

We remark that since $\langle x_i\rangle_{t,\le j-1}^{2l-2} \le 1$, we have the inequality $\mathsf{Corr}(l) \le \mathsf{Corr}(1)$. Since the probability of the event, $\mathsf{T} = N_d(i)$ is a tree is greater than $1-o_n(1)$, for $l = 1$ we get

$$\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_t] = \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_{t,N_d(i)} \mid \mathsf{T}] + \mathsf{Corr}(1) + o_n(1). \quad (3.64)$$

Without loss of generality assume that the $q_1 > 0$. Choose d_1 such that $|\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_{t,N_{d_1}(i)} \mid \mathsf{T}] - q_1| < \delta/(2r(r-1))$. This is possible because of the hypothesis of the lemma, which states that q_1 is the BP fixed-point of DE.

Recall that $R(a,b) = (r-1)a^r - ra^{r-1}b + b^r$. If $b \ge a$, then we have

$$\begin{aligned} R(a,b) &= (r-1)a^r - ra^{r-1}b + b^r \\ &= (b-a)[b^{r-1} + \cdots + a^{r-2}b + a^{r-1} - ra^{r-1}] \\ &\ge (b-a)(a^{r-2}b - a^{r-1}) = (b-a)^2a^{r-2}. \end{aligned} \quad (3.65)$$

Using $a = q_1 - \delta/(2r(r-1))$ and $b = \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_t]$ in (3.65) we find

$$R\Big(q_1 - \frac{\delta}{2r(r-1)}\,,\ \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_t]\Big) \ge \Big(\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_t] - \Big(q_1 - \frac{\delta}{2r(r-1)}\Big)\Big)^2\Big(q_1 - \frac{\delta}{2r(r-1)}\Big)^{r-2}. \quad (3.66)$$

Using $\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_{t,N_{d_1}(i)} \mid \mathsf{T}] > q_1 - \delta/(2r(r-1))$ in the r.h.s of (3.66) and using (3.64) we obtain

$$R\Big(q_1 - \frac{\delta}{2r(r-1)}\,,\ \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i\rangle_t]\Big) \ge (\mathsf{Corr}(1))^2\Big(q_1 - \frac{\delta}{2r(r-1)}\Big)^{r-2} + o_n(1). \quad (3.67)$$

Using the identities $a^r - b^r = (a-b)(a^{r-1} + a^{r-2}b + \cdots + b^{r-1})$ and $|R(a,b) - R(a+\delta', b)| \leq 2r(r-1)|\delta'|$ for $|a|, |a+\delta'|, |b|$ all less than 1, we find

$$\begin{aligned} R(a,b) &= R\Big(a+\delta', b\Big) + R(a,b) - R\Big(a+\delta', b\Big) \\ &\leq R\Big(a+\delta', b\Big) + 2r(r-1)|\delta'|. \end{aligned} \tag{3.68}$$

We remark that the above inequality holds even if δ' is negative.

Thus, using 3.68 with $a = q_1 - \frac{\delta}{2r(r-1)}$, $b = \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t]$ and $\delta' = \frac{\delta}{2r(r-1)}$, we find

$$R\bigg(q_1 - \frac{\delta}{2r(r-1)}\ ,\ \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t]\bigg) \leq R(q_1, \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t]) + \delta. \tag{3.69}$$

Combining (3.67) and (3.69) we get

$$(\mathsf{Corr}(1))^2\Big(q_1 - \frac{\delta}{2r(r-1)}\Big)^{r-2} \leq R(q_1, \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t]) + \delta + o_n(1). \tag{3.70}$$

Again, the hypothesis of the theorem allows us to choose d_2, n large enough such that $\Big|\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_{t,N_{d_2}(i)}^{2l-1} \mid \mathsf{T}] - q_{2l-1}\Big| < \frac{\delta}{2r(r-1)}$, for all $1 \leq l \leq L-1$. Finally, choose d to be the larger of d_1 and d_2.

Using (3.68) for $a = q_{2l-1}$, $b = \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t^{2l-1}]$ and $a + \delta' = \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_{t,N_{d_2}(i)}^{2l-1} \mid \mathsf{T}]$ we get

$$R\big(q_{2l-1}, \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t^{2l-1}]\big) \leq R\big(\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_{t,N_{d_2}(i)}^{2l-1} \mid \mathsf{T}]\ ,\ \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t^{2l-1}]\big) + \delta.$$

Putting everything together we have

$$\begin{aligned} &\frac{1}{2l(2l-1)} \int_{t=0}^{1} dt R(q_{2l-1}, \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t^{2l-1}) \\ &\leq \int_{t=0}^{1} dt R\big(\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_{t,N_{d_2}(i)}^{2l-1} \mid \mathsf{T}]\ ,\ \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t^{2l-1}\big) + \delta \\ &\overset{R(a,b)\leq|b-a|2r \text{ and } (3.62)}{\leq} 2r \int_{t=0}^{1} dt\ \mathsf{Corr}(l) + \delta + o_n(1) \\ &\overset{\text{remark after } (3.63)}{\leq} 2r \int_{t=0}^{1} dt\ \mathsf{Corr}(1) + \delta + o_n(1) \\ &\overset{\text{Cauchy-Schwartz}}{\leq} 2r\Big(\int_{t=0}^{1} dt (\mathsf{Corr}(1))^2\Big)^{1/2} + \delta + o_n(1) \\ &\overset{(3.70)}{\leq} \frac{2r}{(q_1-\delta)^{\frac{r-2}{2}}}\Big(\int_{t=0}^{1} dt R\big(q_1, \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_t]\big)\Big)^{1/2} + \frac{2r\delta^{1/2}}{(q_1-\delta)^{\frac{r-2}{2}}} + \delta + o_n(1) \\ &\overset{\text{Hyp. of theorem}}{\leq} \frac{4r}{(q_1-\delta)^{\frac{r-2}{2}}}\delta^{1/2} + \delta + o_n(1). \end{aligned}$$

Finally, summing over all $l \leq L-1$ and using $L \leq \frac{2}{\delta^{1/4}}$ we obtain

$$\int_0^1 dt \mathcal{R}(t) \leq \frac{4\gamma}{(q_1-\delta)^{\frac{r-2}{2}}}\delta^{1/4} + \frac{2\gamma}{r}\delta^{3/4} + o_n(1).$$

□

It will be useful in the next chapter to have the following lemma.

Lemma 3.11 (Erasing Checks)**.** *Consider transmission over any general BMS channel using any fixed LDPC code C. Then we have, for any positive integer $p \geq 1$,*

$$\mathbb{E}_{\underline{l}}\big[\langle x_i \rangle^{2p}\big] \geq \mathbb{E}_{\underline{l}}\big[\langle x_i \rangle_{C\setminus B}^{2p}\big],$$

where x_i is any code-bit and the average $\langle - \rangle_{C\setminus B}$ is w.r.t. the Gibbs measure obtained by removing a set of parity-check nodes B from the code C.

Proof. This follows immediately from the analysis in the previous proof, by noticing that the derivative (or the remainder term) when we "erase" any check node (see (3.53)) is always positive (see (3.60)). □

3.E Derivation of (3.15)

Due to the continuity of $\Delta_n(\frac{\gamma}{r}, \bar{p})$ with respect to $\bar{p}$, we can bound the difference

$$|\Delta_n(\frac{\gamma}{r}, \bar{p} - \delta) - \Delta_n(\frac{\gamma}{r}, \bar{p})| \leq O(\delta). \tag{3.71}$$

Since δ can be made as small as wanted, it is sufficient to show that $\Delta_n(\frac{\gamma}{r}, \bar{p})$ goes to zero in the limit $n \to +\infty$, for $p = p_{RS}$.

Computation of $\Delta_n(.,.)$ can be done exactly because it does not involve any code constraints. From (3.15) we have

$$\begin{aligned}
\Delta_n(\frac{\gamma}{r}, \bar{p}) &= \frac{1}{n}\mathbb{E}\left[\ln Z(0, \frac{\gamma}{r}, \bar{p}) - \ln Z(0, 0, \bar{p})\right] \\
&= \frac{1}{n}\mathbb{E}\left[\ln \left\langle e^{\frac{\gamma}{r} n(m^r - r\bar{p}^{r-1}m + (r-1)\bar{p}^r)} \right\rangle_0\right] \\
&= \frac{\gamma}{r}(r-1)\bar{p}^r + \frac{1}{n}\mathbb{E}\left[\ln \left\langle e^{\frac{\gamma}{r} n(m^r - r\bar{p}^{r-1}m)} \right\rangle_0\right].
\end{aligned} \tag{3.72}$$

Above, $\langle - \rangle_0$ is the Gibbs average associated to the partition function $Z(0, 0, \bar{p})$. Note that $\langle - \rangle_0$ has no parity-check constraints. Let us now compute the second term in (3.72).

$$\begin{aligned}
\frac{1}{n}\mathbb{E}\ln\langle e^{\frac{\gamma}{r} n(m^r - r\bar{p}^{r-1}m)} \rangle_0 &= \frac{1}{n}\mathbb{E}\ln \sum_{\underline{x}} e^{n\frac{\gamma}{r}m^r} \prod_{i=1}^{n} \left(\frac{e^{(\frac{l_i}{2} + \sum_{a=1}^{e_i} \frac{u_a^i}{2} - \frac{\gamma}{r} r\bar{p}^{r-1})x_i}}{2\cosh(\frac{l_i}{2}) \prod_{a=1}^{e_i} 2\cosh(\frac{u_a^i}{2})} \right) \\
&\quad - \frac{1}{n}\mathbb{E}\ln \sum_{\underline{x}} \prod_{i=1}^{n} \left(\frac{e^{(\frac{l_i}{2} + \sum_{a=1}^{e_i} \frac{u_a^i}{2})x_i}}{2\cosh(\frac{l_i}{2}) \prod_{a=1}^{e_i} 2\cosh(\frac{u_a^i}{2})} \right) \\
&= \frac{1}{n}\mathbb{E}\ln \sum_{\underline{x}} e^{n\frac{\gamma}{r}m^r + \sum_{i=1}^{n} J_i x_i} - \mathbb{E}\ln\left(2\cosh\left(\frac{l_i}{2} + \sum_{a=1}^{e_i} \frac{u_a^i}{2}\right)\right).
\end{aligned}$$

Where we used

$$\frac{1}{n}\mathbb{E}\ln\sum_{\underline{x}}\prod_{i=1}^{n}e^{(\frac{l_i}{2}+\sum_{a=1}^{e_i}\frac{u_a^i}{2})x_i}=\mathbb{E}\ln\Big(2\cosh\Big(\frac{l_i}{2}+\sum_{a=1}^{e_i}\frac{u_a^i}{2}\Big)\Big).$$

We also introduce the notation $J_i=\frac{l_i}{2}+\sum_{a=1}^{e_i}\frac{u_a^i}{2}-r\frac{\gamma}{r}\bar{p}^{r-1}$.

Thus, setting $p=p_{RS}$, we finally obtain

$$\Delta_n(\frac{\gamma}{r},\bar{p}_{RS})=\frac{\gamma}{r}(r-1)\bar{p}_{RS}^r+\frac{1}{n}\mathbb{E}\ln\sum_{\underline{x}}e^{n\frac{\gamma}{r}m^r+\sum_{i=1}^n J_ix_i}-\mathbb{E}\ln 2\cosh(\frac{l_i}{2}+\sum_{a=1}^{e_i}\frac{u_a^i}{2}). \tag{3.73}$$

All the spins have independent and identically distributed J_i's. The spins are coupled only through the term m^r ($r-$ body interactions) which appears in $R(x,m)$. The following lemma of [96] helps us to convert the $r-$body interactions into $2-$body interactions.

Lemma 3.12 (S. Korada, N. Macris [96])**.** *For some* $0<\theta<1$*,*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\ln\sum_{\underline{x}}e^{n\frac{\gamma}{r}m^r+\sum_{i=1}^n J_ix_i}=\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\ln\sum_{\underline{x}}e^{\sum_{i=1}^n J_ix_i}\int_{-1}^{1}dze^{-n^{1+\theta}\left(z-\frac{\sum x_i}{n}\right)^2+n\frac{\gamma}{r}z^r}. \tag{3.74}$$

The square term in the integral in the above lemma can be linearized using the gaussian identity $\sqrt{\pi}e^{-n^{1+\theta}\left(z-\frac{\sum x_i}{n}\right)^2}=\int dye^{-y^2-2in^{\frac{1+\theta}{2}}\left(z-\frac{\sum x_i}{n}\right)y}$. With this, there are only linear terms in the exponent[7]. As a result, the sum over the spins, $\underline{x}$, can be performed. Then the integral over z is estimated performing saddle point computations [96]. Thus, combining (3.73) and (3.71), leads us to the estimate

$$\Delta_n\big(\frac{\gamma}{r},\bar{p}_\delta\big)=\frac{\gamma}{r}(r-1)\bar{p}_{RS}^r+\max_z f(z)+O(\delta)+o_n(1).$$

[7]Note that there will be two integrals in the r.h.s. of (3.74). One over y (obtained via the gaussian identity) and the one over z. But the integrals can be exchanged (see [96] for details).

Decay of Correlations: Low Signal-to-Noise Ratio

4

4.1 Introduction

The study of the behavior of correlations between spins is one of the central aims of statistical physics [98], [99], [100]. The analysis of the correlations gives an insight into whether there is a single Gibbs measure describing the system in the thermodynamic limit or if the measure of the system is a convex combination of many Gibbs measures (which signals a phase transition in the system [99], [101], [102], [103]).

In this chapter we will show that a good deal can be learned by looking at the correlations between code-bits (more precisely the covariance of the MAP decoder), averaged over the channel outputs. Intuitively, studying the correlations should tell us how much "information" one code-bit conveys about another. As we will show, this has an immediate bearing on the performance of the BP decoder. We will have much more to say about the methods we use, but let us say at the outset that our aim is to cover a fairly general class of BMS channels, including the BSC and BIAWGNC.

One of our main results in this chapter is that, for sufficiently high noise, the correlations between two code-bits of an LDGM code and a class of LDPC codes decay exponentially fast as a function of the graph distance between these code-bits. Further the convergence is uniform in the block-length n. The sparsity of the underlying graph then implies that, if furthermore the decay rate beats the local expansion of the graph, the asymptotic average MAP-GEXIT curve can be computed via density evolution. Another interpretation of this result is that the solutions provided by the replica method of spin glass theory are exact. For LDGM codes transmitted over the BSC, the decay of correlations has another consequence. It implies that the limits $d, n \to +\infty$ can be exchanged for the computation of the asymptotic average BP-GEXIT

curve.

For lattice spin systems (e.g. the Ising model) an important criterion that ensures correlation decay is the Dobrushin's criterion [104], [99], [62]. The second most common method is based on suitable expansions in powers of "the strength of interactions". There exists a host of such expansions collectively called "cluster expansions". In the context of spin systems the historically first and the simplest such expansion is the so-called "polymer expansion" [98], [105].

The main rule of thumb is that all these methods work if the spins/code-bits are weakly interacting. In the language of coding: the methods would work only for bounded LLRs. However, when we consider transmission over the BIAWGNC, the LLRs are unbounded. As a result, these methods are not suitable for the analysis. It turns out that a cluster expansion of Dreifus, Klein and Perez can be carried out for the case of transmission over the BIAWGNC. This method was first developed in [101], in the context of lattice systems. We adapt their expansion to our case of Tanner graphs and demonstrate decay of correlations.

The rest of the chapter is organized as follows. In Section 4.2 we formulate the channel models and the encoding schemes. In Section 4.3 we state our main results for LDGM as well as LDPC codes. The main strategy of the proofs is also explained there. Section 4.4 contains the proof of correlation decay and its consequences: computing the asymptotic average MAP-GEXIT curve via density evolution and exchanging limits of large block-length and iteration number of computing the average BP-GEXIT. The proof of decay of correlations for the special class of LDPC code ensemble is provided in Section 4.5. We conclude the chapter in Section 4.6 by some discussion and by pointing out some open problems. Appendix 4.A derives the cluster expansion, in the Tanner graph setting, which is used in our proofs.

4.2 Set-Up – Channels and Codes

In this section we introduce the channel models and the codes for which we will state and prove our main results. We first describe a fairly general class of BMS channels. This includes the important case of the BIAWGNC. We will then describe the codes that we use for transmission, namely LDGM codes and a special class of LDPC codes.

4.2.1 Low SNR Channel Models

Consider a BMS(ϵ) channel defined by a transition p.d.f. $p_{Y|X}(y \mid x)$ with inputs $x \in \{-1, +1\}$ and outputs belonging to $\bar{\mathbb{R}}$. *In this chapter, it will be convenient to replace the channel* LLR *with the half-loglikelihood* ($\frac{1}{2}$LLR)

$$l = \frac{1}{2} \ln \left[\frac{p_{Y|X}(y| + 1)}{p_{Y|X}(y| - 1)} \right], \tag{4.1}$$

which, by abuse of notation, we also denote by l.

Also, we denote the p.d.f. of the $\frac{1}{2}\mathsf{LLR}$, assuming that the all-one codeword was transmitted, by $c(l)$. Recall from Section 1.2 that, the noise parameter varies in an interval $[0, \epsilon_{\max}]$ ($\epsilon_{\max}$ is possibly infinite). For example, we have, $\epsilon_{\max} = \frac{1}{2}$ for the BSC, $\epsilon_{\max} = +\infty$ for the BIAWGNC and $\epsilon_{\max} = 1$ for the BEC. Note that $\epsilon \to 0$ corresponds to low noise and $\epsilon \to \epsilon_{\max}$ corresponds to high noise. We now define the class of channels for which our main results hold.

Definition 4.1 (Class of Channels – $\mathcal{K}$)**.** *A channel belongs to the class $\mathcal{K}$ if it satisfies:*

1. *The numbers $T_{2p}(\epsilon) = \frac{d}{d\epsilon}\int_{-\infty}^{\infty} dl\, c(l)(\tanh l)^{2p}$ are uniformly bounded with respect to the positive integer p ($p \geq 1$).*

2. *For any finite $m > 0$ we have $\mathbb{E}[e^{m|l|}] \leq c_m < +\infty$.*

3. *(High noise condition): Set $\delta(\epsilon, H) = e^{4H} - 1 + \mathbb{P}(|l| > H)$. One can find $H(\epsilon) \geq 0$ such that $\lim_{\epsilon \to \epsilon_{\max}} \delta(\epsilon, H(\epsilon)) = 0$, where $\mathbb{P}(\cdot)$ is the c.d.f. (associated to $c(l)$).*

Let us give some examples of channels which belong to the class $\mathcal{K}$.

Example 4.1 (BSC)**.** *Consider the BSC(ϵ) with $\epsilon \in (0, \frac{1}{2})$. We have*

$$p_{Y|X}(y|x) = (1-\epsilon)\delta(y-x) + \epsilon\delta(y+x),$$
$$c(l) = (1-\epsilon)\delta\left(l - \frac{1}{2}\ln\left[\frac{1-\epsilon}{\epsilon}\right]\right) + \epsilon\delta\left((l - \frac{1}{2}\ln\left[\frac{\epsilon}{1-\epsilon}\right]\right). \qquad (4.2)$$

From above we have, $(\tanh l)^{2p} = (1-2\epsilon)^{2p}$ for all values of l. As a consequence, we obtain

$$T_{2p}(\epsilon) = \frac{d}{d\epsilon}[(1-2\epsilon)^{2p}] = -4p(1-2\epsilon)^{2p-1}.$$

It is easy to see that, for any $0 < \epsilon < \frac{1}{2}$, one can find a constant, which depends only on ϵ, which upper bounds T_{2p} for every $p \geq 1$. Since $\mathbb{E}[e^{m|l|}] = (\frac{1-\epsilon}{\epsilon})^{\frac{m}{2}}$, the second condition is also met. Let us turn our attention to the last condition. If we choose $H(\epsilon) = \frac{1}{2}\ln\frac{1-\epsilon}{\epsilon}$ for $\epsilon \in (0, 1/2)$, we find from the distribution $c(l)$ that $\mathbb{P}(|l| > H) = 0$. As a result, $\delta(\epsilon, H) = e^{2\ln\left[\frac{1-\epsilon}{\epsilon}\right]} - 1$, and $\lim_{\epsilon \to \frac{1}{2}} \delta(\epsilon, H) = 0$. Let us say a few more words on the high noise condition. The high noise condition states that we can make the term $\delta(\epsilon, H)$ as small as we want, by making the channel sufficiently bad ($\epsilon \to \epsilon_{\max}$). The factor $\delta(\epsilon, H)$ appears in the analysis of the correlation between code-bits. By choosing the channel noise large enough we gain control over its estimate.

Example 4.2 (BIAWGNC)**.** *The most important example is the BIAWGNC. Let us show that BIAWGNC* $\in \mathcal{K}$*. We have*

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\epsilon}\exp\left(-\frac{(y-x)^2}{2\epsilon^2}\right),\ c(l) = \frac{\epsilon}{\sqrt{2\pi}}\exp\left(-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}\right),$$
$$\frac{dc(l)}{d\epsilon} = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}\right) + \frac{\epsilon}{\sqrt{2\pi}}\exp\left(-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}\right)(\epsilon^{-3} - \epsilon l^2). \quad (4.3)$$

Thus we have

$$\begin{aligned}
|T_{2p}| &\leq \int_{-\infty}^{+\infty} dl \left|\frac{dc(l)}{d\epsilon}\right| \\
&\leq \epsilon^{-1}\int_{-\infty}^{+\infty} dl \frac{\epsilon}{\sqrt{2\pi}}\exp\left(-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}\right) + \epsilon^{-3}\int_{-\infty}^{+\infty} dl \frac{\epsilon}{\sqrt{2\pi}}\exp\left(-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}\right) \\
&\quad + \epsilon\int_{-\infty}^{+\infty} dl \frac{\epsilon}{\sqrt{2\pi}}\exp\left(-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}\right) l^2 \\
&= 2(\epsilon^{-1} + \epsilon^{-3}).
\end{aligned}$$

This implies a uniform bound on T_{2p} *with respect to* p*. Since* $\mathbb{E}[e^{m|l|}] \leq \mathbb{E}[e^{ml}] + \mathbb{E}[e^{-ml}]$*, we use the characteristic function for the Gaussian distribution to get* $\mathbb{E}[e^{m|l|}] \leq e^{\epsilon^{-2}(m+\frac{m^2}{2})} + e^{\epsilon^{-2}(-m+\frac{m^2}{2})} = c_m < \infty$*. Let us test the high noise condition. Choose* $H(\epsilon) = \epsilon^{-3/4}$*. Clearly* $\lim_{\epsilon\to+\infty} e^{4H(\epsilon)} - 1 = e^{4\epsilon^{-3/4}} - 1 = 0$*. Using the definition of the* $Q-$*function,* $Q(\alpha) = \frac{1}{\sqrt{2\pi}}\int_{\alpha}^{+\infty} e^{-x^2/2}dx$ *and the bound* $Q(x) \leq \frac{1}{2}e^{-x^2/2}$*, for* $x \geq 0$*, we find,*

$$\begin{aligned}
\mathbb{P}(|l| > H) &= \frac{\epsilon}{\sqrt{2\pi}}\int_{-\infty}^{H} e^{-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}} dl + \frac{\epsilon}{\sqrt{2\pi}}\int_{H}^{+\infty} e^{-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}} dl \\
&\leq \frac{2\epsilon}{\sqrt{2\pi}}\int_{H}^{+\infty} e^{-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}} dl \\
&= \frac{2}{\sqrt{2\pi}}\int_{\frac{H-\epsilon^{-2}}{\sqrt{\epsilon^{-2}}}}^{+\infty} e^{-\frac{x^2}{2}} dx \\
&= 2Q\Big(\frac{H-\epsilon^{-2}}{\sqrt{\epsilon^{-2}}}\Big) \leq e^{-\frac{(H-\epsilon^{-2})^2}{2\epsilon^{-2}}} \\
&= e^{-\epsilon^{1/2}(1-\epsilon^{-5/4})^2/2}.
\end{aligned}$$

From the above we conclude that $\lim_{\epsilon\to+\infty}\delta(\epsilon, H) = 0$*.*

The case of unbounded likelihoods turns out to be significantly more difficult than that of bounded ones. For channels, such as the BSC, we could use the simpler polymer expansion or the Dobrushin theory [99], [62]. But since we want to also cover the BIAWGNC, we directly use the more powerful cluster expansion.

Note that the BEC is not contained in the class $\mathcal{K}$. The BEC violates the second condition. Nevertheless, due to the special nature of this channel our methods can easily be adapted. We show this in Section 4.4.1.

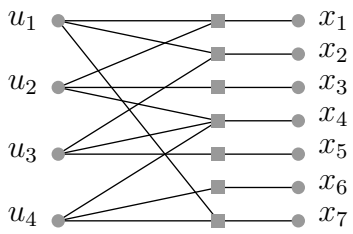

Figure 4.1: The variable nodes on the left represent the information bits and the variable nodes on the right represent the code-bits. The information bits are denoted by u_1, u_2, u_3 and u_4. The code-bits are denoted by x_1, x_2, x_3, x_4, x_5, x_6 and x_7.

4.2.2 Codes

We consider transmission over channels belonging to the class $\mathcal{K}$ using two types of codes: LDGM codes and a special class of LDPC codes.

Recall that LDGM codes are constructed from a fixed bipartite graph with $n + m$ variable nodes and n check nodes. Edges only connect variable and check nodes. Refer to Figure 4.1 for an illustration. The information bits $u_1, ..., u_m \in \{-1, +1\}^m$ are represented by the variable nodes on the left. The code-bits, $x_1, ..., x_n$ are represented by the variable nodes on the right. The check nodes are numbered in the same manner as the code-bits. Thus, code-bit i and check node i are fungible. The code-bit x_i is obtained by $x_i = \prod_{a \in \partial i} u_a \triangleq u_{\partial i}$, $i = 1, ..., n$. Note that here ∂i denotes the neighborhood of information bits of the check node i. The design rate of the code $\mathtt{R} = \frac{m}{n}$ is kept fixed.

In the sequel, we will consider the Tanner graph without the variable nodes on the right (representing the code-bits). See Section 4.4 for details. In this modified Tanner graph, the check node i has only information bits, u_a, as its neighbors. Recall that the set of neighbors of a variable node a is denoted by ∂a. Also, ∂i denotes the neighborhood (of information bits) of the check node i, in the modified Tanner graph. We consider graphs with *bounded node degrees* $|\partial a| \leq l_{\max}$ and $|\partial i| \leq k_{\max}$. As usual, we use $\Lambda(x)$ and $P(x)$ to denote the variable and check node degree distributions.

We also consider a special class of LDPC code ensembles, which we define as follows.

Definition 4.2 (Class of Ensemble of LDPC Codes – $\mathfrak{K}$)**.** *The ensemble of LDPC codes $\mathfrak{K}$ is defined as follows. Consider a code from the standard ensemble LDPC$(n, \Lambda(x), P(x))$ with design rate $\mathtt{R} = 1 - \frac{\Lambda'(1)}{P'(1)}$ and bounded maximum variable and check node degrees. We denote the set of variable nodes of this code by $\mathfrak{V}$. Denote the set of check nodes by $\mathfrak{C}$. To every check node belonging to $\mathfrak{C}$, attach a degree-one variable node with probability $\mathfrak{p}$. Call the set of new degree-one variable nodes $\mathfrak{V}_1$. The set of check nodes which contain a new degree-one variable node is $\mathfrak{C}_1$ and the remaining ones are denoted by*

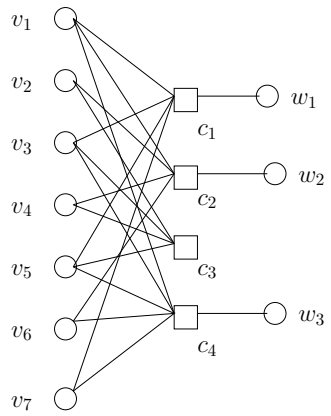

Figure 4.2: A code from the ensemble $\mathfrak{K}$ is constructed from a standard LDPC graph, with 7 variable nodes and 4 check nodes denoted by $v_1, v_2, \ldots, v_7$ and c_1, c_2, c_3, c_4, respectively. To each check node in the underlying LDPC code, we attach a degree-one variable node with probability $\mathfrak{p}$. In the figure, w_1, w_2 and w_3 denote the degree-one variable nodes belonging to the set $\mathfrak{V}_1$. The check nodes c_1, c_2 and c_4 form the set $\mathfrak{C}_1$.

$\mathfrak{C}_0 = \mathfrak{C} \setminus \mathfrak{C}_1$. *On an average the resulting Tanner graph has* $n + m\mathfrak{p}$ *variable nodes. The design rate of the code is* $\mathfrak{R} = \frac{1-(1-R)(1-\mathfrak{p})}{1+\mathfrak{p}(1-R)}$. *See Figure 4.2 for an illustration.*

Note that this LDPC code construction has two kinds of randomness. We denote by $\mathbb{E}_{\mathcal{C}}$, the average with respect to the randomness in the underlying code ensemble LDPC(Λ, P). We denote by $\underline{S} = \{S_1, \ldots, S_m\}$ the vector of Bernoulli$(\mathfrak{p})$ random variables. We associate S_a to each check node a where the value $S_a = 1$ indicates that $a \in \mathfrak{C}_1$ and $S_a = 0$ indicates that $a \in \mathfrak{C}_0$. We will denote the average w.r.t. to the ensemble $\mathfrak{K}$ by $\mathbb{E}_{\mathcal{C},\underline{S}} \triangleq \mathbb{E}_{\mathcal{C}}\mathbb{E}_{\underline{S}}$.

Recall that we use the Gibbs bracket $\langle - \rangle$ to denote the averages with respect to the a posteriori (MAP) measure when transmitting over a BMS channel using any of the above codes.

4.3 Main Theorem on Correlation Decay

Before we set out to state our main theorems on the decay of correlations for the two types of codes mentioned above, we define the notion of self-avoiding random walks and the graph distance.

Definition 4.3 (Self-Avoiding Walk)**.** *Consider a Tanner graph (representing either an LDGM code or an LDPC code). A self-avoiding walk* w *between two variable nodes* a *and* b*, belonging to the Tanner graph, is a sequence* $v_1, c_1, v_2, c_2, \ldots, c_l, v_{l+1}$ *of variable nodes (denoted* $v_1, v_2, \ldots, v_{l+1}$*) and check nodes (denoted* $c_1, c_2, \ldots, c_l$*), such that* $v_1 = a, v_{l+1} = b$*,* $\{v_m, v_{m+1}\} \in \partial c_m$*, and* $v_m \neq v_n, c_m \neq c_n$ *for* $m \neq n$*.*

Thus on a self-avoiding walk we do not repeat variable and check nodes. The length of the walk, denoted by $|w|$, is the number of *edges* present in it. If $a = b$ then the self-avoiding walk from a to b is the trivial walk a. The length of such a walk is zero.

Clearly, two variable nodes a and b are connected on the Tanner graph, if and only if there exists a self-avoiding walk from a to b. Also, it is not hard to see that from any general walk between a and b, we can extract a self-avoiding walk w, between a and b, which has all its check nodes belonging to the parent walk (this is done by chopping off all the loops of the general walk).

Definition 4.4 (Graph Distance – $\mathrm{dist}(A, B)$)**.** *Consider two non-empty sets of variable nodes A and B. Let W_{ab} denote the set of all self-avoiding walks between variable nodes a and b and let $W_{AB} = \cup_{a\in A, b\in B} W_{ab}$. We denote the graph distance between A and B by $dist(A, B)$ and set it to*

$$dist(A, B) = \min_{w\in W_{AB}} |w|.$$

Recall that $\mathbb{E}_{\underline{l}}$ denotes the expectation w.r.t. the vector of $\frac{1}{2}\mathsf{LLR}$ realizations $\underline{l}$. Also, recall that $\langle -\rangle$ denotes the average w.r.t. the Gibbs (MAP) measure.

Theorem 4.1 (Decay of Correlations for the MAP Decoder and LDGM Codes)**.** *Consider communication over a channel from the class $\mathcal{K}$ at high enough noise, i.e. $\epsilon_g < \epsilon < \epsilon_{\max}$, where $\epsilon_g > 0$ depends only on $l_{\max}$ and $k_{\max}$. Consider transmission using a fixed LDGM code. Then*

$$\mathbb{E}_{\underline{l}}\big[|\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|\big] \le c_1 e^{-\frac{\mathrm{dist}(\partial i,\partial j)}{\xi(\epsilon)}}, \tag{4.4}$$

where i and j are any two code-bits, c_1 is a positive constant, and $\xi(\epsilon)$ is a strictly positive constant depending only on ϵ, $l_{\max}$ and $k_{\max}$.

Theorem 4.2 (Decay of Correlations for the MAP Decoder and the Ensemble of Codes $\mathfrak{K}$)**.** *Consider communication over a channel from the class $\mathcal{K}$. Consider transmission using a code picked from the ensemble $\mathfrak{K}$ of LDPC codes with large enough fraction $\mathfrak{p}$ of the extra degree-one variable nodes, $0 < \mathfrak{p}_0 < \mathfrak{p} < 1$. Here $\mathfrak{p}_0$ depends only on $\mathfrak{l}_{max}$ and $\mathfrak{r}_{max}$. Let the channel noise ϵ be high enough, i.e. $\epsilon_g < \epsilon < \epsilon_{\max}$, where $\epsilon_g > 0$ depends only on $\mathfrak{p}_0$, $l_{\max}$ and $k_{\max}$. Then for i and j belonging to $\mathfrak{V}$ (c.f. Definition 4.2)*

$$\mathbb{E}_{\underline{S},\underline{l}}\big[|\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|\big] \le c_1 e^{-\frac{\mathrm{dist}(i,j)}{\xi(\epsilon)}}, \tag{4.5}$$

where c_1 is a positive constant and $\xi(\epsilon)$ is a strictly positive constant depending only on ϵ, $l_{\max}$, $k_{\max}$ and $\mathfrak{p}$. If $i \in \mathfrak{V}_1$ and $j \notin \mathfrak{V}_1$ ($j \in \mathfrak{V}_1$) (c.f. Definition 4.2), then we replace $dist(i, j)$ by $dist(\partial i, j)$ ($dist(\partial i, \partial j)$) in (4.5). If both i and j belong to $\mathfrak{V}_1$ then the theorem holds with $dist(i, j)$ replaced by $dist(\partial i, \partial j)$.

We first remark that in the above theorem, the "underlying" code from the ensemble LDPC($\Lambda(x), P(x)$) (c.f. Definition 4.2) remains the same. The randomness is only in the noise realizations and the addition of the extra degree-one variable nodes.

In the statement of Theorem 4.2, when $i \in \mathfrak{V}_1$, then ∂i is the set of variable nodes of the underlying LDPC code, which belong to the corresponding check node. Also, notice that in the above theorem if either $i \in \mathfrak{V}_1$ ($j \in \mathfrak{V}_1$), then the expectation in (4.5) should be replaced with $\mathbb{E}_{\underline{S}^{\sim i},\underline{l}}$ ($\mathbb{E}_{\underline{S}^{\sim j},\underline{l}}$), where $\underline{S}^{\sim i}$ ($\underline{S}^{\sim j}$) denotes the expectation w.r.t. all $\underline{S}$ except S_i (S_j). If both i and j belong to the set $\mathfrak{V}_1$, then the expectation in (4.5) should be replaced with $\mathbb{E}_{\underline{S}^{\sim ij},\underline{l}}$ where $\underline{S}^{\sim ij}$ denotes the expectation w.r.t. all $\underline{S}$ except S_i and S_j. We also remark that $\xi^{-1}(\epsilon)$ grows with $\epsilon \to \epsilon_{\max}$ and also as $\mathfrak{p} \to 1$ for the case of the ensemble $\mathfrak{K}$. We furnish the proofs of these theorems in Sections 4.4 and 4.5.

4.3.1 Consequences of Decay of Correlations

We first describe the BP decoder from the point of view of Gibbs measures. Given a graph G defining a given LDGM or LDPC code with the block-length n fixed (large), we choose a code-bit i. Construct the computation tree $T_d(i)$ of depth d (even) [23]. This is the universal covering tree truncated at distance d from node i. We label the variable/check nodes of this tree with labels denoted by ν. Let $\pi : T_d(i) \to G$ be the projection from the covering tree to the original graph. A node $\nu \in T_d(i)$ has an image $\pi(\nu)$, and due to the loops in G this projection is a many to one map: one may have $\pi(\nu) = \pi(\nu')$ for $\nu \neq \nu'$. Now, consider a tree-code defined in the usual way on the tree-graph $T_d(i)$. One can view the BP decoder ($d-$ iterations) for node x_i as a MAP decoder for this tree-code. In other words the BP decoder uses the Gibbs measure on $T_d(i)$: one crucial point is that for this Gibbs measure, the $\frac{1}{2}\mathsf{LLR}$ attached to the nodes are no longer independent.

LDGM: Recall from Section 1.7 that the Gibbs measure for LDGM codes is given by

$$p_{\underline{U}|\underline{Y}}(\underline{u} \mid \underline{y}) = \frac{1}{Z} \prod_{i=1}^{n} e^{l_i u_{\partial i}}. \tag{4.6}$$

For the LDGM case the measure on the computation tree, $T_d(i)$, is

$$\frac{1}{Z_{T_d(i)}} \prod_{k \in T_d(i)} e^{l_{\pi(k)} u_{\partial k}}, \tag{4.7}$$

where $Z_{T_d(i)}$ is the proper normalization factor. We call $\langle - \rangle_d^{BP}$ the Gibbs bracket with respect to this measure. The extrinsic BP soft estimate for the code-bit i is $\langle x_i \rangle_{0,d}^{BP}$. Recall that the average (with respect to the ensemble)

MAP-GEXIT function is given by

$$g_n(\epsilon) = \mathbb{E}_{\mathcal{C}}\left[\int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim i}} \ln\left\{\frac{1+\langle x_i\rangle_0 \tanh l_i}{1+\tanh l_i}\right\}\right]. \tag{4.8}$$

Recall that $\langle - \rangle_0$ denotes the (extrinsic) Gibbs average with $l_i = 0$. Also, recall from (1.18) that we can write the average (with respect to the ensemble) BP-GEXIT function, for LDGM case as

$$g^{BP}_{n,d}(\epsilon) = \mathbb{E}_{\mathcal{C}}\left[\int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim i}} \ln\left\{\frac{1+\langle x_i\rangle^{BP}_{0,d} \tanh l_i}{1+\tanh l_i}\right\}\right]. \tag{4.9}$$

Notice that we have abused the notation to denote the average BP-GEXIT curve also by $g^{BP}_{n,d}(\epsilon)$. The soft code-bit estimate $\langle x_i\rangle^{BP}_d$ can be computed exactly by summing the spins starting from the leaves of $T_d(i)$ all the way up to the root i. This computation is left to the reader and yields the usual message passing BP algorithm.

Example 4.3 (MAP-GEXIT for LDGM over BEC)**.** *Figure 4.3 shows the asymptotic average (over code ensemble) MAP-GEXIT curve for the case of transmission over the BEC using the ensemble LDGM*$(n, \Lambda(x) = x^3, P(x) = 0.3x + 0.7x^3)$*. We also show the asymptotic average BP-GEXIT. Notice that there is a jump (discontinuity) in the asymptotic average MAP-GEXIT curve as well as in the asymptotic average BP-GEXIT curve. Away from the jumps the two curves match.*

One of the consequences of our main result on decay of correlations enables us to compute the asymptotic average MAP-GEXIT curve via density evolution, for large enough erasure fractions (see Corollary 4.1). This can be seen form the figure, where for large erasure fractions the asymptotic average MAP-GEXIT and BP-GEXIT match.

Codes from $\mathfrak{K}$*:* Consider now the Gibbs measure associated to a code in the ensemble $\mathfrak{K}$. Under MAP decoding the a posteriori probability distribution of the input to the channel $\underline{X} = (X_1, \ldots, X_N)$ given the output $\underline{Y} = (Y_1, \ldots, Y_N)$ (where $N = |\mathfrak{V} \cup \mathfrak{V}_1|$ is the block-length of the code) is

$$p(\underline{X}|\underline{Y}) = \frac{1}{Z}\prod_{a\in\mathfrak{C}_0}\frac{1}{2}(1+x_{\partial a})\prod_{a\in\mathfrak{C}_1}\frac{1}{2}(1+x_{\partial a})\prod_{i\in\mathfrak{V}}e^{l_i x_i}\prod_{i\in\mathfrak{V}_1}e^{l_i x_i},$$

with

$$Z = \sum_{\underline{x}}\prod_{a\in\mathfrak{C}_0}\frac{1}{2}(1+x_{\partial a})\prod_{a\in\mathfrak{C}_1}\frac{1}{2}(1+x_{\partial a})\prod_{i\in\mathfrak{V}}e^{l_i x_i}\prod_{i\in\mathfrak{V}_1}e^{l_i x_i}.$$

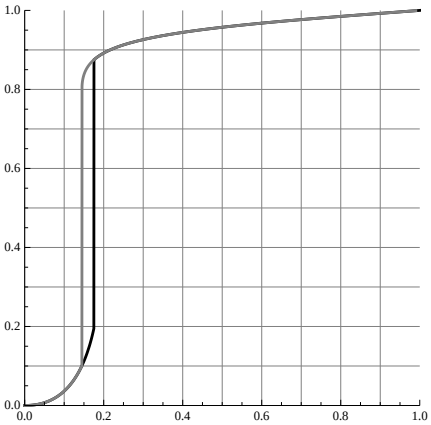

Figure 4.3: This figure illustrates the asymptotic average MAP-GEXIT curve (dark curve) for the case of transmission over a BEC using the ensemble LDGM$(n, \Lambda(x) = x^3, P(x) = 0.3x + 0.7x^3)$. We also show the asymptotic average BP-GEXIT curve (gray curve). Note that away from the jumps, the curves match.

As usual, $h_N = \frac{1}{N} H(\underline{X} \mid \underline{Y})$. Recall from Section 1.5 that the MAP-GEXIT is given by

$$\begin{aligned}\frac{dh_N}{d\epsilon} =& \frac{1}{N} \sum_{i \in \mathfrak{V}} \int dl_i \frac{dc(l_i)}{d\epsilon} \mathbb{E}_{\underline{l}^{\sim i}} \ln\left\{ \frac{1 + \langle x_i \rangle_0 \tanh l_i}{1 + \tanh l_i} \right\} \\ &+ \frac{1}{N} \sum_{a \in \mathfrak{V}_1} \int dl_a \frac{dc(l_a)}{d\epsilon} \mathbb{E}_{\underline{l}^{\sim a}} \ln\left\{ \frac{1 + \langle x_a \rangle_0 \tanh l_a}{1 + \tanh l_a} \right\}.\end{aligned}$$

Note that $\underline{l} = (l_1, \ldots, l_{|\mathfrak{V} \cup \mathfrak{V}_1|})$ is the vector of $\frac{1}{2}$LLR realizations.

For channels belonging to class $\mathcal{K}$ one can show that $\int dl_i \frac{dc(l_i)}{d\epsilon} \mathbb{E}_{\underline{l}^{\sim i}} \ln\{\frac{1+\langle x_i \rangle_0 \tanh l_i}{1+\tanh l_i}\}$ and $\int dl_a \frac{dc(l_a)}{d\epsilon} \mathbb{E}_{\underline{l}^{\sim a}} \ln\{\frac{1+\langle x_a \rangle_0 \tanh l_a}{1+\tanh l_a}\}$ are bounded. Taking the average w.r.t. the code ensemble $\mathbb{E}_{\mathcal{C},\underline{S}}$, using the symmetry of sites and tail bounds for the Binomial$(m, \mathfrak{p})$ distribution, one can show that for channels belonging to the class $\mathcal{K}$ we have

$$\begin{aligned}\frac{d}{d\epsilon} \mathbb{E}_{\mathcal{C},\underline{S}}[h_N] =& \frac{1}{(1 + (1 - \mathtt{R})\mathfrak{p})} \mathbb{E}_{\mathcal{C},\underline{S}} \left[\int dl_i \frac{dc(l_i)}{d\epsilon} \mathbb{E}_{\underline{l}^{\sim i}} \ln\left\{ \frac{1 + \langle x_i \rangle_0 \tanh l_i}{1 + \tanh l_i} \right\} \right] \\ &+ \frac{(1 - \mathtt{R})\mathfrak{p}}{(1 + (1 - \mathtt{R})\mathfrak{p})} \mathbb{E}_{\mathcal{C},\underline{S}} \left[\int dl_a \frac{dc(l_a)}{d\epsilon} \mathbb{E}_{\underline{l}^{\sim a}} \ln\left\{ \frac{1 + \langle x_a \rangle_0 \tanh l_a}{1 + \tanh l_a} \right\} \right] + o_n(1).\end{aligned} \tag{4.10}$$

Above, x_i denotes a random variable node belonging to the set $\mathfrak{V}$ and x_a denotes a random variable node in the set $\mathfrak{V}_1$. Recall that $\langle x_i \rangle_0 (\langle x_a \rangle_0)$ denotes the (extrinsic) Gibbs average with $l_i = 0 (l_a = 0)$. Again, we get $g^{BP}_{n,d}(\epsilon)$ by replacing the two soft bit-MAP estimates, $\langle x_i \rangle_0$ and $\langle x_a \rangle_0$ by $\langle x_i \rangle^{BP}_{0,d}$ and $\langle x_a \rangle^{BP}_{0,d}$

respectively. The Gibbs measure $\langle x_i \rangle_{0,d}^{BP}$ is obtained from the computation tree, $T_d(i)$, of depth d around the code-bit i in the set $\mathfrak{V}$. Similarly, the Gibbs measure $\langle x_a \rangle_{0,d}^{BP}$ is obtained from the computation tree $T_d(a)$ of depth d around the code-bit $a \in \mathfrak{V}_1$.

Example 4.4 (MAP-GEXIT for Ensemble $\mathfrak{K}$ over BEC)**.** *As an illustration, Figure 4.4 shows the asymptotic average (over the code ensemble) MAP-GEXIT curve for LDPC*$(n, \Lambda(x) = x^3, P(x) = x^{15}, \mathfrak{p} = 0.9)$ *on the BEC. The probability of attaching a degree-one variable node is set to* $\mathfrak{p} = 0.9$*. The asymptotic average MAP-GEXIT curve has been sketched using the Maxwell's construction of [106] on the EBP-GEXIT curve given in the parametric form by*

$$\Big(\epsilon(x), \frac{1}{(1+(1-R)\mathfrak{p})}\Lambda\big(1-(1-\epsilon(x)\mathfrak{p})\rho(1-x)\big)+\frac{(1-R)\mathfrak{p}}{(1+(1-R)\mathfrak{p})}\big(1-P(1-x)\big)\Big),$$

where $\epsilon(x)$ *is the unique positive real solution of the density evolution equation (for the erasure fraction emanating from a variable node in the set* $\mathfrak{V}$*) given by* $x = \epsilon(1-(1-\epsilon\mathfrak{p})(1-x)^{14})^2$.

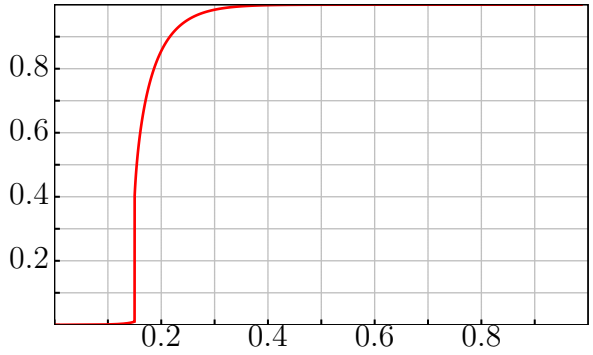

Figure 4.4: This figure illustrates the asymptotic average MAP-GEXIT curve for the case of transmission over a BEC with LDPC$(n, \Lambda(x) = x^3, P(x) = x^{15}, \mathfrak{p} = 0.9)$ code ensemble as defined in Section 4.2.2.

The following two corollaries say that the asymptotic average MAP-GEXIT function can be computed by the density evolution equations in the high noise regime. They also show that the replica expressions computed at the appropriate fixed-point are exact.

Corollary 4.1 (Density Evolution Allows to Compute the Asymptotic Average MAP-GEXIT for Ensemble of LDGM Codes)**.** *Consider communication over a channel belonging to the class* $\mathcal{K}$ *at high enough noise, i.e.* $\epsilon'_g < \epsilon < \epsilon_{\max}$ *where* ϵ'_g *depends only on* $l_{\max}$ *and* $k_{\max}$*. Then for the LDGM*(Λ, P) *ensemble we have*

$$\lim_{n\to+\infty} g_n(\epsilon) = \lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon). \tag{4.11}$$

Corollary 4.2 (Density Evolution Allows to Compute the Asymptotic Average MAP-GEXIT for Ensemble $\mathfrak{K}$)**.** *Consider communication over a channel from the class $\mathcal{K}$. Consider transmission using the ensemble $\mathfrak{K}$ of LDPC codes with large enough fraction $\mathfrak{p}$ of the extra degree-one variable nodes, $0 < \mathfrak{p}_0 < \mathfrak{p} < 1$. Here $\mathfrak{p}_0$ depends only on $\mathfrak{l}_{max}$ and $\mathfrak{r}_{max}$. Let the channel noise ϵ be high enough, i.e. $\epsilon'_g < \epsilon < \epsilon_{\max}$, where $\epsilon'_g > 0$ depends only on $\mathfrak{p}_0$, $l_{\max}$ and $k_{\max}$. Then we have*

$$\lim_{n\to+\infty} g_n(\epsilon) = \lim_{d\to+\infty} \lim_{n\to+\infty} g^{BP}_{n,d}(\epsilon). \tag{4.12}$$

The above corollaries extend the results previously known for the BEC [23], [106], to the class of channels $\mathcal{K}$. *Note that the result applies irrespective whether or not there is a phase transition (e.g a jump discontinuity in the GEXIT curves): so it applies even in situations where the area theorem [36], [107] does not allow to prove* (4.12). The values obtained for ϵ'_g are in general worse than those ϵ_g obtained in Theorem 4.1. One reason for this is that in our proof technique to show the exactness of the BP estimate we use the decay of correlations. We obtain the result only if the decay (of correlations) is fast enough to beat the expansion of the graph.

Finally, concerning the exchange of limits $d, n \to +\infty$ for the BP algorithm we prove

Corollary 4.3 (Exchange of Limits for Computing Average BP-GEXIT)**.** *Consider transmission over the BSC(ϵ) using LDGM(Λ, P) ensembles with bounded degrees. For high enough noise, i.e. $\epsilon''_g < \epsilon < \epsilon_{\max}$, with ϵ''_g depending only on $l_{\max}, k_{\max}$, we have*

$$\lim_{d\to+\infty} \lim_{n\to+\infty} g^{BP}_{n,d}(\epsilon) = \lim_{n\to+\infty} \limsup_{d\to+\infty} g^{BP}_{n,d}(\epsilon) = \lim_{n\to+\infty} \liminf_{d\to+\infty} g^{BP}_{n,d}(\epsilon). \tag{4.13}$$

The proof is a simple application of the decay of correlations. We present it only for the BSC. But it can also be extended to cover any convex combination of such channels. More generally, as long as $c(l)$ has bounded support that diminishes as the noise parameter increases ($\epsilon \to \epsilon_{\max}$) the present result applies. The present result complements the recent work [37] which concerns the bit-error-rate of LDPC codes for message-passing decoders in the regime where the error rate vanishes (low noise regime).

4.3.2 Proof Strategy

In this section we outline the strategy to prove the above statements. As explained in the introduction, for LDGM at high noise and for channels with bounded LLR , (4.5) follows from Dobrushin's criterion or from the polymer expansion [98], [99], [62], [108]. These techniques however do not work when the LLR are unbounded. This is because, roughly speaking, overlapping interactions involve moments $\mathbb{E}[l^m]$ which can spoil the convergence as $m \to +\infty$.

More physically, what happens is that even in the high noise regime there always exists, with positive probability, large portions of the graph that are at low noise (or "low temperature"). This might result in large correlations between the nodes.

Therefore, we use instead a cluster expansion of Dreifus, Klein and Perez [101], developed first in the context of lattices, and adapt it to the Tanner graph setting. This overcomes the "overlapping of moments" problem by organizing the expansion over *self-avoiding random walks* on the graph. Since the walks are self-avoiding, the moment problem does not occur and we can treat unbounded loglikelihoods.

For LDPC codes the situation is more subtle because of the hard parity-check constraints. The hard parity-check constraints give an inherently "low temperature" flavor to the problem. For the moment we have no approach for dealing with general LDPC ensembles. But for codes in the ensemble $\mathfrak{K}$, the addition of the extra degree-one variable nodes helps create a system which is a mixture of LDPC type hard check nodes and LDGM type soft check nodes. If there is a sufficiently large number of soft check nodes, then we can again provide an estimate of the average correlation via the cluster expansion. We will see in the concluding remarks of this chapter that this ensemble provides a possible approach towards proving our results for a more general class of LDPC codes.

4.4 Proof of Main Theorem for LDGM Codes

In this section we prove Theorem 4.1 and Corollary 4.1. Let us recall here the Gibbs measure for LDGM codes. We have

$$p_{\underline{U}|\underline{Y}}(\underline{u} \mid \underline{y}) = \frac{1}{Z} \prod_{i=1}^{n} e^{l_i u_{\partial i}}. \tag{4.14}$$

The code-bit x_i, which is transmitted over the channel, equals $u_{\partial i}$. Thus from (4.14) we see that it will be convenient to think of the LDGM Tanner graph as consisting of m variable nodes and n check nodes. The variable nodes, as usual, correspond to the m information-bits $u_1, \ldots, u_m$. Check node i, has the function $e^{l_i u_{\partial i}}$ associated to it, and corresponds to the code-bit i which is transmitted over the channel. So in this modified Tanner graph, the variable nodes corresponding to the code-bits are not present. Refer to Figure 4.5 for an illustration.

It is convenient to set $K = l_{\max} k_{\max}$.

Proof of Theorem 4.1, LDGM. Refer to the Tanner graph of the LDGM code as described just above (c.f. Figure 4.5). Let W_{ab} denote the set of all self-avoiding walks between variable nodes a and b and let $W_{AB} = \cup_{a \in A, b \in B} W_{ab}$ (see Figure 4.6). In words, W_{AB} is the set of all self-avoiding walks between any variable node in A and any variable node in B. Fix some number $H > 0$.

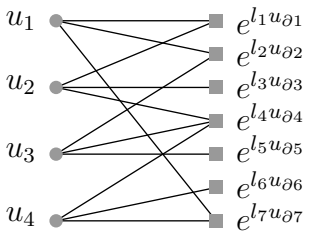

Figure 4.5: The figure illustrates the modified LDGM Tanner graph with variable nodes representing the information-bits, u_1, u_2, u_3 and u_4. The number of check nodes is equal to the block-length. Each check node i has associated to it the function $e^{l_i u_{\partial i}}$. Here, l_i is the LLR value for the code-bit i.

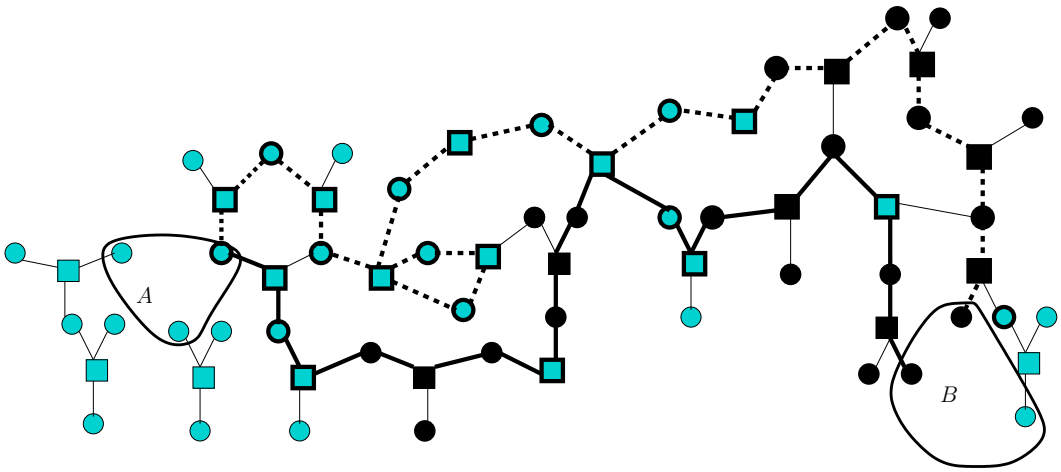

Figure 4.6: The sets A and B each contain three variable nodes. The light squares denote the code-bits or check nodes in the complement of $\mathcal{B}$ and the dark squares denote the check nodes in $\mathcal{B}$. The thick path is an example of a self-avoiding path between A and B. The dashed path is an example of non-self-avoiding path. As shown in Appendix 4.A, the self-avoiding walks contribute towards estimating the correlation between the sets A and B.

Denote by $\mathcal{B}$ the set of all check nodes i, such that $|l_i| > H$. In words, the set $\mathcal{B}$ consists of all check nodes, whose associated "code-bit" has $\frac{1}{2}$LLR greater than the chosen number H. Recall that $\langle - \rangle$ denotes the average w.r.t. the Gibbs (MAP) measure (c.f. (4.14)). The proof of the following lemma can be found in Appendix 4.A.

Lemma 4.1 (Bound on the Correlation)**.** *Consider an LDGM code with bounded left and right degree. Consider two sets of variable nodes A and B. We have*

$$\left| \left\langle \prod_{a \in A} u_a \prod_{b \in B} u_b \right\rangle - \left\langle \prod_{a \in A} u_a \right\rangle \left\langle \prod_{b \in B} u_b \right\rangle \right| \leq 2 \sum_{w \in W_{AB}} \prod_{i \in w} \rho_i, \tag{4.15}$$

where the product $\prod_{i\in w}$ *is over all the check nodes of the self-avoiding walk* w*. Above,*

$$\rho_i = \begin{cases} 1, & i \in \mathcal{B}, \\ e^{4|l_i|} - 1, & i \notin \mathcal{B}. \end{cases}$$

Averaging over the noise realization in (4.15) we get

$$\mathbb{E}_{\underline{l}} \left| \Big\langle \prod_{a\in A} u_a \prod_{b\in B} u_b \Big\rangle - \Big\langle \prod_{a\in A} u_a \Big\rangle \Big\langle \prod_{b\in B} u_b \Big\rangle \right| \leq 2 \sum_{w\in W_{AB}} \prod_{i\in w} \mathbb{E}_{l_i}[\rho_i]. \tag{4.16}$$

In the r.h.s. of the above inequality, we took the expectation inside the product because each $w \in W_{AB}$ is a self-avoiding walk. As a result, all the check nodes on the walk are *distinct* and hence all the corresponding $\frac{1}{2}\mathsf{LLR}$ are independent random variables.

Now,

$$\begin{aligned} \mathbb{E}_{l_i}[\rho_i] &= \mathbb{E}_{l_i}\big[\rho_i \mid i \notin \mathcal{B}\big]\ \mathbb{P}\big(i \notin \mathcal{B}\big) + \mathbb{E}_{l_i}\big[\rho_i \mid i \in \mathcal{B}\big]\ \mathbb{P}\big(i \in \mathcal{B}\big) \\ &\leq (e^{4H} - 1) + \mathbb{P}\big(|l| > H\big) = \delta(\epsilon, H). \end{aligned} \tag{4.17}$$

Since we consider transmission over a channel chosen from the class of channels $\mathcal{K}$, condition (3) of the Definition 4.1 implies that we can choose $H = H(\epsilon)$ such that $K(\delta(\epsilon, H(\epsilon)))^{1/2} < 1$.

As a consequence, we get

$$\begin{aligned} \mathbb{E}_{\underline{l}} \left| \Big\langle \prod_{a\in A} u_a \prod_{b\in B} u_b \Big\rangle - \Big\langle \prod_{a\in A} u_a \Big\rangle \Big\langle \prod_{b\in B} u_b \Big\rangle \right| &\leq 2 \sum_{w\in W_{AB}} \Big(\delta(\epsilon, H)\Big)^{|w|/2} \\ &\leq 2|A||B| \sum_{d=\mathrm{dist}(A,B)}^{\infty} (K(\delta(\epsilon, H))^{1/2})^d \\ &\leq \frac{2|A||B|}{1 - K(\delta(\epsilon, H))^{1/2}} (K(\delta(\epsilon, H))^{1/2})^{\mathrm{dist}(A,B)}. \end{aligned} \tag{4.18}$$

The second inequality is obtained by noticing that the number of self-avoiding random walks of length $|w|$, with the same initial node, is bounded by $K^{|w|}$. Also, we have that the factor $\delta(\epsilon, H)$ is raised to a $1/2$, because a walk of length $|w|$ (which is the number of edges present in the walk) has $|w|/2$ check nodes. The factor $|A||B|$ accounts for the number of initial and final vertices. The correlation decay of the theorem is in fact a special case of this last bound for the choice $A = \partial i$ and $B = \partial j$ for the check nodes i and j (associated to code-bits i and j respectively). □

4.4.1 Exactness of BP Estimate

We now look at GEXIT functions of the MAP and BP decoders. For LDGM codes recall from (1.17) that the average MAP-GEXIT function is given by

$$g_n(\epsilon) = \mathbb{E}_{\mathcal{C}}\left[\int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim i}} \ln\left\{\frac{1+\langle x_i\rangle_0 \tanh l_i}{1+\tanh l_i}\right\}\right]. \tag{4.19}$$

The average BP-GEXIT curve is given by the same function with $\langle x_i\rangle_0$ replaced by $\langle x_i\rangle_{0,d}^{BP}$. Recall that $\langle x_i\rangle_{0,d}^{BP}$ corresponds to the estimate provided by the BP decoder after d iterations. Consider $N_d(i)$, the neighborhood of node i of radius d, an even integer (all the vertices at graph distance less or equal to d from i). As is well known for an LDGM(Λ, P) ensemble with bounded degrees, given d, if n is large enough, the probability that $N_d(i)$ is a tree is $1 - O(\frac{\gamma^d}{n})$ (where γ depends only on $l_{\max}$ and $k_{\max}$). Thus when d is fixed and $n \to +\infty$ the computation tree $T_d(i)$ and the neighborhood $N_d(i)$ match with high probability. This implies that the asymptotic average BP-GEXIT function is given by

$$\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \mathbb{E}_{\mathcal{C}}\left[\int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim i}} \ln\left\{\frac{1+\langle x_i\rangle_{0,N_d(i)} \tanh l_i}{1+\tanh l_i}\right\} \middle| N_d(i) \text{ is a tree}\right], \tag{4.20}$$

where $\langle x_i\rangle_{0,N_d(i)}$ is the Gibbs bracket associated to the subgraph $N_d(i)$. The right hand side can computed exactly by performing the statistical mechanical sums on a tree and it yields the density evolution formulas

$$\lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \lim_{d\to\infty}\int dl \frac{dc(l)}{d\epsilon}\mathbb{E}_{\Delta^{(d)}}\left[\ln\left\{\frac{1+\tanh\Delta^{(d)}\tanh l}{1+\tanh l}\right\}\right], \tag{4.21}$$

where both limits exist and $\mathbb{E}_{\Delta^{(d)}}$ is expectation w.r.t. the random variable $\Delta^{(d)}$, given by

$$\Delta^{(d)} = \tanh^{-1}\left(\prod_{i=1}^{k}\tanh v_i^{(d)}\right). \tag{4.22}$$

The $v_i^{(d)}$ are i.i.d. random variables with distribution obtained from the iterative system of density evolution equations

$$\eta^{(d)}(v) = \sum_l \lambda_l \int \prod_{i=1}^{l-1} du_i\, \widehat{\eta}^{(d)}(u_i)\delta(v - \sum_{i=1}^{l-1} u_i), \tag{4.23}$$

$$\widehat{\eta}^{(d)}(u) = \sum_k \rho_k \int dl c(l) \prod_{a=1}^{k-1} dv_a\, \eta^{(d-1)}(v_a)\delta(u - \tanh^{-1}(\tanh l \prod_{i=1}^{k-1}\tanh v_a), \tag{4.24}$$

with the initial condition $\eta^{(0)}(v) = \delta(v)$. Note that the degree k in (4.22) is random with distribution $P(x)$. Recall that these equations are also an iterative version of the stationary point equation of the replica solution.

Remark 4.1. *Note that for a LDGM code the BP decoder fails to start and is always stuck at the trivial fixed-point $\delta(v)$. But for an LDGM code with an arbitrary small fraction of degree-one check nodes, the BP decoder goes to a non-trivial fixed-point. Thus this corollary gives a non-trivial result when the fraction of degree-one check nodes present in the LDGM code ensemble is non-zero.*

Proof of Corollary 4.1. We want to prove that

$$\lim_{n\to+\infty} g_n(\epsilon) = \lim_{d\to+\infty} \lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon).$$

Recall that the Gibbs bracket $\langle - \rangle_0$ denotes the Gibbs average when the $\frac{1}{2}\mathsf{LLR}$ l_i is not present. Since we are considering the limit $n \to +\infty$, we have $\langle x_i \rangle_{0,d}^{BP} = \langle x_i \rangle_{0,N_d(i)}$ for any fixed d. Thus, in order to prove the corollary, we would like to prove that $\langle x_i \rangle_0 \approx \langle x_i \rangle_{0,N_d(i)}$. The idea is to use the decay of correlations to prove that

$$\langle x_i \rangle_0 = \langle x_i \rangle_{0,N_d(i)} + o_d(1).$$

Taking the limit $d \to +\infty$ allows us then to prove the corollary.

Let us make these informal statements more precise. Note that the MAP estimate, $\langle x_i \rangle_0$ appears in the logarithm in the average MAP-GEXIT (c.f. (4.19)). This makes it quite cumbersome to handle. So we first expand the logarithm (in both the numerator and the denominator) using $\log(1+z) = \sum_{p=1} \frac{(-1)^{p+1} z^p}{p}$ for $|z| \leq 1$ to find

$$\begin{aligned} g_n(\epsilon) &= \mathbb{E}_{\mathcal{C}} \left[\int dl_i \frac{dc(l_i)}{d\epsilon} \mathbb{E}_{\underline{l}^{\sim i}} \left\{ \sum_{p=1} \frac{(-1)^{p+1}}{p} \left(\left(\langle x_i \rangle_0 \tanh l_i \right)^p - \left(\tanh l_i \right)^p \right\} \right] \\ &= \mathbb{E}_{\mathcal{C}} \left[\sum_{p=1} \frac{(-1)^{p+1}}{p} \left(\int dl_i \frac{dc(l_i)}{d\epsilon} \left(\tanh l_i \right)^p \right) \left(\mathbb{E}_{\underline{l}^{\sim i}} \left[\langle x_i \rangle_0^p \right] - 1 \right) \right] \end{aligned}$$

where we used that the $\frac{1}{2}\mathsf{LLR}$ l_i does not appear in $\langle - \rangle_0$.

We now use the following Nishimori identities (see 2.A.2)

$$\begin{aligned} &\mathbb{E}_{\underline{l}^{\sim i}} \left[\langle x_i \rangle_0^{2p} \right] = \mathbb{E}_{\underline{l}^{\sim i}} \left[\langle x_i \rangle_0^{2p-1} \right] \\ &\mathbb{E}_{l_i} \left[(\tanh l_i)^{2p} \right] = \mathbb{E}_{l_i} \left[(\tanh l_i)^{2p-1} \right] \Rightarrow \frac{d\mathbb{E}_{l_i} \left[(\tanh l_i)^{2p} \right]}{d\epsilon} = \frac{d\mathbb{E}_{l_i} \left[(\tanh l_i)^{2p-1} \right]}{d\epsilon}, \end{aligned}$$

to combine the even and odd terms in the sum over p. Thus we obtain,

$$g_n(\epsilon) = \frac{\Lambda'(1)}{P'(1)} \sum_{p=1}^{+\infty} \frac{T_{2p}(\epsilon)}{2p(2p-1)} \left(\mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} [\langle x_i \rangle_0^{2p}] - 1 \right), \tag{4.25}$$

for the MAP-GEXIT for any BMS channel. We recall that

$$T_{2p}(\epsilon) = \frac{d}{d\epsilon} \int_{-\infty}^{+\infty} dl\, c(l) (\tanh l)^{2p}. \tag{4.26}$$

Recall that for the class $\mathcal{K}$ that we consider here, the condition (1) in Definition 4.1 implies that T_{2p} is uniformly bounded in p. Also, $|\mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}}[\langle x_i \rangle_0^{2p}] - 1| < 2$. As a result, the series (4.25) behaves like the series $\sum_p \frac{1}{p^2}$ which is uniformly convergent. Thus the series (4.25) is absolutely convergent, uniformly with respect to n, for the class $\mathcal{K}$. Thus by dominated convergence, the proof will be complete if we show that for every p,

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} \big[\langle x_i \rangle_0^{2p} \big] = \lim_{d\to+\infty} \lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} \big[\langle x_i \rangle_{0,N_d(i)}^{2p} | N_d(i) \text{ is a tree} \big]. \tag{4.27}$$

Indeed one can then compute the $n \to +\infty$ limit term by term in (4.25) and then resum the resulting series (which is again absolutely convergent, uniformly with respect to d) to obtain (4.20). Since we consider codes with bounded maximum degrees we have $\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} \big[\langle x_i \rangle_0^{2p} \big] = \lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} \big[\langle x_i \rangle_0^{2p} \mid N_d(i) \text{ is a tree} \big]$. Thus it is enough to show

$$\begin{aligned} &\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} \big[\langle x_i \rangle_0^{2p} \, | N_d(i) \text{ is a tree} \big] \\ &\qquad = \lim_{d\to+\infty} \lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} \big[\langle x_i \rangle_{0,N_d(i)}^{2p} \mid N_d(i) \text{ is a tree} \big]. \end{aligned} \tag{4.28}$$

Notice that all paths connecting the node i with those outside $N_d(i)$ have a length at least equal to d, so because of Theorem 4.1 in the high noise regime x_i is very weakly correlated to the complement of $N_d(i)$. Therefore we may expect that (4.28) holds.

Let us formally prove (4.28). Recall that $\mathring{N}_d(i)$ is the boundary of checks that are at distance d from i. We order the checks $\in \mathring{N}_d(i)$ in a given (arbitrary) way, and call $\langle - \rangle_{0;\le k}$ the Gibbs average with $l_k = 0$ for the k first checks of $\mathring{N}_d(i)$ (and $l_i = 0$ for the root node). For the first one (call it 1) we use $e^{l_1 x_1} = (\cosh l_1 + x_1 \sinh l_1)$ to get

$$\langle x_i \rangle_0 = \langle x_i \rangle_{0;\le 1} + \frac{\tanh l_1 \big(\langle x_i x_1 \rangle_{0;\le 1} - \langle x_i \rangle_{0;\le 1} \langle x_1 \rangle_{0;\le 1} \big)}{1 + \langle x_1 \rangle_{0;\le 1} \tanh l_1}. \tag{4.29}$$

Therefore

$$\begin{aligned} |\langle x_i \rangle_0^{2p} - \langle x_i \rangle_{0;\le 1}^{2p}| &\le 2p |\langle x_i \rangle_0 - \langle x_i \rangle_{0;\le 1}| \\ &\le 2p\, t_1 |\langle x_i x_1 \rangle_{0;\le 1} - \langle x_i \rangle_{0;\le 1} \langle x_1 \rangle_{0;\le 1}|, \end{aligned} \tag{4.30}$$

where

$$t_k = \frac{|\tanh l_k|}{1 - |\tanh l_k|}. \tag{4.31}$$

We remark that when $|l_k| \to +\infty$, the factor t_k diverges.We can now take the second check of $\mathring{N}_d(i)$ (call it 2) and show

$$|\langle x_i\rangle_{0;\leq 1}^{2p} - \langle x_i\rangle_{0;\leq 2}^{2p}| \leq 2p\, t_2 |\langle x_i x_2\rangle_{0;\leq 2} - \langle x_i\rangle_{0;\leq 2}\langle x_2\rangle_{0;\leq 2}|. \tag{4.32}$$

We can repeat this argument for all nodes of $\mathring{N}_d(i)$ and use the triangle inequality to obtain

$$|\langle x_i\rangle_0^{2p} - \langle x_i\rangle_{0,N_d(i)}^{2p}| \leq 2p \sum_{k\in\mathring{N}_d(i)} t_k |\langle x_i x_k\rangle_{0;\leq k} - \langle x_i\rangle_{0;\leq k}\langle x_k\rangle_{0;\leq k}|. \tag{4.33}$$

Indeed the Gibbs average with all $l_k = 0$ for all $k \in \mathring{N}_d(i)$ is equal to $\langle x_i\rangle_{0,N_d(i)}$. Now using the bound (4.18) in the proof of Theorem 4.1 for $K(\delta(\epsilon))^{1/2} < 1$, the last inequality implies

$$\mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}}\big[|\langle x_i\rangle_0^{2p} - \langle x_i\rangle_{0,N_d(i)}^{2p}| \mid N_d(i)\,\text{tree}\big] \leq \frac{4p\mathtt{r}_{\max}^2\mathbb{E}[t]}{1-K(\delta(\epsilon))^{1/2}} K^d (K(\delta(\epsilon))^{1/2})^d. \tag{4.34}$$

Note that since we are considering channels from the class $\mathcal{K}$, we use the condition (2) of Definition 4.1 to get

$$\mathbb{E}[t] = \mathbb{E}\left[\frac{|\tanh l|}{1-|\tanh l|}\right] \leq \mathbb{E}[e^{2|l|}] < \infty. \tag{4.35}$$

The right hand side of (4.34) does not depend on n, so it is immediate that $\lim_{d\to+\infty}\lim_{n\to+\infty}$ vanishes as long as the noise is high enough such that $K^2\delta(\epsilon) < 1$. This is again possible because of the condition (3) of Definition 4.1, which allows us to make $\delta(\epsilon)$ as small as desired. This proves (4.27) and the corollary. □

To conclude, let us remark that, for the BIAWGNC the GEXIT formulas simplify considerably and there is a clear relationship to the magnetization,

$$\begin{aligned} g_n(\epsilon) &= \frac{1}{\epsilon^3}\big(1 - \mathbb{E}_{\underline{l}}[\langle x_i\rangle]\big) \\ &= \frac{1}{\epsilon^3}(1 - \mathbb{E}_{\underline{l}}[\tanh(l + \tanh^{-1}\langle x_i\rangle_0)]), \end{aligned} \tag{4.36}$$

and our corollary gives

$$\lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \frac{1}{\epsilon^3}\big(1 - \mathbb{E}_{l,\Delta^{(d)}}\big[\tanh(l+\Delta^{(d)})\big]\big). \tag{4.37}$$

The proof of Corollary 1 for the BIAWGNC can thus proceed without expansions and is slightly simpler. Next, we show that for the BEC there are similar simplifications and this allows us to make a proof which avoids the second condition in the class $\mathcal{K}$.

Example 4.5 (Computations for the BEC). *For the BEC the factor* $\delta(\epsilon, H(\epsilon)) = e^{4H} - 1 + \mathbb{P}(|l| > H)$, *for any* $H > 0$ *equals* $e^{4H} - 1 + (1 - \epsilon)$. *So we can choose* ϵ, H *such that* $K(\delta(\epsilon, H))^{1/2} < 1$ *and we can argue exactly as in the proof of Theorem 4.1 in Section 4.4 to prove the correlation decay in this case also. To get a feel of the numbers* H, ϵ_g, *it is easy to check that for all* $\epsilon > 1 - \frac{1}{4K^2} \triangleq \epsilon_g$ *and for any* $H < \frac{1}{4} \ln \left(1 + \frac{1}{4K^2}\right)$, *we have* $K(\delta(\epsilon, H))^{1/2} < 1$.

To prove the exactness of the BP estimate for the case of BEC, we have from (2.26),

$$\begin{aligned} g_n(\epsilon) &= \ln 2(1 - \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_0]) \\ &= \ln 2\big(1 - \mathbb{E}_{\mathcal{C},\underline{l}}\big[\langle x_i \rangle_0 \mid N_d(i) \text{ is a tree}\big]\big) + o_n(1). \end{aligned}$$

We now proceed exactly as in (4.29), use the GKS inequality ($\langle x_1 \rangle_{0;\leq k} \geq 0$) to bound

$$\frac{\tanh l_k}{1 + \langle x_1 \rangle_{0;\leq k} \tanh l_k} \leq 1,$$

and conclude the Corollary 4.1 for the BEC as well.

4.4.2 Large Block Length Versus Large Number of Iterations for Computing BP-GEXIT

In the LDGM case we prove the exchange of limits $d, n \to +\infty$ for the BSC channel. As will become clear one needs the decay of correlations (or covariance) of the Gibbs measure on the computation tree for $d \gg n$. Hence the likelihoods are not independent r.v.: the proof of Theorem 4.1 still goes through in the case of the BSC when we consider the computation tree instead of the Tanner graph. Indeed, in Lemma 4.1 we can take $H > \frac{1}{2} \ln \frac{1-\epsilon}{\epsilon}$ such that $\mathcal{B} = \emptyset$ and $\rho_{\pi(j)} = e^{4|l_{\pi(j)}|} - 1 = \frac{4|1-2\epsilon|}{(1-|1-2\epsilon|)^2}$ for all $j \in T_d(i)$.

Lemma 4.2 (Decay of Correlations for the BP decoder, LDGM on BSC). *Consider communication with a fixed LDGM code with block-length* n *and bounded degrees* $l_{\max}$ *and* $k_{\max}$, *over the BSC*(ϵ). *We can find* $c > 0$, *a small enough numerical constant, such that for* $l_{\max} k_{\max} |1 - 2\epsilon| < c$ *we have, for any realization of the channel outputs,*

$$|\langle x_i x_j \rangle_d^{BP} - \langle x_i \rangle_d^{BP} \langle x_j \rangle_d^{BP}| \leq c_1 e^{-c_2(\epsilon) \text{dist}(\partial i, \partial j)}, \tag{4.38}$$

where i *is the root code-bit of the computation tree,* j *and arbitrary code-bit,* $c_1 > 0$ *a numerical constant and* $c_2(\epsilon) > 0$ *depending only on* ϵ, $l_{\max}$, $k_{\max}$. *Moreover,* $c_2(\epsilon)$ *increases like* $\ln |1 - 2\epsilon|$ *as* $\epsilon \to \frac{1}{2}$.

We remark the Tanner graph in this case should also be thought of in the same way as in Section 4.4 Basically, this result is contained in [62] where it is obtained by Dobrushin's criterion. Note that it is valid for fixed noise realizations

and not only on average. The unbounded case would require to take averages but then, on the computation tree one has to control moments $\mathbb{E}[\rho_\pi(i)^m]$ and this requires more work. The following proof is a simple application of this lemma.

Proof of Corollary 4.3. We consider $d \gg n$ iterations of the BP decoder. We consider the subtree of root i and depth $d' \ll n$ on the computation tree $T_d(i)$. This subtree is a smaller computation tree $T_{d'}(i) \subset T_d(i)$ and $d' \ll n \ll d$. Let $\mathring{T}_{d'}(i)$ be the boundary of check nodes k with $\mathrm{dist}(i,k) = d'$ and order them in an arbitrary way. Consider the Gibbs measure $\langle - \rangle^{BP}_{d;\leq k}$ where for the first k checks of $\mathring{T}_{d'}(i)$ we set $l_{\pi(k)} = 0$ in (4.7). Proceeding as in Section 4.4 we get

$$|\langle x_i \rangle^{BP}_{d} - \langle x_i \rangle^{BP}_{d'}| \leq \sum_{k \in \mathring{T}_{d'}(i)} t_{\pi(k)} |\langle x_i x_k \rangle^{BP}_{d;\leq k} - \langle x_i \rangle^{BP}_{d;\leq k} \langle x_k \rangle^{BP}_{d;\leq k}|. \tag{4.39}$$

For the BSC, $t_{\pi(k)} = \frac{|1-2\epsilon|}{1-|1-2\epsilon|}$. From Lemma 4.2, for $|1-2\epsilon|$ small enough (but independent of n, d)

$$|\langle x_i \rangle^{BP}_{d} - \langle x_i \rangle^{BP}_{d'}| = O(K^{d'} e^{-c_2(\epsilon) d'}). \tag{4.40}$$

In this equation $O(-)$ is uniformly bounded with respect to n and d (and the noise realizations of course). Recall the GEXIT function of the BP decoder

$$g_{n,d}(\epsilon) = \frac{\Lambda'(1)}{P'(1)} \int dl_i \frac{dc(l_i)}{d\epsilon} \mathbb{E}_{\mathcal{C}, \underline{l}^{\sim i}} \left[\ln \left\{ \frac{1 + \langle x_i \rangle^{BP}_{0,d} \tanh l_i}{1 + \tanh l_i} \right\} \right]. \tag{4.41}$$

Since for $|1-2\epsilon| \ll 1$ we have $|\tanh l_i| = \frac{1}{2} |\ln \frac{1-\epsilon}{\epsilon}| \ll 1$, one can easily show (by following same arguments as in the proof of Corollary 4.1 done in the previous section 4.4.1: by expanding the ln in powers of $|\tanh l_i|$ and estimate the series term by term.)

$$g_{n,d}(\epsilon) = g_{n,d'}(\epsilon) + O(K^{d'} e^{-c_2(\epsilon) d'}). \tag{4.42}$$

Now since $O(-)$ is uniformly bounded with respect to n, d, (4.42) implies for d' fixed

$$\lim_{n \to +\infty} \liminf_{d \to \infty} g_{n,d}(\epsilon) = \lim_{n \to +\infty} g_{n,d'}(\epsilon) + O(K^{d'} e^{-c_2(\epsilon) d'}). \tag{4.43}$$

Now we take the limit $d' \to +\infty$,

$$\lim_{n \to +\infty} \liminf_{d \to \infty} g_{n,d}(\epsilon) = \lim_{d' \to +\infty} \lim_{n \to +\infty} g_{n,d'}(\epsilon). \tag{4.44}$$

A similar result with lim sup replacing lim inf is derived in the same way. □

4.5 Proof of Main Theorem for the Ensemble $\mathfrak{K}$ of LDPC Codes

In this section we first prove Theorem 4.2 on the decay of correlations and then its implication in the form of the Corollary 4.2. Recall that the partition function for a code picked from the ensemble $\mathfrak{K}$ is given by

$$Z = \sum_{\underline{x}} \prod_{a \in \mathfrak{C}_0} \frac{1}{2}(1 + x_{\partial a}) \prod_{a \in \mathfrak{C}_1} \frac{1}{2}(1 + x_{\partial a}) \prod_{i \in \mathfrak{V}} e^{l_i x_i} \prod_{i \in \mathfrak{V}_1} e^{l_i x_i}.$$

We make the following crucial observation. Consider a check node $a \in \mathfrak{C}_1$ and its associated degree-one variable node. Together their contribution to the Gibbs measure is given by $\big(\frac{1+x_1 x_{\partial a \setminus 1}}{2}\big) e^{l_1 x_1}$, where x_1 denotes the degree-one variable node $\in \mathfrak{V}_1$. It is not hard to see that the effective contribution of this combination is given by

$$e^{l_1 x_{\partial a \setminus 1}}.$$

Indeed, $\big(\frac{1+x_1 x_{\partial a \setminus 1}}{2}\big)$ is non-zero if and only if $x_1 = x_{\partial a \setminus 1}$. One more way to see this is that, in deriving the cluster expansion (see Appendix 4.A), we can *sum out the degree-one variable nodes* in $\mathfrak{V}_1$, in both the numerator and denominator of equation (4.64), in an independent manner. We thus obtain the previous claim by noticing that

$$\sum_{x_1} \Big(\frac{1 + x_1 x_{\partial a \setminus 1}}{2}\Big) e^{l_1 x_1} = e^{l_1 x_{\partial a \setminus 1}}. \tag{4.45}$$

Remark 4.2. *With above observation, any code from our ensemble $\mathfrak{K}$, is basically a mixture of "LDGM-type" soft checks (those belonging to $\mathfrak{C}_1$) and "LDPC-type" hard checks (those belonging to $\mathfrak{C}_0$).*

Proof of Theorem 4.2, LDPC. We prove the main theorem for i and j both belonging to $\mathfrak{V}$ (c.f. Definition 4.2). We again refer to Section 4.4, and view the Tanner graph for a code from ensemble $\mathfrak{K}$ in a similar manner (each check node a in $\mathfrak{C}_1$ is associated with the function $e^{l_1 x_{\partial a \setminus 1}}$ (c.f. (4.45)) and the variable nodes corresponding to the code-bits in $\mathfrak{V}_1$ are not present in the modified Tanner graph).

We make use of the bound on the correlation provided by the Lemma 4.1. We choose the set $\mathcal{B}$ of check nodes as follows. We put all the check nodes belonging to the set $\mathfrak{C}_0$, in the set $\mathcal{B}$. These are all the hard parity-check constraints. We also include in the set $\mathcal{B}$, all those check nodes $a \in \mathfrak{C}_1$, for which the absolute value of the associated LLR (c.f. (4.45)) is greater than some constant H. With this, we find in the bound of Lemma 4.1, $\rho_a = e^{|4l_a|} - 1 \leq e^{4H} - 1$ for $a \notin \mathcal{B}$ and $\rho_a = 1$ for $a \in \mathcal{B}$.

Applying now the correlation decay Lemma 4.1, for $A = i$ and $B = j$ (where i, j are any two code-bits belonging to $\mathfrak{V}$), we obtain

$$\mathbb{E}_{\underline{l}}\big[\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|\big] \leq 2\mathbb{E}_{\underline{l}}\Bigg[\sum_{w\in W_{ij}}\prod_{c\in w}\Big((e^{4H}-1)\mathbb{1}(c\notin\mathcal{B})+\mathbb{1}(c\in\mathcal{B})\Big)\Bigg],$$

where $\mathbb{1}(\cdot)$ is the indicator function. Using the definition of the set $\mathcal{B}$ in our case, we write the indicator function explicitly as,

$$\begin{aligned}\mathbb{1}(c\in\mathcal{B}) &= \mathbb{1}(c\in\mathfrak{C}_1, |l_c|>H) + \mathbb{1}(c\in\mathfrak{C}_0),\\ &= \mathbb{1}(|l_c|>H)\ \mathbb{1}(c\notin\mathfrak{C}_0) + \mathbb{1}(c\in\mathfrak{C}_0). \end{aligned} \tag{4.46}$$

Now since the walks are self-avoiding, all the check nodes in a given path are independent and we can take the expectation w.r.t. the noise inside the product yielding

$$\begin{aligned}&\mathbb{E}_{\underline{l}}\big[|\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|\big]\\ &\leq 2\sum_{w\in W_{ij}}\prod_{c\in w}\Big(\Big[(e^{4H}-1)+\mathbb{P}(|l_c|>H)\Big]\mathbb{1}(c\notin\mathfrak{C}_0)+\mathbb{1}(c\in\mathfrak{C}_0)\Big)\\ &\leq 2\sum_{w\in W_{ij}}\prod_{c\in w}\Big(\delta(\epsilon,H)\mathbb{1}(c\notin\mathfrak{C}_0)+\mathbb{1}(c\in\mathfrak{C}_0)\Big) \qquad (4.47)\\ &\leq 2\sum_{w\in W_{ij}}\prod_{c\in w}\Big(\delta(\epsilon,H)+\mathbb{1}(c\in\mathfrak{C}_0)\Big),\end{aligned}$$

where we use $\delta(\epsilon,H)\triangleq e^{4H}-1+\mathbb{P}(|l_c|>H)$. In the first inequality we used (4.46) coupled with the fact that since $\mathfrak{C}_0\subseteq\mathcal{B}$, we have $\mathbb{1}(c\notin\mathcal{B})\leq\mathbb{1}(c\notin\mathfrak{C}_0)$. In words, if a check node does not belong to $\mathcal{B}$, then it cannot belong to $\mathfrak{C}_0$. Now taking the average w.r.t. the randomness in $\underline{S}$ we find

$$\begin{aligned}\mathbb{E}_{\underline{S}}\mathbb{E}_{\underline{l}}\big[|\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|\big] &\leq 2\sum_{w\in W_{ij}}\prod_{c\in w}\big(\delta(\epsilon,H)+\mathbb{E}_{\underline{S}}[\mathbb{1}_{c\in\mathfrak{C}_0}(c)]\big)\\ &= 2\sum_{w\in W_{ij}}\big(\delta(\epsilon,H)+1-\mathfrak{p}\big)^{|w|/2}\\ &\leq 2\sum_{d\geq\mathrm{dist}(i,j)}\Big(K\Big[\delta(\epsilon,H)+1-\mathfrak{p}\Big]^{1/2}\Big)^d,\end{aligned}$$

where we again upper bound the number of walks of length d with the same initial node by K^d. In the first inequality we could take the expectation w.r.t. $\underline{S}$ inside the product, because each check node has a degree-one variable node attached to itself independent of any other check node. For our class $\mathcal{K}$, we can choose ϵ and p to be large enough such that $K[\delta(\epsilon,H(\epsilon))+(1-\mathfrak{p})]^{1/2}<1$ (c.f. the condition (3) of Definition 4.1). As a result, we get

$$\mathbb{E}_{\underline{S}}\mathbb{E}_{\underline{l}}\big[|\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|\big] = \frac{2\big(K[\delta(\epsilon,H(\epsilon))+1-\mathfrak{p}]^{1/2}\big)^{\mathrm{dist}(i,j)}}{1-K[\delta(\epsilon,H(\epsilon))+1-\mathfrak{p}]^{1/2}},$$

proving the decay of correlation theorem. □

4.5.1 Exactness of BP Estimate

Arguing on the same lines as in Section 4.4.1, we have the average BP-GEXIT, for fixed d and in the limit $n \to +\infty$, is given by,

$$\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \frac{1}{(1+(1-\mathtt{R})\mathfrak{p})}\mathbb{E}_{\mathcal{C},\underline{S}}\Bigg[\int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim i}}\Big\{\ln\Big(\frac{1+\langle x_i\rangle_{0,N_d(i)}\tanh l_i}{1+\tanh l_i}\Big)\Big\}\Bigg] + \frac{(1-\mathtt{R})\mathfrak{p}}{(1+(1-\mathtt{R})\mathfrak{p})}\mathbb{E}_{\mathcal{C},\underline{S}}\Bigg[\int dl_a \frac{dc(l_a)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim a}}\Big\{\ln\Big(\frac{1+\langle x_a\rangle_{0,N_d(a)}\tanh l_a}{1+\tanh l_a}\Big)\Big\}\Bigg]. \tag{4.48}$$

From (4.48) and arguing on the same lines as in Section 4.4.1 we get

$$\lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \lim_{d\to+\infty}\Bigg\{\frac{1}{(1+(1-\mathtt{R})p)}\int dl\frac{dc(l)}{d\epsilon}\mathbb{E}_{\Delta_0^{(d)}}\Bigg[\ln\Big(\frac{1+\tanh\Delta_0^{(d)}\tanh l}{1+\tanh l}\Big)\Bigg] + \frac{(1-\mathtt{R})p}{(1+(1-\mathtt{R})p)}\int dl\frac{dc(l)}{d\epsilon}\mathbb{E}_{\Delta_1^{(d)}}\Bigg[\ln\Big(\frac{1+\tanh\Delta_1^{(d)}\tanh l}{1+\tanh l}\Big)\Bigg]\Bigg\}, \tag{4.49}$$

where $\mathbb{E}_{\Delta_0^{(d)}}$ ($\mathbb{E}_{\Delta_1^{(d)}}$) denotes the expectation w.r.t. $\Delta_0^{(d)}$ ($\Delta_1^{(d)}$), where the random variables $\Delta_0^{(d)}$ and $\Delta_1^{(d)}$ are given by

$$\Delta_1^{(d)} = \tanh^{-1}\Big(\prod_{i=1}^{k}\tanh v_i^{(d)}\Big),$$
$$\Delta_0^{(d)} = \sum_{i=1}^{l} u_i^{(d)}. \tag{4.50}$$

The $v_i^{(d)}$ are i.i.d. random variables with distribution obtained from the iterative system of density evolution equations

$$\widehat{\eta}^{(d)}(u) = \sum_k \rho_k\Big\{\mathfrak{p}\int dlc(l)\prod_{i=1}^{k-1}dv_i\eta^{(d-1)}(v_i) \times\delta\big(u-\tanh^{-1}\big(\tanh l\prod_{i=1}^{k-1}\tanh v_i\big)\big)+(1-\mathfrak{p}) \times\int\prod_{i=1}^{k-1}\eta^{(d-1)}(v_i)\delta\big(u-\tanh^{-1}\big(\prod_{i=1}^{k-1}\tanh v_i\big)\big)\Big\}, \tag{4.51}$$

$$\eta^{(d)}(v) = \sum_\ell \lambda_\ell\int dlc(l)\prod_{c=1}^{\ell-1}du_c\widehat{\eta}^{(d)}(u_c)\delta(v-l-\sum_{c=1}^{l-1}u_c), \tag{4.52}$$

where $c(l)$ denotes the distribution of the $\frac{1}{2}\mathsf{LLR}$ assuming transmission of the all zero codeword. The initial condition is $\eta^{(0)}(v) = c(v)$. Here $\eta^{(d)}(v)$ is the density of messages from a variable node in $\mathfrak{V}$ to a check node and $\widehat{\eta}^{(d)}(u)$ the density of messages from a check node to a variable node in $\mathfrak{V}$. Note that in (4.50), the degrees l and k are random with distributions $\Lambda(x)$ and $P(x)$ respectively.

To prove our main theorem we utilize the following lemma. It is stated for a variable node belonging to $\mathfrak{V}$ but is also valid for a variable node belonging to $\mathfrak{V}_1$. Recall that $K = \mathtt{l}_{\max}\mathtt{r}_{\max}$. Define $\alpha = \ln \frac{(K-1)^{K-1}}{(1-\mathtt{p})\mathtt{p}^{K-1}K^K}$ we have,

Lemma 4.3 (Existence of Soft Boundary)**.** *Pick a code in the ensemble $\mathfrak{K}$ and a node $i \in \mathfrak{V}$ u.a.r. Take an integer $L > \frac{\ln K}{\alpha}$, a depth d and consider the two neighborhoods $N_d(i)$ and $N_{(L+1)d}(i)$. The integer L is to be considered large enough but fixed and not depending on d and n. With high probability $1 - o_{n,d}(1)$ these two neighborhoods are trees and there exists a check node boundary $\mathring{\mathsf{B}}_S$ between $\mathring{N}_d(i)$ and $\mathring{N}_{(L+1)d}(i)$ such that all the check nodes in $\mathring{\mathsf{B}}_S$ belong to the set $\mathfrak{C}_1$. Here $o_{n,d}(1)$ is a function of d, n which goes to zero as $n \to +\infty$ followed by $d \to +\infty$.*

Proof. The fact that the neighborhoods are tree like with high probability is well known for sparse graphs (see for example [23]). The existence of a boundary $\mathring{\mathsf{B}}_S$ of soft checks follows from the standard analysis of a Birth-Death process and is presented in Appendix 4.B. □

Let us denote the graph, enclosed by the boundary $\mathring{\mathsf{B}}_S$ (which includes the root variable node i), by B_S.

Proof of Corollary 4.2. As in the proof of Corollary 4.1, we will be done if (c.f. (4.10))

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\left[\langle x_i\rangle_0^{2p}\right] = \lim_{d\to+\infty}\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\left[\langle x_i\rangle_{0,N_d(i)}^{2p} | N_d(i) \text{ is a tree}\right] \quad (4.53)$$

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim a}}\left[\langle x_a\rangle_0^{2p}\right] = \lim_{d\to+\infty}\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim a}}\left[\langle x_a\rangle_{0,N_d(a)}^{2p} | N_d(a) \text{ is a tree}\right] \quad (4.54)$$

We provide the proof for the first term in (4.53) (estimate for variable nodes in $\mathfrak{V}$), since for the other one all arguments are similar.

Consider a code-bit $x_i \in \mathfrak{V}$ and the neighborhoods $N_d(i), N_{(L+1)d}(i)$ around it, where L is the same constant as in Lemma 4.3. Let us denote T the event that $N_{(L+1)d}(i)$ is a tree. Since this is a high probability event for bounded degree distributions we have,

$$\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}[\langle x_i\rangle_0^{2p}] = \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}[\langle x_i\rangle_0^{2p} \mid \mathsf{T}] + o_n(1).$$

Let $\mathcal{S}$ denote the set of all $\underline{S}$ such that there exists a boundary between $\mathring{N}_d(i)$ and $\mathring{N}_{(L+1)d}(i)$ such that all the check nodes in the boundary belong to $\mathfrak{C}_1$.

From the previous lemma we know that $\mathbb{P}(\mathcal{S}) \geq 1 - o_{n,d}(1)$. Thus we have

$$\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}}[\langle x_i \rangle_0^{2p}] = \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}}[\langle x_i \rangle_0^{2p} \mid \mathsf{T}, \mathcal{S}] + o_{n,d}(1),$$

Let us now compute $\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}}[\langle x_i \rangle_0^{2p} \mid \mathsf{T}, \mathcal{S}]$. For a $\underline{S} \in \mathcal{S}$, let us denote the boundary[1] which contains only check nodes belonging to $\mathfrak{C}_1$, by $\mathring{\mathsf{B}}_S$. Recall that for any check node c in the boundary $\mathring{\mathsf{B}}_S$ the effective contribution of $\frac{1}{2}(1 + x_1 x_{\partial c \setminus 1}) e^{l_1 x_1}$ to the Gibbs measure is given by $e^{l_1 x_{\partial c \setminus 1}}$.

As in the proof in the case of LDGM codes in Section 4.4.1, we order the checks belonging to $\mathring{\mathsf{B}}_S$ in a given (arbitrary) way, and call $\langle - \rangle_{0;\leq k}$ the Gibbs average with $l_k = 0$ for the k first checks of $\mathring{\mathsf{B}}_S$ (and $l_i = 0$ for the root node). For the first one (call it 1) we again use $e^{l_1 x_1} = (\cosh l_1 + x_1 \sinh l_1)$ to get

$$\langle x_i \rangle_0 = \langle x_i \rangle_{0;\leq 1} + \frac{\tanh l_1 \big(\langle x_i x_1 \rangle_{0;\leq 1} - \langle x_i \rangle_{0;\leq 1} \langle x_1 \rangle_{0;\leq 1} \big)}{1 + \langle x_1 \rangle_{0;\leq 1} \tanh l_1}. \tag{4.55}$$

We repeat this argument for all check nodes of $\mathring{\mathsf{B}}_S$ to obtain

$$|\langle x_i \rangle_0^{2p} - \langle x_i \rangle_{0,\mathsf{B}_S}^{2p}| \leq 2p \sum_{k \in \mathring{\mathsf{B}}_S} t_k |\langle x_i x_k \rangle_{0;\leq k} - \langle x_i \rangle_{0;\leq k} \langle x_k \rangle_{0;\leq k}|, \tag{4.56}$$

where t_k is defined in (4.31). Indeed the Gibbs average with all $l_k = 0$ for all $k \in \mathring{\mathsf{B}}_S$ is equal to $\langle x_i \rangle_{0,\mathsf{B}_S}$.

Proceeding as in the proof of the correlation decay Theorem 4.2 (c.f. (4.47)) we get

$$\begin{aligned} &\mathbb{E}_{\underline{l}^{\sim k}} |\langle x_i x_k \rangle_{0;\leq k} - \langle x_i \rangle_{0;\leq k} \langle x_k \rangle_{0;\leq k}| \\ &\qquad \leq 2 \Big(\sum_{w \in W_{ik}} \prod_{c \in w} \Big[\delta(\epsilon, H) \mathbb{1}(c \notin \mathfrak{C}_0) + \mathbb{1}(c \in \mathfrak{C}_0) \Big] \Big), \end{aligned} \tag{4.57}$$

where $\mathcal{B}$, $\delta(\epsilon, H)$ are the same as defined in the proof of Theorem 4.2 on decay of correlations.

Now we consider the average w.r.t. the randomness in $\underline{S}$. Since we restrict the random vector $\underline{S}$ to lie in $\mathcal{S}$ we "lose" the randomness in the region between $\mathring{N}_d(i)$ and $\mathring{N}_{(L+1)d}(i)$ (but fortunately the number of such check nodes on any self-avoiding path of any length l is upper bounded by Ld). Thus for check nodes belonging to this intermediate region we bound $[\delta(\epsilon, H) \mathbb{1}(c \notin \mathfrak{C}_0) + \mathbb{1}(c \in \mathfrak{C}_0)]$ by 1. As a result, for any self-avoiding path w we have

$$\prod_{c \in w} \Big[\delta(\epsilon, H) \mathbb{1}(c \notin \mathfrak{C}_0) + \mathbb{1}(c \in \mathfrak{C}_0) \Big] \leq \prod_{\substack{c \in w \\ c \notin N_{(L+1)d}(i) \setminus N_d(i)}} \Big[\delta(\epsilon, H) \mathbb{1}(c \notin \mathfrak{C}_0) + \mathbb{1}(c \in \mathfrak{C}_0) \Big].$$

[1] There can be many soft boundaries between $N_d(i)$ and $N_{(L+1)d}(i)$. We choose any one of them for the purpose of our analysis and call is $\mathring{\mathsf{B}}_S$.

Now taking expectation, w.r.t. the randomness in $\underline{S} \in \mathcal{S}$, of the r.h.s. of (4.57), we get

$$
\begin{aligned}
&2\mathbb{E}_{\underline{S}}\Big[\sum_{w\in W_{ik}} \prod_{\substack{c\in w\\ c\notin N_{(L+1)d}(i)\setminus N_d(i)}} \Big(\delta(\epsilon,H)\mathbb{1}(c\notin\mathfrak{C}_0)+\mathbb{1}(c\in\mathfrak{C}_0)\Big) \mid \mathcal{S}\Big]\\
&\overset{(a)}{\leq} 2\sum_{w\in W_{ik}} \prod_{\substack{c\in w\\ c\notin N_{(L+1)d}(i)\setminus N_d(i)}} \Big(\delta(\epsilon,H)+\mathbb{E}_{\underline{S}}\big[\mathbb{1}(c\in\mathfrak{C}_0)\mid\mathcal{S}\big]\Big)\\
&= 2\sum_{w\in W_{ik}} \prod_{\substack{c\in w\\ c\notin N_{(L+1)d}(i)\setminus N_d(i)}} \Big(\delta(\epsilon,H)+1-\mathfrak{p}\Big)\\
&\leq 2\sum_{l\geq(L+1)d} K^l\Big(\big[\delta(\epsilon,H)+1-\mathfrak{p}\big]^{1/2}\Big)^{l-Ld}+\sum_{l=d}^{(L+1)d} K^l\Big(\big[\delta(\epsilon,H)+1-\mathfrak{p}\big]^{1/2}\Big)^{d},
\end{aligned}
\tag{4.58}
$$

where in (a) we used the fact that any check node $c \notin N_{(L+1)d}(i)\setminus N_d(i)$ belongs to $\mathfrak{C}_0$ independent of any other check node, with the probability $1-\mathfrak{p}$. We also utilize the fact that along any self-avoiding walk w we have distinct check nodes. In the last inequality we did several things. Firstly, we upper bound the total number of paths of length l with the same initial node by K^l. Secondly, for paths with length $l \geq (L+1)d$ we do not consider the check nodes belonging to $N_{(L+1)d}(i)\setminus N_d(i)$ which gives $\big([\delta(\epsilon,H)+1-\mathfrak{p}]^{1/2}\big)^{l-Ld}$ as a contribution of such paths. Finally, for paths with length $l \leq Ld$ we do not consider check nodes outside $N_d(i)$, which results in a contribution of $\big([\delta(\epsilon,H)+1-\mathfrak{p}]^{1/2}\big)^{d}$ by such paths. The power of $1/2$ arises because the number of check nodes in a path of length $|l|$ is equal to $|l|/2$.

Choose $p > 1-\frac{1}{2(K^2)^{2(L+1)}}$ and ϵ such that $\delta(\epsilon,H(\epsilon)) < (1-\mathfrak{p})$. This would imply that $\sqrt{2(1-\mathfrak{p})}K^{2(L+1)} < 1$. Putting all together, from (4.56), we get

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}[|\langle x_i\rangle_0^{2p}-\langle x_i\rangle_{0,\mathsf{B}_S}^{2p}| \mid \mathsf{T},\mathcal{S}] \leq 4p\mathbb{E}[t]\Big(K^{2(L+1)}\sqrt{2(1-\mathfrak{p})}\Big)^{d}\\
\times\Big(\frac{1}{1-(K\sqrt{2(1-\mathfrak{p})})}+Ld\Big),
\end{aligned}
\tag{4.59}
$$

where we bound the number of check nodes in the boundary $\mathring{\mathsf{B}}_S$ by $K^{(L+1)d}$ and the second sum in (4.58) by $LdK^{(L+1)d}(\sqrt{2(1-\mathfrak{p})})^d$. Since for the class $\mathcal{K}$, the term $\mathbb{E}[t]$ is bounded (c.f. condition (3) of Definition 4.1), we find that the r.h.s in the above inequality is less than ζ^d for some $\zeta < 1$.

We will be done if we can show that

$$
\lim_{d\to+\infty}\lim_{n\to+\infty}\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,\mathsf{B}_S}^{2p}\mid\mathsf{T},\mathcal{S}\big]=\lim_{d\to+\infty}\lim_{n\to+\infty}\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,N_d(i)}^{2p}\mid\mathsf{T},\mathcal{S}\big]
$$

To evaluate $\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,\mathsf{B}_S}^{2p} \mid \mathsf{T},\mathcal{S}\big]$, it is sufficient to compute $\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,\mathsf{B}_S}^{2p} \mid \mathsf{T}\big]$. Indeed, since $\mathbb{P}(\mathcal{S}^c) \le o_{n,d}(1)$ and $|\langle x_i\rangle_{\mathsf{B}_S}^{2p}| \le 1$ we have

$$\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,\mathsf{B}_S}^{2p} \mid \mathsf{T}\big] = \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,\mathsf{B}_S}^{2p} \mid \mathsf{T},\mathcal{S}\big] + o_{n,d}(1).$$

Define

$$\langle x_i\rangle_{0,B}^{2p} = \begin{cases} \langle x_i\rangle_{0,N_{(L+1)d}(i)}^{2p}, & \underline{S} \notin \mathcal{S}, \\ \langle x_i\rangle_{0,\mathsf{B}_S}^{2p}, & \underline{S} \in \mathcal{S}. \end{cases}$$

Arguing as above we also have $\big|\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}[\langle x_i\rangle_{0,\mathsf{B}_S}^{2p}] - \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}[\langle x_i\rangle_{0,B}^{2p}]\big| \le o_{n,d}(1)$.

Applying Lemma 3.11 in Appendix 3.D of Chapter 3, we get

$$\begin{aligned}\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,N_d(i)}^{2p} \mid \mathsf{T}\big] &\le \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,B}^{2p} \mid \mathsf{T}\big] \\ &\le \mathbb{E}_{\mathcal{C},\underline{S},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,N_{(L+1)d}(i)}^{2p} \mid \mathsf{T}\big]. \qquad (4.60)\end{aligned}$$

Since L is a constant which depends only on code parameters, taking the limit $d \to +\infty$ in (4.60) we obtain the result. □

For the BIAWGNC the average MAP-GEXIT formula simplifies to

$$\begin{aligned}\frac{d}{d\epsilon}\mathbb{E}_{\mathcal{C},\underline{S}}[h_N] = \frac{1}{\epsilon^{-3}(1+(1-\mathtt{R})\mathfrak{p})}\Big(1 - \mathbb{E}_{\mathcal{C},\underline{S}}\big[\langle x_i\rangle\big]\Big) \\ + \frac{(1-\mathtt{R})\mathfrak{p}}{\epsilon^{-3}(1+(1-\mathtt{R})\mathfrak{p})}\Big(1 - \mathbb{E}_{\mathcal{C},\underline{S}}\big[\langle x_a\rangle\big]\Big) + o_n(1),\end{aligned}$$

Thus our corollary implies that, for high enough noise, the two soft code-bit MAP estimates can be computed using the density evolution equations as

$$\begin{aligned}\lim_{n\to+\infty}\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}}\big[\langle x_i\rangle\big] &= \lim_{d\to+\infty}\mathbb{E}_{l,\Delta_0^{(d)}}\Big[\tanh\Big(l+\Delta_0^{(d)}\Big)\Big], \\ \lim_{n\to+\infty}\mathbb{E}_{\mathcal{C},\underline{S},\underline{l}}\big[\langle x_a\rangle\big] &= \lim_{d\to+\infty}\mathbb{E}_{l,\Delta_1^{(d)}}\Big[\tanh\Big(l+\Delta_1^{(d)}\Big)\Big],\end{aligned}$$

where $\Delta_0^{(d)}$ and $\Delta_1^{(d)}$ are defined in (4.50).

4.6 Discussion and Open Questions

In this chapter we have shown that cluster expansion techniques of statistical mechanics are a valuable tool for the theory of error correcting codes on graphs. We showed that for high enough noise regimes of a fairly general class of BMS channels and the MAP decoder, the average correlation between any two code-bits decays exponentially with their graph distance. We showed this result for two class of codes, namely LDGM codes with bounded maximum degrees and the ensemble of LDPC codes, $\mathfrak{K}$. As an important consequence we showed that the asymptotic average MAP-GEXIT function can be computed via density

evolution equations. In other words, we showed that the replica solution is exact.

There are several open issues here. Firstly, it would be nice to extend our results to a wider class of channels. Secondly, we were able to show decay of correlations and consequently the exactness of the replica solution only for small regimes of noise. We believe that the limitedness of the regime of noise in which we could prove our results is an artifact of our techniques and that the correlation should decay in a larger regime.

We have not investigated the regimes of low noise for LDGM codes. In this regime there will be very large areas of "singularity" (equivalently the set $\mathcal{B}$ will be very large). As a result, the cluster expansion technique we use here would fail to give us any meaningful estimate. Nevertheless, we still believe that the correlations should decay exponentially in the low noise regime also. Indeed one can argue informally as follows. Consider transmission over the BIAWGNC using any fixed LDGM code. For *any* noise variance of the gaussian channel we can write, using our sum-rule technique, for the special case of the BIAWGNC,

$$\mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle] = \mathbb{E}_{\mathcal{C},\underline{l}}[\langle x_i \rangle_{N_d(i)}] + 2\mathbb{E}_{\mathcal{C}} \sum_{k \in \mathring{N}_d(i)} \int_{\epsilon}^{\infty} \frac{d\nu}{\nu^3} \mathbb{E}_{\underline{l}}[\big(\langle x_i x_k \rangle_{\leq k} - \langle x_i \rangle_{\leq k} \langle x_k \rangle_{\leq k}\big)^2].$$

Empirically[2] for large block-lengths, we know that the BP estimate is equal to the MAP estimate for some values of noise even in the low noise regime. Combining this observation with the above sum-rule we conclude that the correlations vanish. In fact the correlation decay is strong enough to even kill the local graph expansion. It would be interesting to determine if the BP threshold signals a change in the nature of decay of correlations (of the MAP decoder). This would relate the BP decoder and the MAP decoder in an intimate manner, different from the Maxwell's construction of [36].

In the case of LDPC codes and high noise we are able to prove decay of correlations for the special ensemble of LDPC codes defined by $\mathfrak{K}$. *This is the only example known to us for which such a rigorous derivation is made beyond the BEC with non-combinatorial methods.* In fact one can extend this result for any LDPC code ensemble that contains a sufficient fraction of degree one variable nodes. This would include the Poisson-LDPC code ensemble. We do not present the result here as the computations are technically involved. The idea behind the proof was to eliminate the degree-one nodes and convert the problem to a new graphical model containing a mixture of hard parity-check constraints and soft LDGM type weights. We then show the decay of correlations if the density of soft check nodes is high enough. This method of "summing" out the degree-one variable nodes to convert a low-temperature problem (LDPC) to a high temperature problem (LDGM) resembles the renormalization-group

[2]For BEC we in fact can prove that the BP estimate is equal to the MAP estimate in the large block-length limit for low noise regime [23], [18], [19], [106].

technique of statistical mechanics. It would be interesting to see if the "summing" out technique can be applied to a larger class of LDPC codes. For example, one could "sum out" the degree two variable nodes. A potential hurdle in this approach is that the degrees of check nodes in the resulting graph can increase unbounded. Dobrushin and Bassalygo [109] also developed a technique to show uniqueness of Gibbs measure in a random spin-glass with large values of the spin-spin interactions. In their technique one also "converts" the low temperature system to a high temperature system at the price of unbounded degrees.

One of the consequences of the decay of correlations was the proof of the equivalence of $\lim_{n\to+\infty}\lim_{d\to+\infty}$ and $\lim_{d\to+\infty}\lim_{n\to+\infty}$ while evaluating the average BP-GEXIT function, in the high noise regime for the case of transmission over the BSC channel. Firstly, it would be nice to extend this result to the more general case of channels with unbounded LLR , like the BIAWGNC. Recently it was shown in [37], that the limits of large block-length and iteration number can be exchanged for computing the bit-error rate of the BP decoder. This was shown for various message-passing algorithms for the case of transmission over general BMS channels. Their result holds only in the regime of noise where the bit-error rate vanishes, i.e. in the *low noise regime*. We point out that although the BP-GEXIT function and the bit-error rate of BP decoder are related, at the moment our techniques do not extend to the case of bit-error rate. More precisely, the BP-GEXIT function is a "soft" estimate and the bit-error rate is a "hard" estimate which makes its analysis more complicated. Thus a second goal here would be to see if one could extend our methods to prove the exchange of limits for computing the bit-error rate of the BP decoder in the high noise regime.

We hope that the ideas and techniques investigated in the present work will have other applications in coding theory and, more broadly, random graphical models. We mentioned in Section 1.6 of Chapter 2, that decay of correlation was crucially used in the investigation of the random k-SAT problem at low constraint density in [58]. This has allowed the authors to prove that the replica symmetric solution is exact at low constraint density. It is not hard to show that the decay of correlations which we show is stronger than the worst case decay of correlation used by Montanari and Shah in [58]. We hope that this analysis can be extended to the random k-satisfiability problem and counting the number of independent set problem, as mentioned in the Section 1.6 of Chapter 1. More precisely, suppose we are able to associate an appropriate Gibbs distribution to the above mentioned problems. And suppose we consider an ensemble of graph/formulas from which we pick a graph/formula uniformly at random. Then demonstrating that the correlations of the form $\mathbb{E}\big[|\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|\big]$ decay exponentially with the graph distance, could help improve the range of the parameters for which the results for the above problems hold (see [58], [59]).

As mentioned in Section 1.6 of Chapter 1, in [55], [56] the authors derive a new type of expansion called "loop expansion" in an attempt to compute

corrections to BP equations. The link to traditional cluster expansions is unclear, and also it would be interesting to develop rigorous methods to control the loop expansions.

4.A Cluster Expansion for LDGM Codes

Here we adapt the cluster expansion of Dreifus, Klein and Perez in [101] to our setting of Tanner graphs. In the process we prove Lemma 4.1 stated in Section 4.4. Using the notation $u_{\partial i} = \prod_{a\in\partial i} u_a$ for $i = 1, ..., n$, we find that the correlation (of Lemma 4.1) becomes

$$\langle u_A u_B\rangle - \langle u_A\rangle\langle u_B\rangle. \tag{4.61}$$

It is first necessary to rewrite the Gibbs measure (4.6) in a form such that the exponent is positive

$$\frac{1}{Z}\prod_{i=1}^{m} e^{l_i x_i} = \frac{1}{Z'}\prod_{i=1}^{m} e^{l_i u_{\partial i}+|l_i|}, \tag{4.62}$$

where Z' is the appropriately modified partition function. Notice that with this modification we do not change any averages, since the added factors will be canceled. We introduce the replicated measure, which is the product of two identical and independent copies,

$$\frac{1}{Z'^2}\prod_{i=1}^{m} e^{l_i(u^{(1)}_{\partial i}+u^{(2)}_{\partial i})+2|l_i|}. \tag{4.63}$$

Thus we now have two replicas of the information-bits $u^{(1)}_1, ..., u^{(1)}_m$ and $u^{(2)}_1, ..., u^{(2)}_m$. The Gibbs bracket for the replicated measure is denoted by $\langle -\rangle_{12}$. It is not hard to see that

$$\langle u_A u_B\rangle - \langle u_A\rangle\langle u_B\rangle = \frac{1}{2}\langle (u^{(1)}_A - u^{(2)}_A)(u^{(1)}_B - u^{(2)}_B)\rangle_{12}. \tag{4.64}$$

Indeed, expand the r.h.s. of (4.64) to obtain

$$\frac{1}{2}\langle (u^{(1)}_A - u^{(2)}_A)(u^{(1)}_B - u^{(2)}_B)\rangle_{12} = \frac{1}{2}\langle (u^{(1)}_A u^{(1)}_B + u^{(2)}_A u^{(2)}_B - u^{(1)}_A u^{(2)}_B - u^{(2)}_A u^{(1)}_B)\rangle_{12}$$

Using the linearity of the Gibbs bracket we get

$$\begin{aligned}\frac{1}{2}\langle (u^{(1)}_A - u^{(2)}_A)(u^{(1)}_B - u^{(2)}_B)\rangle_{12} =& \frac{1}{2}\Big(\big\langle u^{(1)}_A u^{(1)}_B\big\rangle_{12} + \big\langle u^{(2)}_A u^{(2)}_B\big\rangle_{12}\Big)\\ &- \frac{1}{2}\Big(\big\langle u^{(1)}_A u^{(2)}_B\big\rangle_{12} + \big\langle u^{(2)}_A u^{(1)}_B\big\rangle_{12}\Big).\end{aligned}$$

From (4.63) we see that

$$\begin{aligned}\big\langle u^{(1)}_A u^{(1)}_B\big\rangle_{12} &= \big\langle u^{(1)}_A u^{(1)}_B\big\rangle_{1} = \big\langle u_A u_B\big\rangle,\\ \big\langle u^{(2)}_A u^{(2)}_B\big\rangle_{12} &= \big\langle u^{(2)}_A u^{(2)}_B\big\rangle_{2} = \big\langle u_A u_B\big\rangle.\end{aligned}$$

Note that we have dropped one of the replicas in the Gibbs averages in the second equality above.

Also, since the replicated measure is a product measure (c.f. (4.63)), we have

$$\langle u_A^{(1)} u_B^{(2)} \rangle_{12} = \langle u_A^{(1)} \rangle_1 \langle u_B^{(2)} \rangle_2 = \langle u_A \rangle \langle u_B \rangle,$$
$$\langle u_A^{(2)} u_B^{(1)} \rangle_{12} = \langle u_A^{(2)} \rangle_2 \langle u_B^{(1)} \rangle_1 = \langle u_A \rangle \langle u_B \rangle.$$

Thus combining all of the above we find (4.64).

Recall that $\mathcal{B} = \{i \mid |l_i| > H\}$ for some fixed number H, and set

$$e^{l_i(u_{\partial i}^{(1)} + u_{\partial i}^{(2)}) + 2|l_i|} - 1 = K_i. \tag{4.65}$$

It will be important to keep in mind later that $K_i \geq 0$. Expanding the product $\prod_{i \in \mathcal{B}^c}(1 + K_i)$ we find

$$\begin{aligned}
&\frac{1}{2}\langle (u_A^{(1)} - u_A^{(2)})(u_B^{(1)} - u_B^{(2)}) \rangle_{12} \\
&\qquad = \frac{1}{2Z'^2} \sum_{u^{(1)}, u^{(2)}} f_A f_B \prod_{i \in \mathcal{B}} e^{l_i(u_{\partial i}^{(1)} + u_{\partial i}^{(2)}) + 2|l_i|} \prod_{i \in \mathcal{B}^c} (1 + K_i) \\
&\qquad = \frac{1}{2Z'^2} \sum_{u^{(1)}, u^{(2)}} f_A f_B \prod_{i \in \mathcal{B}} e^{l_i(u_{\partial i}^{(1)} + u_{\partial i}^{(2)}) + 2|l_i|} \sum_{G \subseteq \mathcal{B}^c} \prod_{i \in G} K_i \\
&\qquad = \frac{1}{2Z'^2} \sum_{G \subseteq \mathcal{B}^c} \sum_{u^{(1)}, u^{(2)}} f_A f_B \prod_{i \in \mathcal{B}} e^{l_i(u_{\partial i}^{(1)} + u_{\partial i}^{(2)}) + 2|l_i|} \prod_{i \in G} K_i,
\end{aligned} \tag{4.66}$$

where $f_X = u_X^{(1)} - u_X^{(2)}$, $X = A, B$ and G is any subset of distinct check nodes of the set $\mathcal{B}^c$.

Take a term with given $G \subseteq \mathcal{B}^c$ in the last sum. We say that "G connects A and B" if and only if there exist a self-avoiding walk[3] w_{ab} with initial variable node $a \in A$, final variable node $b \in B$ and such that all check nodes of w_{ab} are in $G \cup \mathcal{B}$[4]. The crucial point is that: if a set G does not connect A and B, then it gives a vanishing contribution to the sum. We defer the proof of this fact to the end of this section. For the moment let us show that it implies the bound in Lemma 4.1.

Consider a self-avoiding walk w_{ab} connecting A to B and lying entirely in $G \cup \mathcal{B}$. Note that there can be many such self-avoiding walks. We split the contribution of $G \cup \mathcal{B}$, for a given G, as follows

$$\prod_{i \in \mathcal{B}} e^{l_i(u_{\partial i}^{(1)} + u_{\partial i}^{(2)}) + 2|l_i|} \prod_{i \in G} K_i = \prod_{i \in \mathcal{B}} e^{l_i(u_{\partial i}^{(1)} + u_{\partial i}^{(2)}) + 2|l_i|} \prod_{i \in w_{ab} \setminus \mathcal{B}} K_i \prod_{i \in G'} K_i,$$

[3]See Section 4.4 for the definition of these walks.

[4]Note that it is really $G \cup \mathcal{B}$ that connects A and B. Since $\mathcal{B}$ is fixed our definition is valid

where it is easy to see that $G' \subset \mathcal{B}^c \setminus w_{ab}$. Thus using the positivity of K_i we have

$$\begin{aligned}
&|\langle u_A u_B\rangle - \langle u_A\rangle\langle u_B\rangle| \\
&\quad \le \frac{2}{Z'^2} \sum_{\substack{G\subseteq\mathcal{B}^c \\ G \text{ connects A and B}}} \sum_{u^{(1)},u^{(2)}} \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in G} K_i \\
&\quad \le \frac{2}{Z'^2} \sum_{w\in W_{AB}} \sum_{G'\subseteq\mathcal{B}^c\setminus w} \sum_{u^{(1)},u^{(2)}} \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in w\setminus\mathcal{B}} \mathfrak{h}_i \prod_{i\in G'} K_i, \quad (4.67)
\end{aligned}$$

where in the first inequality we use the trivial bound $|f_X| \le 2$ for both $X = A, B$. In the second inequality we used $K_i \le e^{4|l_i|} - 1 \equiv \mathfrak{h}_i$ for $i \in w \setminus \mathcal{B}$. Now resumming over $G' \subseteq \mathcal{B}^c \setminus w$ we obtain

$$\begin{aligned}
&|\langle u_A u_B\rangle - \langle u_A\rangle\langle u_B\rangle| \\
&\quad \le \frac{2}{Z'^2} \sum_{w\in W_{AB}} \prod_{i\in w\setminus\mathcal{B}} \mathfrak{h}_i \sum_{u^{(1)},u^{(2)}} \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in\mathcal{B}^c\setminus w} (1+K_i). \quad (4.68)
\end{aligned}$$

Inserting extra terms $(1 + K_i) \ge 1$ for $i \in w \setminus \mathcal{B}$ we find

$$\begin{aligned}
&|\langle u_A u_B\rangle - \langle u_A\rangle\langle u_B\rangle| \quad (4.69) \\
&\quad \le \frac{2}{Z'^2} \sum_{w\in W_{AB}} \prod_{i\in w\setminus\mathcal{B}} \mathfrak{h}_i \sum_{u^{(1)},u^{(2)}} \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in w\setminus\mathcal{B}} (1+K_i) \prod_{i\in\mathcal{B}^c\setminus w} (1+K_i) \\
&\quad = 2 \sum_{w\in W_{AB}} \prod_{i\in w\setminus\mathcal{B}} \mathfrak{h}_i. \quad (4.70)
\end{aligned}$$

The last equality follows by reconstituting Z'^2 in the numerator. Indeed, for *any* walk we have

$$\prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in w\setminus\mathcal{B}} (1+K_i) \prod_{i\in\mathcal{B}^c\setminus w} (1+K_i) = \prod_{i=1}^{n} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|}.$$

Using

$$2 \sum_{w\in W_{AB}} \prod_{i\in w\setminus\mathcal{B}} \mathfrak{h}_i = 2 \sum_{w\in W_{AB}} \prod_{i\in w} \rho_i,$$

where $\rho_i = 1$ for $i \in \mathcal{B}$ and $\rho_i = \mathfrak{h}_i$ for $i \notin \mathcal{B}$ we get the bound (4.15).

It remains to explain why, if G does not connect A and B, the G-term does not contribute to (4.66). Let $\partial G \cup \partial\mathcal{B}$ be the set of variable nodes connected to the check nodes $G \cup \mathcal{B}$. We define a partition $\partial G \cup \partial\mathcal{B} = \mathfrak{V}_A \cup \mathfrak{V}_C \cup \mathfrak{V}_B$ into three sets of variable nodes. $\mathfrak{V}_A$ is the set of all variable nodes $v \in \partial G \cup \partial\mathcal{B}$ such that there exist a self-avoiding walk w_{av} connecting some $a \in A$ to v, and such that all check nodes of w_{av} are in $G \cup \mathcal{B}$. $\mathfrak{V}_B$ is similarly defined with B

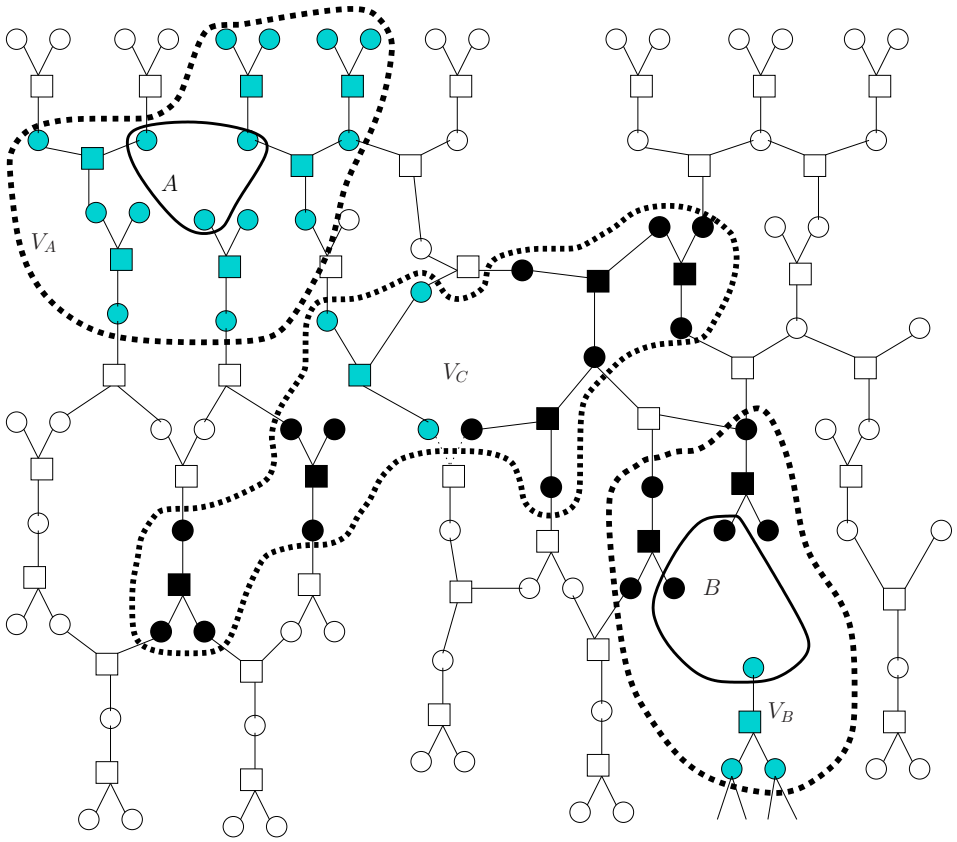

Figure 4.7: In the figure we show a part of the graph representing a LDGM code. The set $\mathcal{B}$ is depicted by the dark squares. The set $G \subseteq \mathcal{B}^c$ is depicted by the light squares. A and B contain both three nodes and are demarcated by the continuous curves. From the figure we see that there does not exist a self-avoiding walk that connects A to B with all its check nodes in $G \cup B$. The sets of variable nodes $\mathfrak{V}_A$, $\mathfrak{V}_B$ and $\mathfrak{V}_C$ are disjoint as well as the sets of check nodes G_A, G_B and G_C: these sets are enclosed in the dotted areas.

and $b \in B$ instead of A. Finally $\mathfrak{V}_C = (\partial G \cup \partial \mathcal{B}) \setminus (\mathfrak{V}_A \cup \mathfrak{V}_B)$. By construction $\mathfrak{V}_C \cap \mathfrak{V}_A = \mathfrak{V}_C \cap \mathfrak{V}_B = \emptyset$. The point is that if G does not connect A and B, then $\mathfrak{V}_A \cap \mathfrak{V}_B = \emptyset$. Indeed, otherwise there would be a $u \in \mathfrak{V}_A \cap \mathfrak{V}_B$ with a walk w_{au} and a walk w_{ub} both with all check nodes in $G \cup \mathcal{B}$, but this would mean that G connects A and B through the walk $w_{au} \cup w_{ub}$. We also define three sets of check nodes $\mathfrak{C}_A = (G \cup \mathcal{B}) \cap \partial \mathfrak{V}_A$, $\mathfrak{C}_B = (G \cup \mathcal{B}) \cap \partial \mathfrak{V}_B$ and $\mathfrak{C}_C = (G \cup \mathcal{B}) \setminus (G_A \cup G_B)$. Again the three sets are disjoint when G does not connect A and B: indeed if there exists $c \in \mathfrak{C}_A \cap \mathfrak{C}_B$ then ∂c has an intersection with both $\mathfrak{V}_A$ and $\mathfrak{V}_B$ which would again imply that G connects A and B. This situation is depicted in Figure 4.7.

Now we examine a term of (4.66) for a G that does not connect A and B. Expanding the product $f_A f_B$ and using linearity of the bracket, the term

equals

$$\frac{1}{2Z'^2}\Big(\sum_{u^{(1)},u^{(2)}} u_A^{(1)}u_B^{(1)}\prod_{i\in G\cup\mathcal{B}}\mathbb{H}_i(u^{(1)},u^{(2)})+\sum_{u^{(1)},u^{(2)}} u_A^{(2)}u_B^{(2)}\prod_{i\in G\cup\mathcal{B}}\mathbb{H}_i(u^{(1)},u^{(2)})$$
$$-\sum_{u^{(1)},u^{(2)}} u_A^{(1)}u_B^{(2)}\prod_{i\in G\cup\mathcal{B}}\mathbb{H}_i(u^{(1)},u^{(2)})-\sum_{u^{(1)},u^{(2)}} u_A^{(2)}u_B^{(1)}\prod_{i\in G\cup\mathcal{B}}\mathbb{H}_i(u^{(1)},u^{(2)})\Big),$$

where

$$\mathbb{H}_i(u^{(1)},u^{(2)})=\begin{cases}e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|}, & i\in\mathcal{B},\\ \prod_{i\in G}K_i, & i\in G.\end{cases}$$

Observe that $\prod_{i\in G\cup\mathcal{B}}\mathbb{H}_i(u^{(1)},u^{(2)})$ is symmetric under exchange of replicas $(1)\leftrightarrow(2)$ for any i. More precisely we have, for any i, $\mathbb{H}_i(u^{(1)},u^{(2)})=\mathbb{H}_i(u^{(2)},u^{(1)})$. As a consequence, the $G-$term contribution is is equal to the difference $I-II$ where

$$I=\frac{1}{Z'^2}\sum_{u^{(1)},u^{(2)}} u_A^{(1)}u_B^{(1)}\prod_{i\in\mathcal{B}}e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|}\prod_{i\in G}K_i, \tag{4.71}$$

$$II=\frac{1}{Z'^2}\sum_{u^{(1)},u^{(2)}} u_A^{(1)}u_B^{(2)}\prod_{i\in\mathcal{B}}e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|}\prod_{i\in G}K_i. \tag{4.72}$$

Because of the disjointness of the sets $\mathfrak{V}_A$, $\mathfrak{V}_B$ and $\mathfrak{V}_C$ and the sets $\mathfrak{C}_A$, $\mathfrak{C}_B$ and $\mathfrak{C}_C$ (the areas enclosed in dotted lines, see Figure 4.7) one can, in I and II, factor the sums $\sum_{u^{(1)},u^{(2)}}$ in a product of three terms (in fact there is a fourth trivial term which is a power of 2 coming from the bits outside the dotted areas) as

$$I=\frac{2^{2|\mathfrak{V}\setminus(\mathfrak{V}_A\cup\mathfrak{V}_B\cup\mathfrak{V}_C)|}}{Z'^2}\Big(\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ a\in\mathfrak{V}_A}} u_A^{(1)}\prod_{i\in\mathfrak{C}_A}\mathbb{H}_i\Big)\Big(\sum_{\substack{u_b^{(1)},u_b^{(2)}\\ b\in\mathfrak{V}_B}} u_B^{(1)}\prod_{i\in\mathfrak{C}_B}\mathbb{H}_i\Big)\Big(\sum_{\substack{u_c^{(1)},u_c^{(2)}\\ c\in\mathfrak{V}_C}}\prod_{i\in\mathfrak{C}_C}\mathbb{H}_i\Big),$$

$$II=\frac{2^{2|\mathfrak{V}\setminus(\mathfrak{V}_A\cup\mathfrak{V}_B\cup\mathfrak{V}_C)|}}{Z'^2}\Big(\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ a\in\mathfrak{V}_A}} u_A^{(1)}\prod_{i\in\mathfrak{C}_A}\mathbb{H}_i\Big)\Big(\sum_{\substack{u_b^{(1)},u_b^{(2)}\\ b\in\mathfrak{V}_B}} u_B^{(2)}\prod_{i\in\mathfrak{C}_B}\mathbb{H}_i\Big)\Big(\sum_{\substack{u_c^{(1)},u_c^{(2)}\\ c\in\mathfrak{V}_C}}\prod_{i\in\mathfrak{C}_C}\mathbb{H}_i\Big),$$

where for convenience we drop the dependence on $u^{(1)}$, $u^{(2)}$ of $\mathbb{H}_i$. Then by symmetry $(1)\leftrightarrow(2)$ one recognizes that $I=II$. Thus $I-II=0$ and this proves that G does not contribute to (4.66) when it does not connect A and B.

4.B Existence of Soft Boundary

In this appendix we prove Lemma 4.3. This lemma is crucial for us to apply the sum-rule to relate the MAP and BP estimates. Let us point out that the

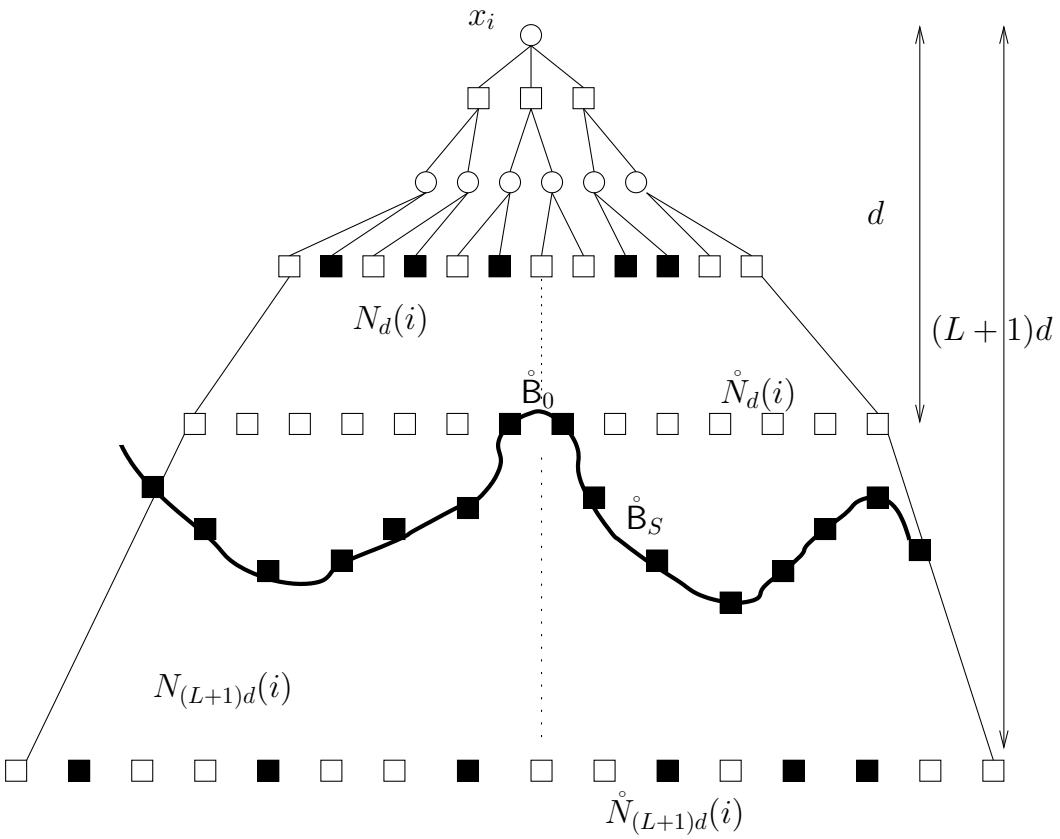

Figure 4.8: Figure shows the two tree neighborhoods $N_d(i), N_{(L+1)d}(i)$ as well as their boundaries $\mathring{N}_d(0), \mathring{N}_{(L+1)d}(i)$. Check nodes belonging to $\mathfrak{C}_1$ are shaded black and those belonging to $\mathfrak{C}_0$ are white. We also show the irregular boundary $\mathring{\mathsf{B}}_S$ consisting of only "soft" check nodes belonging to $\mathfrak{C}_1$ and lying between the two neighborhoods.

notion of the boundary is to be thought of as a set of check nodes, which when appropriately "removed" would separate the region "inside" the boundary from the one "outside".

Proof. Since we consider codes with bounded maximum degrees, from standard arguments [23] the neighborhood $N_{(L+1)d}(i)$ is a tree with high probability. Pick a code randomly from the ensemble $\mathfrak{K}$ such that $N_{(L+1)d}(i)$ is a tree and consider a variable node i in $\mathfrak{V}$. Consider the check nodes belonging to $\mathring{N}_d(i)$. We define the following process of finding the required "soft" boundary $\mathring{\mathsf{B}}_S$ consisting only of check nodes belonging to the set $\mathfrak{C}_1$. Initially we set boundary $\mathring{\mathsf{B}}_S = \mathring{\mathsf{B}}_0 \triangleq \{c | c \in \mathring{N}_d(i) \text{ and } c \in \mathfrak{C}_1\}$. In words, consider the check nodes in the boundary $\mathring{N}_d(i)$, belonging to the set $\mathfrak{C}_1$, as the initial constituents of the soft boundary.

Next, pick a check node $c \in \mathring{N}_d(i) \setminus \mathring{\mathsf{B}}_0$. Traverse down along all possible paths emanating from c till we hit a check node belonging to $\mathfrak{C}_1$ on all the possible paths. We include all these check nodes belonging to $\mathfrak{C}_1$ in our

boundary $\mathring{\mathsf{B}}_S$. Now pick another check node $c \in \mathring{N}_d(i) \setminus \mathring{\mathsf{B}}_0$ and follow down all the possible paths originating from this check node, till we encounter a "soft" check node. Again include all these "soft" check nodes in $\mathring{\mathsf{B}}_S$. Continue this process for all the check nodes in $\mathring{N}_d(i) \setminus \mathring{\mathsf{B}}_0$. At the end we have our soft boundary $\mathring{\mathsf{B}}_S$. The region enclosed by $\mathring{\mathsf{B}}_S$, containing the root variable node i, is denoted by B_S. Clearly, if we "remove" all the check nodes from $\mathring{\mathsf{B}}_S$ we would end up disconnecting B_S from the rest of the graph.

We now associate the following Birth-Death (B-D) process, independently to each check node in the set $\mathring{N}_d(i) \setminus \mathring{\mathsf{B}}_0$. This would model the above procedure for finding the soft-boundary. We say that a check node is "alive" if it has no degree-one variable node belonging to $\mathfrak{V}_1$ attached to it and "dead" otherwise. A check node is alive with probability $1 - \mathfrak{p}$ and dead with probability $\mathfrak{p}$. At each time, an alive check node gives rise to the next generation of check nodes and then dies. The number of check nodes fathered by any check node is given by $(\mathtt{l}-1)(\mathtt{r}-1)$, where $\mathtt{l}$ corresponds to the variable node degree and $\mathtt{r}$ is the check node degree. Initially there is only one alive check node ($\in \mathring{N}_d(i) \setminus \mathring{\mathsf{B}}_0$). It gives rise to a random number of check nodes and then dies. The B-D process continues like this till the time becomes $Ld+1$, at which we stop the B-D process, irrespective of whether there are alive check nodes at that time. If the B-D process stops before reaching time $Ld+1$, then we say that the B-D process dies naturally. Suppose that all the B-D processes, running parallely for all check nodes in $\mathring{N}_d(i) \setminus \mathring{\mathsf{B}}_0$, stop naturally. Then we claim that we would have found our boundary of soft check nodes. Indeed, the soft-boundary consists of all the dead check nodes from every B-D process. Also, this soft-boundary would clearly lie between $\mathring{N}_d(i)$ and $\mathring{N}_{(L+1)d}(i)$. Also, since the B-D processes are independent, we have the region enclosed by the boundary to be a tree.

Let us give a bound on the probability that a single B-D process does not stop naturally, i.e., the extinction time is strictly greater than Ld. Let Y_{t-1} denote the number of alive check nodes at time $t-1$. Let Z_t denote the number of alive check nodes generated at time t by one alive check node at time $t-1$. Thus the number of alive check nodes at time t is given by $Y_t = Y_{t-1} + Z_t - 1$. Here Z_t is distributed as $\text{Binomial}\big((\mathtt{r}-1)(\mathtt{l}-1), 1-\mathfrak{p}\big)$. Thus from standard arguments [110], [111], [112] we have

$$\mathbb{P}(T > Ld) \leq \mathbb{P}(Y_{Ld} > 0) = \mathbb{P}(Z_1 + Z_2 + \cdots + Z_{Ld} \geq Ld).$$

From the Markov inequality and the independence of $\{Z_i\}$, we have for $s \geq 0$,

$$\begin{aligned}
\mathbb{P}(Z_1 + Z_2 + \cdots + Z_{Ld} \geq Ld) &\leq \mathbb{E}[e^{s(Z_1+Z_2+\cdots+Z_{Ld})}]e^{-sLd} \\
&= (\mathbb{E}[e^{sZ_1}])^{Ld} e^{-sLd} \\
&\overset{(a)}{=} \mathbb{E}_{\mathtt{l},\mathtt{r}}\Big[\big(\mathfrak{p} + e^s(1-\mathfrak{p})\big)^{Ld(\mathtt{l}-1)(\mathtt{r}-1)}\Big] e^{-sLd} \\
&\overset{(b)}{\leq} \big(\mathfrak{p} + e^s(1-\mathfrak{p})\big)^{LdK} e^{-sLd},
\end{aligned}$$

where in (a) we used the characteristic function of a Binomial$\big((\mathtt{r}-1)(\mathtt{l}-1), 1-\mathfrak{p}\big)$ distribution. In (b) we use $s \geq 0$ to find $\mathfrak{p} + e^s(1-\mathfrak{p}) \geq 1$, which gives us the inequality since $(\mathtt{l}-1)(\mathtt{r}-1) \leq K$. Above, $\mathbb{E}_{\mathtt{l},\mathtt{r}}$ denotes expectation w.r.t. the degrees $\mathtt{l}$ and $\mathtt{r}$ which are distributed as $\Lambda(x)$ and $P(x)$ respectively.For $\mathfrak{p} > 1 - \frac{1}{K}$ the B-D process extincts in finite time. Also, it is not hard to see that for $\mathfrak{p} > 1 - \frac{1}{K}$, setting $s = \ln\big(\frac{\mathfrak{p}}{(1-\mathfrak{p})(K-1)}\big)$ minimizes the above bound and we get

$$\mathbb{P}(T > Ld) \leq e^{-Ld\alpha},$$

where $\alpha = \ln \frac{(K-1)^{K-1}}{(1-\mathfrak{p})\mathfrak{p}^{K-1}K^K} > 0$.

Since $|\mathring{N}_d(i) \setminus \mathsf{B}_0| \leq K^d$, and each check node $c \in \mathring{N}_d(i) \setminus \mathsf{B}_0$ gives rise to an independent and identical B-D process, the probability that all the B-D processes extinct within time Ld is lower bounded by

$$\begin{aligned}
&\mathbb{P}(\text{all birth-death processes are extinct within depth } Ld) \\
&\quad = \big[1 - \mathbb{P}(T > Ld)\big]^{|\mathring{N}_d(i)\setminus \mathsf{B}_0|} \\
&\quad \geq (1 - e^{-\alpha Ld})^{K^d}.
\end{aligned}$$

Note that since $L > \frac{\ln K}{\alpha}$, the probability that all birth-death processes are extinct within depth Ld is lower bounded by $1 - o_d(1)$. □

Decay of Correlations: High Signal-to-Noise Ratio

5

5.1 Introduction

In this chapter we study the correlations between code-bits of an LDPC code when transmitting over general BMS channels in the low noise regime.

The Gibbs measures associated with the optimal decoder of LDPC codes confront us with new challenges which invalidate the direct use of the standard methods. For example it is easy to see that the standard Dobrushin type methods [99], [62] fail due to the presence of hard parity-check constraints. As seen in the last chapter, for the high noise regime we were able to convert the problem, in the special case of LDPC code ensemble with large fraction of degree one variable nodes, to a spin-glass containing a mixture of soft and hard constraints for which appropriate cluster expansions can be applied to show decay of correlations. The low noise regime which is our interest here is a truly "low-temperature" spin-glass problem for which all the above methods fail. Our method entails transforming LDPC codes to a dual one that involves "negative weights". As a result we cannot proceed by probabilistic methods and we resort to cluster expansion techniques different from those employed in the previous chapter.

The rest of the chapter is organized as follows. In the next section we set-up our channel models and encoding schemes. In Section 5.3 we state our main result on decay of correlation and its corollary for computing the asymptotic average MAP-GEXIT via the density evolution equations. Before going on to prove our main theorem, we present duality formulas for LDPC codes in Section 5.4. The duality idea coupled with cluster expansion technique allows us to prove our main theorem in Section 5.5. In Section 5.6 we sketch our main application of our result to the asymptotic average MAP-GEXIT function. We prove that in the low noise regime where the average correlation decays

(fast enough) the asymptotic average MAP-GEXIT function can be exactly computed from the density evolution analysis. These curves remain non-trivial all the way down to zero noise as long as there are degree one variable nodes (e.g Poisson LDPC codes). This proves that a non-trivial replica solution is the exact expression for the conditional entropy of a class of LDPC codes (containing a fraction of degree one variable nodes) on our class of general channels. Previously the result was known only for the BEC [23], [18], [19], [113], [114], [106], [115] and for a class of codes in the high noise regime as shown in the previous chapter. We end the chapter with some discussion and open problems. Appendix 5.D contains the derivation of the cluster expansion for LDPC codes.

5.2 Set-Up – Channels and Codes

In this section we describe the high SNR channel models (low noise regime) for which we will state and prove our results. The class of channels that we define here is essentially the same as the one in Chapter 4, except for the high noise condition. Consider a BMS(ϵ) channel defined by a transition p.d.f. $p_{Y|X}(y \mid x)$ with inputs $x \in \{-1, +1\}$ and outputs belonging to $\bar{\mathbb{R}}$. As in the previous chapter, we use the $\frac{1}{2}\mathsf{LLR}$ convention. The $\frac{1}{2}\mathsf{LLR}$ is given by

$$l = \frac{1}{2} \ln \left[\frac{p_{Y|X}(y|+1)}{p_{Y|X}(y|-1)} \right], \tag{5.1}$$

We denote the distribution of the $\frac{1}{2}\mathsf{LLR}$, under the all-one codeword assumption, by $c(l)$. As usual the noise value varies in the interval $[0, \epsilon_{\max}]$ where $\epsilon_{max} = \frac{1}{2}$ for the BSC, $\epsilon_{max} = +\infty$ for the BIAWGNC and $\epsilon_{max} = 1$ for the BEC. We now define the general class of channels for which our main results hold.

Definition 5.1 (Class of Channels – $\mathcal{K}$)**.** *A channel belongs to the class $\mathcal{K}$ of BMS channels if it satisfies:*

1. *The numbers $T_{2p}(\epsilon) = \frac{d}{d\epsilon} \int_{-\infty}^{\infty} dl\, c(l) (\tanh l)^{2p}$ are bounded uniformly with respect to the positive integer p ($p \geq 1$).*

2. *For any finite $m > 0$ we have $\mathbb{E}[e^{m|l|}] \leq c_m < +\infty$.*

3. *(Low noise condition) There exists $s_0 > 0$ small enough such that for $0 < s \leq s_0$ we have $\lim_{\epsilon \to 0} \mathbb{E}[e^{-sl}] = 0$.*

Notice that the first and second condition are exactly the same as in the definition of class of general BMS channels in the previous chapter. The only change is the modification of the high noise condition to a low noise condition.

Let us provide some examples of the channels which belong to this class.

Example 5.1 (BSC)**.** *Consider the BSC*(ϵ) *with* $\epsilon \in (0, \frac{1}{2})$*. We recall from the previous chapter that*

$$p_{Y|X}(y|x) = (1-\epsilon)\delta(y-x) + \epsilon\delta(y+x),$$
$$c(l) = (1-\epsilon)\delta\left(l - \frac{1}{2}\ln\left[\frac{1-\epsilon}{\epsilon}\right]\right) + \epsilon\delta\left(l - \frac{1}{2}\ln\left[\frac{\epsilon}{1-\epsilon}\right]\right). \quad (5.2)$$

As seen from the Example 4.1 in Section 4.2, the first two conditions are met. Let us turn our attention to the low noise condition. From the expression for the $\frac{1}{2}$LLR *distribution we get*

$$\mathbb{E}[e^{-sl}] = \epsilon^{\frac{s}{2}}(1-\epsilon)^{1-\frac{s}{2}} + (1-\epsilon)^{\frac{s}{2}}\epsilon^{1-\frac{s}{2}}.$$

As a consequence, as long as $0 < s < 2$*, we get* $\lim_{\epsilon\to 0}\mathbb{E}[e^{-sl}] = 0$*. Thus the low noise condition is satisfied.*

Example 5.2 (BIAWGNC)**.** *This example shows that our class of channels,* $\mathcal{K}$*, covers the BIAWGNC. Recall from Section 4.2 that the transition p.d.f,* $p_{Y|X}(y \mid x)$*, and the* $\frac{1}{2}$LLR *distribution* $c(l)$*, assuming the all-one codeword transmission are given by,*

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\epsilon}}\exp\left(-\frac{(y-x)^2}{2\epsilon^2}\right),\ c(l) = \frac{\epsilon}{\sqrt{2\pi}}\exp\left(-\frac{(l-\epsilon^{-2})^2}{2\epsilon^{-2}}\right). \quad (5.3)$$

Again the first two conditions are met as shown in Example 4.2. All that remains is the low noise condition. Using the characteristic function formula for the gaussian distribution we find

$$\mathbb{E}[e^{-sl}] = e^{-s\epsilon^{-2}+\frac{s^2\epsilon^{-2}}{2}}$$
$$= e^{-s\epsilon^{-2}(1-\frac{s}{2})}.$$

Clearly, for any $0 < s < 2$ *we have* $\lim_{\epsilon\to 0}\mathbb{E}[e^{-sl}] = 0$*, thus satisfying the low noise condition.*

Note that the BEC is not contained in the class $\mathcal{K}$ because of the second condition. Nevertheless, due to the special nature of this channel our methods can easily be adapted.

Codes

We use for transmission a fixed LDPC code or a code picked u.a.r. from a standard LDPC$(\Lambda(x), P(x))$ ensemble with bounded maximum degrees.

Recall the definition of graph distance $\mathrm{dist}(i,j)$ provided by Definition 4.4 in Chapter 4. Also, recall that $\mathbb{E}_{\underline{l}}$ denotes the expectation w.r.t. the output LLRs denote by the vector $\underline{l}$. Further recall that $\mathbb{E}_{\underline{l}^{\sim i}}$ denotes the expectation w.r.t. all output LLRs except l_i. We are now ready to state our main theorem on decay of correlations.

5.3 Main Theorem on Correlation Decay

Theorem 5.1 (Decay of Correlations for the MAP Decoder and LDPC codes)**.** *Consider transmission using a fixed LDPC code with bounded maximum degrees* $\mathtt{l}_{max}$ *and* $\mathtt{r}_{max}$*. Consider communication over a channel, from the class of channels* $\mathcal{K}$*, at low enough noise, i.e.,* $0 < \epsilon < \epsilon_g$*, where* $\epsilon_g > 0$ *depends only on* $\mathtt{l}_{max}$ *and* $\mathtt{r}_{max}$*. Then for any two code-bits with dist*$(i,j) > 4\,\mathtt{l}_{max}$ *we have*

$$\mathbb{E}_{\underline{l}}\big[|\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|\big] \leq c_1 e^{-\frac{\text{dist}(i,j)}{\xi(\epsilon)}}, \tag{5.4}$$

where c_1 *is a finite positive numerical constant and* $\xi(\epsilon)$ *is a strictly positive constant depending only on* ϵ*,* $\mathtt{l}_{max}$ *and* $\mathtt{r}_{max}$*.*

We will find that $\xi^{-1}(\epsilon)$ grows as $\epsilon \to 0$. Before we set out to prove this theorem, let us look at an important consequence.

5.3.1 Consequence of Decay of Correlation

Recall from 1.7 that the average MAP-GEXIT for any standard LDPC$(\Lambda(x), P(x))$ code ensemble is given by

$$g_n(\epsilon) = \mathbb{E}_{\mathcal{C}}\left[\int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim i}} \ln\left\{\frac{1+\langle x_i\rangle_0 \tanh l_i}{1+\tanh l_i}\right\}\right].$$

As in Section 4.3.1 from Chapter 4, we define the soft code-bit estimate of the BP decoder after d iterations as the Gibbs average on the computation tree, $T_d(i)$ of depth d. Recall that the average with respect to this Gibbs measure is denoted by $\langle - \rangle_d^{BP}$. Thus the BP-GEXIT function is given by

$$g_{n,d}^{BP}(\epsilon) = \mathbb{E}_{\mathcal{C}}\left[\int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim i}} \ln\left\{\frac{1+\langle x_i\rangle_{0,d}^{BP} \tanh l_i}{1+\tanh l_i}\right\}\right]. \tag{5.5}$$

The following corollary to the main theorem states that the asymptotic average MAP-GEXIT function can be computed by using the density evolution equation. Alternatively, the corollary proves that in the low noise regime, the replica solution evaluated at the appropriate fixed-point is exact.

Corollary 5.1 (Density Evolution Allows to Compute MAP-GEXIT for Ensemble of LDPC Codes)**.** *Consider communication over a channel belonging to the class* $\mathcal{K}$ *at low enough noise, i.e.,* $0 < \epsilon < \epsilon_g'$ *where* ϵ_g' *depends only on* $\mathtt{l}_{max}$*,* $\mathtt{r}_{max}$*. Then for LDPC*(Λ, P) *ensemble we have*

$$\lim_{n\to+\infty} g_n(\epsilon) = \lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon). \tag{5.6}$$

We point out that the above corollary is true even when there are jump discontinuities in the GEXIT functions, which would invalidate the use of area theorem to prove such a statement as done in [36]. It is well known [23] that with the presence of a non-zero fraction of degree one variable nodes, the estimates are non-zero and the corollary allows us to compute the asymptotic average MAP-GEXIT function via density evolution.

5.3.2 Strategy of Proof

The general idea of our strategy is to apply a *duality transformation* to the Gibbs measure (MAP measure) of the LDPC code. It turns out that the dual problem does not correspond to a well defined communications problem, and in fact it does not even correspond to a well defined Gibbs measure because the "weight" takes positive as well as *negative values*. Nevertheless, the dual problem has the flavor of a high noise LDGM system (or high temperature spin glass) and we are able to treat it through cluster expansions. There exist a host of such expansions [105], but we wish to stress that the simplest ones do not apply to the present situation for at least two reasons. The first is that there exist arbitrarily large portions of the dual system which are in a low noise (or low temperature) phase with positive probability. The second, is that the weights of the dual problem are not positive. As a result the cluster expansion method of the previous chapter as well as methods related to the Dobrushin criterion are ruled out.

It turns out that a cluster expansion originally devised by Berretti [116] is very well suited to overcome all these problems. We would like to point out here that the duality idea was used in the context of BEC to show that the asymptotic average MAP-GEXIT can be computed via density evolution in some regimes of noise for some class of LDPC code ensembles in [115]. The case of the BEC is special because under duality the Gibbs weight remains positive and the communication problem using LDPC codes on BEC(ϵ) transforms to a real communication problem using LDGM codes on the BEC($1-\epsilon$) [38], [115].

5.4 Duality Formulas

As explained in Section 5.3.2 we first transform the Gibbs measure of the LDPC code to a dual measure. The duality transformation reviewed here is an application of Poisson's summation formula over commutative groups, and has been thoroughly discussed in the context of codes on graphs in [117]. Here we need to know how the correlations transform under the duality, a point that does not seem to appear in the related literature.

Recall that for a linear code C with the codewords given by the $n-$tuples, $\underline{u} \in \{0,1\}^n$, the dual code, $C^\perp$ is defined as the set of $n-$tuples $\underline{v} \in \{0,1\}^n$ such that $\underline{u} \cdot \underline{v}^\top = 0$. We use the mapping $x_i = 1 - 2u_i$ and $\tau_i = 1 - 2v_i$ for $1 \le i \le n$ to denote the corresponding codewords in the spin language.

For any function $f : C \to \mathbb{R}$, the Poisson summation formula gives

$$\sum_{\underline{x} \in C} f(\underline{x}) = \frac{1}{|C^\perp|} \sum_{\underline{\tau} \in C^\perp} \widehat{f}(\underline{\tau}), \tag{5.7}$$

where the Fourier (or Hadamard) transform is given by,

$$\widehat{f}(\underline{\tau}) = \sum_{\underline{x}\in\{-1,+1\}^n} f(\underline{x}) e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^n (1-\tau_j)(1-x_j)}. \tag{5.8}$$

The dual code $C^\perp$ is an LDGM code with codewords given by $\underline{\tau}$ where

$$\tau_i = \prod_{a\in\partial i} u_a, \tag{5.9}$$

and u_a are the m information bits.

We now apply the Poisson formula to the partition function Z of the LDPC code with $f(\underline{x}) = \prod_{i=1}^n e^{l_i x_i}$. This yields the extended form of the MacWilliams identity,

$$Z = \frac{1}{|C^\perp|} e^{\sum_{j=1}^n l_j} Z_\perp, \tag{5.10}$$

where

$$Z_\perp = \sum_{\underline{u}\in\{-1,+1\}^m} \prod_{i=1}^n (1 + e^{-2l_i} \prod_{a\in\partial i} u_a). \tag{5.11}$$

For completeness, we provide a derivation of the Poisson summation formula and (5.10) in Appendix 5.A. The expression (5.10) formally looks like the partition function of an LDGM code with "channel half-loglikelihoods" g_i such that $\tanh g_i = e^{-2l_i}$. Indeed, we can write $(1 + e^{-2l_i} \prod_{a\in\partial i} u_a)$ as $e^{g_i u_{\partial i}}$ (upto a $\cosh g_i$ factor) and compare it with 1.25. This is truly the case only for the BEC(ϵ) where $l_i = 0, +\infty$ and hence $g_i = +\infty, 0$ which still correspond to a BEC($1-\epsilon$) and one recovers (taking the ϵ derivative of the logarithm of the partition functions) the well known duality relation between EXIT functions of a code and its dual on the BEC [38], [23]. For other channels however this is at best a formal (but still useful) analogy since the weights are negative for $l_i < 0$ (and g_i takes complex values). We introduce a bracket $\langle - \rangle_\perp$ which is not a true probabilistic expectation (but it is still linear)

$$\langle f \rangle_\perp = \frac{1}{Z_\perp} \sum_{\underline{u}\in\{-1,+1\}^m} f(\underline{u}) \prod_{i=1}^n (1 + e^{-2l_i} \prod_{a\in i} u_a). \tag{5.12}$$

The denominator may vanish, but it can be shown that when this happens the numerator also does so in a way that ensures the finiteness of the ratio (this becomes clear in subsequent calculations). Taking logarithm of (5.10) we find

$$\ln Z = -\ln |C^\perp| + \sum_{j=1}^n l_j + \ln Z_\perp.$$

We now take the the derivative with respect to l_i of the above equation to find,

$$\frac{\partial \ln Z}{\partial l_i} = 1 + \frac{\partial \ln Z_\perp}{\partial l_i}. \tag{5.13}$$

We now show how to compute $\frac{\partial \ln Z_\perp}{\partial l_i}$. For convenience we use τ_i to denote $\prod_{a\in\partial i} u_a$ (c.f. 5.9). From (5.11) we can write

$$\begin{aligned}\frac{\partial \ln Z_\perp}{\partial l_i} &= \frac{1}{Z_\perp}\sum_{\underline{\tau}} -2e^{-2l_i}\tau_i \prod_{j\neq i}(1+\tau_j e^{-2l_j}) \\ &= \frac{1}{Z_\perp}\sum_{\underline{\tau}} \frac{-2e^{-2l_i}}{1+\tau_i e^{-2l_i}}\tau_i \prod_{j=1}^{n}(1+\tau_j e^{-2l_j}) \\ &= \frac{1}{Z_\perp}\sum_{\underline{\tau}} \frac{-2e^{-2l_i}(1-\tau_i e^{-2l_i})}{1-e^{-4l_i}}\tau_i \prod_{j=1}^{n}(1+\tau_j e^{-2l_j}),\end{aligned} \tag{5.14}$$

where in the last equality we utilized $\tau_i^2 = 1$. Recall that $\langle - \rangle$ denotes the average with respect to the Gibbs measure corresponding to the LDPC partition function Z. Using the notation (5.12) we find

$$\begin{aligned}\langle x_i \rangle &= 1 - \frac{2e^{-2l_i}\langle \tau_i \rangle_\perp}{1-e^{-4l_i}} + \frac{2e^{-4l_i}}{1-e^{-4l_i}} \\ &= \frac{1}{\tanh 2l_i} - \frac{\langle \tau_i \rangle_\perp}{\sinh 2l_i}.\end{aligned} \tag{5.15}$$

Now differentiating once more with respect to l_j, $j \neq i$

$$\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = \frac{\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp}{\sinh 2l_i \sinh 2l_j}. \tag{5.16}$$

The second derivative is evaluated in the same way as the first derivative, and we defer its computation to the Appendix 5.B. We stress that in (5.15), (5.16), τ_i and τ_j are given by products of information bits (5.9). The left hand side of (5.15) is obviously bounded. It is less obvious to see this directly on the right hand side and here we just note that the pole at $l_i = 0$ is harmless since, for $l_i = 0$, the bracket has all its "weight" on configurations with $\tau_i = 1$. Similar remarks apply to (5.16). In any case, we will beat the poles by using the following trick. For any $0 < s < 1$ and $|a| \leq 1$ we have $|a| \leq |a|^s$. Also since $\langle - \rangle$ is a true Gibbs measure we must have $|\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle| \leq 2$. Combining we get

$$\mathbb{E}_{\underline{l}}\big[|\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle|\big] \leq 2^{1-s}\mathbb{E}_{\underline{l}}\big[|\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle|^s\big],$$

where recall that $\mathbb{E}_{\underline{l}}$ is the expectation w.r.t. the channel LLR realizations, $\underline{l}$. We stress here that $\langle - \rangle$ are true averages as opposed to $\langle - \rangle_\perp$. Using (5.16) and Cauchy-Schwarz we find

$$\mathbb{E}_{\underline{l}}\big[|\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle|\big] \leq 2^{1-s}\mathbb{E}[(\sinh 2l)^{-2s}]\ \mathbb{E}_{\underline{l}}\big[|\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp|^{2s}\big]^{1/2}. \tag{5.17}$$

In the above derivation we used the independence of the identically distributed random variables l_i and l_j and denote them simply by l. The prefactor is always

finite for $0 < s < \frac{1}{2}$ for our class of channels $\mathcal{K}$. For example for the BIAWGNC one can show that

$$\mathbb{E}[(\sinh 2l)^{-2s}] \leq \frac{c}{|1-2s|} e^{-c' \frac{s(1-2s)}{\epsilon^2}}, \tag{5.18}$$

for some positive constants c and c' and for the BSC we have

$$\mathbb{E}[(\sinh 2l)^{-2s}] \leq \left(\frac{2\epsilon(1-\epsilon)}{1-2\epsilon}\right)^{2s}. \tag{5.19}$$

5.5 Proof of Main Theorem on Decay of Correlations

We will prove the decay of correlations by applying a high temperature cluster expansion technique to estimate $\mathbb{E}_{\underline{l}}\big[|\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp|^{2s}\big]$. As explained in Section 5.3.2 we need a technique that does not use the positivity of the Gibbs weights. In Appendix 5.D we give a streamlined derivation of an adaptation of Berretti's expansion which gives us the following relation,

$$\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp = \frac{1}{2} \sum_{\hat{X}} K_{i,j}(\hat{X}) \Big(\frac{Z_\perp(\hat{X}^c)}{Z_\perp} \Big)^2, \tag{5.20}$$

where

$$K_{i,j}(\hat{X}) \equiv \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}}} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with} \hat{X}}} (\tau_i^{(1)} - \tau_i^{(2)})(\tau_j^{(1)} - \tau_j^{(2)}) \prod_{k \in \Gamma} E_k, \tag{5.21}$$

and

$$E_k = \tau_k^{(1)} e^{-2l_k} + \tau_k^{(2)} e^{-2l_k} + \tau_k^{(1)} \tau_k^{(2)} e^{-4l_k}. \tag{5.22}$$

Here $u_a^{(1)}$ and $u_a^{(2)}$ are two independent copies of the information bits (replicas) and $\tau_k^{(\alpha)} = \prod_{a \in k} u_a^{(\alpha)}$. Let us explain the various objects appearing in (5.20). Even though the relation (5.20) is in language of the dual code, to explain what are $\hat{X}$ and Γ we will refer to the Tanner graph of the LDPC code. To avoid confusion, we call the check nodes in the Tanner graph representing the LDPC code a-nodes. Also, we call the variable nodes in the Tanner graph representing the LDPC code i-nodes. In this language, i and j appearing in (5.20) are i-nodes. Given a subset S of nodes of the graph we denote by ∂S, the subset of neighboring nodes.

In (5.20), $\hat{X}$ is a set of distinct a-nodes (subset of the entire set of a-nodes) such that "$\hat{X}$ is connected via hyperedges": this means that

(a) $\hat{X} = \partial X$ for some connected subset X of i-nodes,

(b) X is connected. Connected here means that any pair of i-nodes in X can be joined by a path *all of whose i-nodes* lie in X and,

(c) $\hat{X}$ contains both ∂i and ∂j.

We remark that X need not contain the i-nodes i and j. In the sequel we will also refer to $\hat{X}$ as a *cluster* of a-nodes.

In the sum (5.21), Γ is a set of i-nodes (all distinct). We say that "Γ is compatible with $\hat{X}$" if:

(i) $\partial\Gamma \cup \partial i \cup \partial j = \hat{X}$,

(ii) $\partial\Gamma \cap \partial i \neq \phi$ and $\partial\Gamma \cap \partial j \neq \phi$. In words, $\partial\Gamma$ must intersect both ∂i and ∂j and,

(iii) there is a walk connecting ∂i and ∂j such that *all its i-nodes* are in Γ.

Finally we have,

$$Z_{\perp}(\hat{X}^c) = \sum_{\substack{u_a \\ a\in\hat{X}^c}} \prod_{\substack{\text{all } k \text{ s.t.} \\ \partial k\cap\hat{X}=\phi}} (1 + e^{-2l_k} \prod_{a\in k} u_a). \tag{5.23}$$

In words, the sum in (5.23) is carried over all a-nodes which do not belong to the set $\hat{X}$. The product is over all i-nodes such that they do not have an a-node (belonging to their neighborhood) in common with $\hat{X}$. The Figure 5.1 illustrates an example for all the sets appearing above.

We are now ready to prove the theorem on decay of correlations.

Proof of Theorem 5.1. Because of (5.17) it suffices to prove that $\mathbb{E}_{\underline{l}}\big[|\langle\tau_i\tau_j\rangle_{\perp} - \langle\tau_i\rangle_{\perp}\langle\tau_j\rangle_{\perp}|^{2s}\big]$ decays.

The first step is to prove

$$\left|\frac{Z_{\perp}(\hat{X}^c)}{Z_{\perp}}\right| \leq 1. \tag{5.24}$$

This ratio cannot be estimated directly because the weights in $Z_{\perp}$ are not positive. However we can use the duality transformation (5.10) backwards to get a new ratio of partition functions (corresponding to LDPC codes) with positive weights. More precisely, using the duality relation, we have

$$Z(\hat{X}^c) = \frac{1}{|C^{\perp}(\hat{X}^c)|}\Big(\prod_{\substack{\text{all } k \text{ s.t.} \\ \partial k\cap\hat{X}=\phi}} e^{l_k}\Big)\Big(Z_{\perp}(\hat{X}^c)\Big),$$

where

$$Z(\hat{X}^c) = \sum_{\substack{x_k \\ \partial k\cap\hat{X}=\phi}} \prod_{\substack{\text{all } k \text{ s.t} \\ \partial k\cap\hat{X}=\phi}} e^{l_k x_k} \prod_{a\in\hat{X}^c} \frac{1}{2}(1 + \prod_{\substack{k\in a \text{ and} \\ \partial k\cap\hat{X}=\phi}} x_k). \tag{5.25}$$

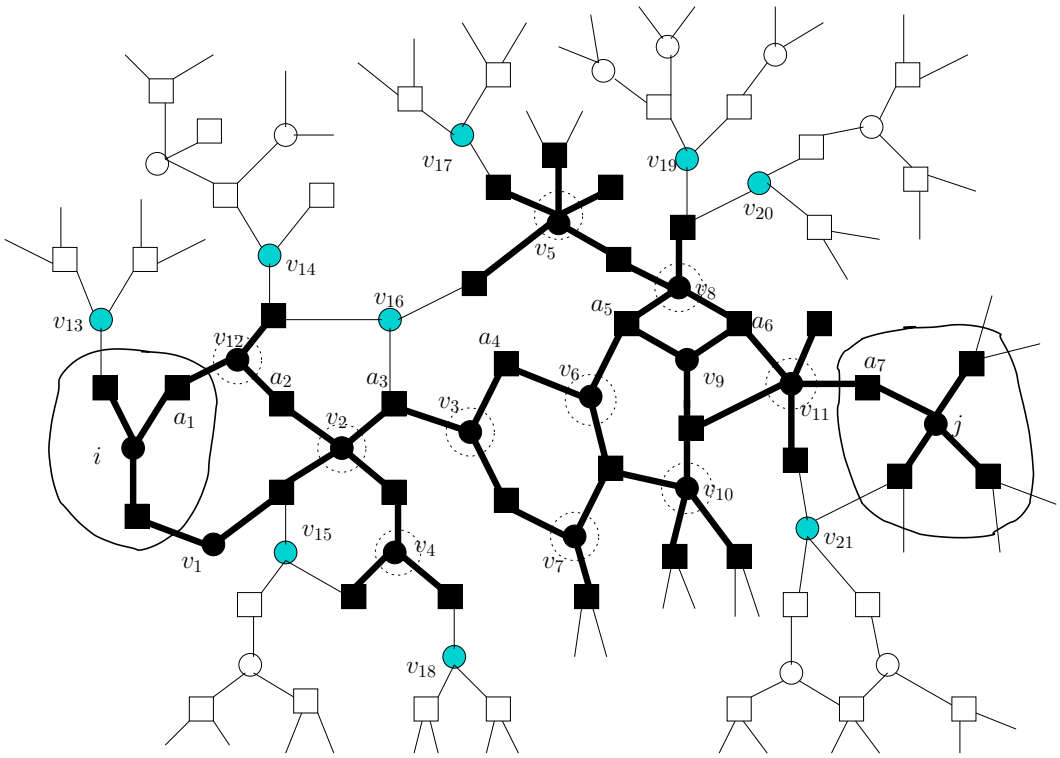

Figure 5.1: In this figure we explain the various sets appearing in the cluster expansion (5.20). The Tanner graph represents the LDPC code with variable nodes (i-nodes) denoted by circles and check nodes (a-nodes) denoted by squares. In this example the set $\hat{X}$ is the set of *dark* check nodes. Let us verify that this choice of $\hat{X}$ satisfies all our conditions. Let $X = \{i, j, v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}\}$ be a set of i-nodes (these are denoted by *dark* circles in the figure). It is easy to check that the set of neighbors of X is given by the *dark* check nodes which is equal to the set $\hat{X}$. Hence $\hat{X} = \partial X$ and we satisfy condition (a). Secondly, any two variable nodes in X are connected by a path all of whose variable nodes lie in X, thus condition (b) is met. Also notice that $\hat{X}$ contains both ∂i and ∂j, thus satisfying condition (c). Let us now show an example of the set Γ which appears in (5.21). Consider the set of distinct i-nodes, $\Gamma = \{v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{12}\}$. In the figure, the set of i-nodes belonging to Γ have a dotted circle around them. Let us verify that Γ is compatible with $\hat{X}$. We meet the first condition (i), since it is easy to check that the set of a-nodes in the union $\partial\Gamma \cup \partial i \cup \partial j$ equals $\hat{X}$. To see that the condition (ii) is satisfied, consider the i-nodes v_{12} and v_{11}. Clearly, $a_1 \in (\partial v_{12} \cap \partial i)$ and $a_7 \in (\partial v_{11} \cap \partial j)$. The walk $\{a_1 v_{12} a_2 v_2 a_3 v_3 a_4 v_6 a_5 v_8 a_6 v_{11} a_7\}$ connects ∂i and ∂j and all its i-nodes lie in Γ, hence the condition (iii) is also satisfied. Another choice for Γ would be the set $\{v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}\}$. In the definition of $Z_\perp(\hat{X}^c), Z(\hat{X}^c)$ the *light* variable nodes, $v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}$, are not present because they have a non-empty intersection with $\hat{X}$ (each of them has a *dark* a-node is their neighborhood). All check nodes which are *not dark* belong to $\hat{X}^c$.

This is the partition function corresponding to the subgraph induced by a-nodes of $\hat{X}^c$ and i-nodes such that $\partial i \cap \hat{X} = \phi$. Moreover $C^\perp(\hat{X}^c)$ is the dual of the code $C(\hat{X}^c)$.

Then combining with (5.10) we find,

$$\frac{Z_\perp(\hat{X}^c)}{Z_\perp} = \left(\exp \sum_{\substack{\text{all } k \text{ s.t} \\ \partial k \cap \hat{X} \neq \phi}} l_k\right)\left(\frac{|C^\perp(\hat{X}^c)|}{|C^\perp|}\right)\left(\frac{Z(\hat{X}^c)}{Z}\right), \tag{5.26}$$

To bound $\frac{Z_\perp(\hat{X}^c)}{Z_\perp}$ we first bound the ratio (of cardinality of the codes), $\frac{|C^\perp(\hat{X}^c)|}{|C^\perp|}$. The rank, $\mathfrak{L}$, of the parity-check matrix of code $C(\hat{X}^c)$, which is obtained by removing rows (checks) and columns (variables) from the parity-check matrix of the original LDPC code C, is smaller than the rank, $\mathfrak{T}$, of the parity-check matrix of C. Also set the block-length of the code $C(\hat{X}^c)$ to be n'. We have $|C| \cdot |C^\perp| = 2^n$ and $|C(\hat{X}^c)| \cdot |C^\perp(\hat{X}^c)| = 2^{n'}$. Dividing the two and using $|C(\hat{X}^c)| = 2^{n'-\mathfrak{L}}$ and $|C| = 2^{n-\mathfrak{T}}$ we get

$$\frac{|C^\perp(\hat{X}^c)|}{|C^\perp|} = \left(\frac{2^{n'}}{|C(\hat{X}^c)|}\right)\left(\frac{|C|}{2^n}\right) = 2^{n'-n}\left(\frac{2^{n-\mathfrak{T}}}{2^{n'-\mathfrak{L}}}\right) = 2^{\mathfrak{L}-\mathfrak{T}} \leq 1,$$

where in the last inequality we used that $\mathfrak{L} \leq \mathfrak{T}$.

Moreover we claim

$$\left(\exp \sum_{\substack{\text{all } k \text{ s.t} \\ \partial k \cap \hat{X} \neq \phi}} l_k\right) Z(\hat{X}^c) \leq Z. \tag{5.27}$$

To see this one must recognize that the left hand side of the inequality is the sum of terms of Z corresponding to $\underline{x}$ such that $x_k = +1$ for $\partial k \cap \hat{X} \neq \phi$ (and all other terms are ≥ 0). These remarks imply (5.24).

Using $|\sum_i a_i|^{2s} \leq \sum_i |a_i|^{2s}$ for $0 < 2s < 1$, and (5.24) we find

$$\begin{aligned}\mathbb{E}_{\underline{l}}[|\langle\tau_i\tau_j\rangle_\perp - \langle\tau_i\rangle_\perp\langle\tau_j\rangle_\perp|^{2s}] &= \frac{1}{2^{2s}}\mathbb{E}_{\underline{l}}\left|\sum_{\hat{X}} K_{i,j}(\hat{X})\Big(\frac{Z_\perp(\hat{X}^c)}{Z_\perp}\Big)^2\right|^{2s} \\ &\leq \frac{1}{2^{2s}}\sum_{\hat{X}} \mathbb{E}_{\underline{l}}\big[|K_{i,j}(\hat{X})|^{2s}\big]. \end{aligned} \tag{5.28}$$

Let us now provide an estimate on r.h.s. of the last inequality. Recall that

$$\begin{aligned} K_{i,j}(\hat{X}) &\equiv \sum_{\substack{u_a^{(1)},u_a^{(2)} \\ a\in\hat{X}}} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with}\hat{X}}} (\tau_i^{(1)} - \tau_i^{(2)})(\tau_j^{(1)} - \tau_j^{(2)}) \prod_{k\in\Gamma} E_k, \\ E_k &= \tau_k^{(1)} e^{-2l_k} + \tau_k^{(2)} e^{-2l_k} + \tau_k^{(1)}\tau_k^{(2)} e^{-4l_k}. \end{aligned} \tag{5.29}$$

To proceed further, first trivially bound the term $|(\tau_i^{(1)} - \tau_i^{(2)})(\tau_j^{(1)} - \tau_j^{(2)})|$ by 4 and $|E_k|$ by $|2e^{-2l_k} + e^{-4l_k}|$. Thus we obtain

$$\begin{aligned}\left|K_{i,j}(\hat{X})\right| &\leq 4 \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}}} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with} \hat{X}}} \prod_{k \in \Gamma} |2e^{-2l_k} + e^{-4l_k}| \\ &= 4^{|\hat{X}|+1} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with} \hat{X}}} \prod_{k \in \Gamma} |2e^{-2l_k} + e^{-4l_k}|,\end{aligned}$$

where, to get the last equality, we first notice that there is no more dependence on $u_a^{(1)}, u_a^{(2)}$ in the expression and we then use $\sum_{u_a^{(1)}, u_a^{(2)}} = 4^{|\hat{X}|}$ for $a \in \hat{X}$.

Then, we again use the identity $|2e^{-2l_k} + e^{-4l_k}|^{2s} \leq 2^{2s}e^{-4sl_k} + e^{-8sl_k}$ for $2s < 1$ to deduce

$$\begin{aligned}\mathbb{E}_{\underline{l}}\big[|K_{i,j}(\hat{X})|^{2s}\big] &\leq 4^{|\hat{X}|+1} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with} \hat{X}}} (2^{2s}\mathbb{E}[e^{-4sl}] + \mathbb{E}[e^{-8sl}])^{|\Gamma|} \\ &\leq 4^{|\hat{X}|+1} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with} \hat{X}}} \Big(\Delta(\epsilon)\Big)^{|\Gamma|}, \qquad (5.30)\end{aligned}$$

where in the first inequality we take the expectation inside the product because all the $k \in \Gamma$ are distinct and hence have independent l_k. We use the notation

$$\Delta(\epsilon) = 2^{2s}\mathbb{E}[e^{-4sl}] + \mathbb{E}[e^{-8sl}]. \qquad (5.31)$$

It only remains to bound the number of Γ in the sum (5.30) and the cardinatlity of Γ which is equal to $|\Gamma|$. First we give a bound on $|\Gamma|$.

Remember that our final aim is to make the r.h.s. of (5.30) small. Recall that the channel over which we are transmitting belongs to the class $\mathcal{K}$. From condition (3) of Definition 5.1 we can choose s and ϵ so that $\Delta(\epsilon)$ is less than one. To further estimate the bound in (5.30), we provide a *lower bound* on $|\Gamma|$. We do this as follows.

Since Γ is compatible with $\hat{X}$, we must have (from condition (i))

$$\partial\Gamma \cup \partial i \cup \partial j = \hat{X}.$$

As a consequence we obtain

$$|\partial\Gamma| \geq |\hat{X}| - |\partial i| - |\partial j|. \qquad (5.32)$$

Since the maximum degree of i-nodes is $\mathtt{l}_{\max}$ we have $|\partial\Gamma| \leq |\Gamma|\mathtt{l}_{\max}$, $|\partial i| \leq \mathtt{l}_{\max}$ and $|\partial j| \leq \mathtt{l}_{\max}$. Thus combining with (5.32), we get

$$|\Gamma| \geq (|\hat{X}| - 2\mathtt{l}_{\max})/\mathtt{l}_{\max}.$$

Let us now bound the number of Γ which are compatible with $\hat{X}$. Clearly, the maximum number of i-nodes which have an intersection with $\hat{X}$ is $|\hat{X}|\mathtt{r}_{\max}$.

Thus there are at most $2^{|\hat{X}|\mathtt{r}_{\max}}$ possible choices for Γ. Combining all the above we obtain

$$\mathbb{E}_{\underline{l}}\big[|K_{i,j}(\hat{X})|^{2s}\big] \leq 4^{(|\hat{X}|+1)}2^{\mathtt{r}_{\max}|\hat{X}|}\Big(\Delta(\epsilon)\Big)^{(|\hat{X}|-2\mathtt{l}_{\max})/\mathtt{l}_{\max}}. \tag{5.33}$$

From (5.28) and (5.33) we get

$$\mathbb{E}_{\underline{l}}\big[|\langle\tau_i\tau_j\rangle_\perp - \langle\tau_i\rangle_\perp\langle\tau_j\rangle_\perp|^{2s}\big] \leq \frac{1}{2^{2s}}\sum_{\hat{X}} 4^{(|\hat{X}|+1)}2^{\mathtt{r}_{\max}|\hat{X}|}\Delta(\epsilon)^{(|\hat{X}|-2\mathtt{l}_{\max})/\mathtt{l}_{\max}}. \tag{5.34}$$

The clusters $\hat{X}$ connect ∂i and ∂j and thus have sizes $|\hat{X}| \geq \frac{1}{2}\mathrm{dist}(i,j)$.

Moreover we claim

Lemma 5.1. *The number of clusters of a given size, $|\hat{X}|$, which satisfies conditions (a), (b), (c), is upper bounded by $\boldsymbol{r}_{max}|\hat{X}|(3^{K\boldsymbol{r}_{max}|\hat{X}|})$ where $K = \boldsymbol{l}_{max}\boldsymbol{r}_{max}$ is fixed for a given LDPC code.*

We prove this claim in the Appendix 5.C. Combining with Lemma 5.1 we can arrange the sum in (5.34) as

$$\begin{aligned}\mathbb{E}_{\underline{l}}\big[|\langle\tau_i\tau_j\rangle_\perp - \langle\tau_i\rangle_\perp\langle\tau_j\rangle_\perp|^{2s}\big] &\leq \frac{1}{2^{2s}}\sum_{|\hat{X}|\geq \mathrm{dist}(i,j)/2} \mathtt{r}_{\max}|\hat{X}|(3^{K\mathtt{r}_{\max}|\hat{X}|})4^{(|\hat{X}|+1)}2^{\mathtt{r}_{\max}|\hat{X}|} \\ &\qquad \times \Delta(\epsilon)^{(|\hat{X}|-2\mathtt{l}_{\max})/\mathtt{l}_{\max}} \\ &\leq \frac{\mathtt{r}_{\max}\big(\Delta(\epsilon)\big)^{-2}}{2^{2s}}\sum_{|\hat{X}|\geq \mathrm{dist}(i,j)/2} \exp\Bigg\{|\hat{X}|\Big(\mathtt{r}_{\max}K\ln 3 + \ln 4 + \mathtt{r}_{\max}\ln 2 \\ &\qquad + \frac{\ln\Big(\Delta(\epsilon)\Big)}{\mathtt{l}_{\max}} + 1\Big)\Bigg\}\end{aligned} \tag{5.35}$$

For channels belonging to the class $\mathcal{K}$ we have, for s small enough, $\mathbb{E}[e^{-sl}] \to 0$ as $\epsilon \to 0$. As a consequence we chose ϵ small enough to make $\Delta(\epsilon)$ small enough. Thus we can make the exponent negative and we conclude the proof. □

5.6 Density Evolution Equals MAP for Low Noise

By the same arguments than in Section 4.4.1 we have for LDPC codes, the average MAP-GEXIT function,

$$g_n(\epsilon) = \mathbb{E}_{\mathcal{C}}\left[\int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{\underline{l}^{\sim i}}\ln\left\{\frac{1+\langle x_i\rangle_0\tanh l_i}{1+\tanh l_i}\right\}\right]. \tag{5.36}$$

For fixed d number of iterations, the average BP-GEXIT function in the limit $n \to +\infty$ is given by

$$\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \mathbb{E}_C \Bigg[\int dl_i \frac{dc(l_i)}{d\epsilon} \mathbb{E}_{\underline{l}^{\sim i}} \ln \Bigg\{ \frac{1 + \langle x_i \rangle_{0,N_d(i)} \tanh l_i}{1 + \tanh l_i} \Bigg\} \Bigg| N_d(i) \text{ is a tree} \Bigg]. \tag{5.37}$$

Above, $\langle x_i \rangle_{0,N_d(i)}$ is the Gibbs bracket associated to the graph $N_d(i)$. When $N_d(i)$ is a tree, the set of leaves in the boundary of $N_d(i)$ are variable nodes. Thus the boundary conditions of $N_d(i)$ are given by the channel outputs. In the sequel, we will call such a boundary condition also as "free" or "natural" boundary. We will also abuse the notation introduced in Section 1.1.3 of Chapter and denote $\mathring{N}_d(i)$ as the variable nodes at a distance d from the root node. Thus $\mathring{N}_d(i)$ denotes the boundary of variable nodes (leaves) of $N_d(i)$. On the tree we evaluate $\langle x_i \rangle_{0,N_d(i)}$ using density evolution equations to get

$$\lim_{d\to+\infty} \lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \lim_{d\to\infty} \int dl \frac{dc(l)}{d\epsilon} \mathbb{E}_{\Delta^{(d)}} \left[\ln \left\{ \frac{1 + \tanh \Delta^{(d)} \tanh l}{1 + \tanh l} \right\} \right], \tag{5.38}$$

where both limits exist and

$$\Delta^{(d)} = \sum_{a=1}^{k} u_a^{(d)}. \tag{5.39}$$

The $u_a^{(d)}$ are i.i.d. random variables with distribution obtained from the iterative system of DE equations

$$\widehat{\eta}^{(d)}(u) = \sum_l \lambda_l \int \prod_{j=1}^{l-1} dv_j \, \eta^{(d)}(v_j) \, \delta(u - \tanh^{-1}(\prod_{j=1}^{l-1} \tanh v_j)),$$
$$\eta^{(d)}(v) = \sum_k \rho_k \int dl \, c(l) \prod_{a=1}^{k-1} du_a \, \widehat{\eta}^{(d-1)}(u_a) \, \delta(v - l - \sum_{a=1}^{k-1} u_a^{(d)}).$$

The initial condition is given by $\eta^{(0)}(v) = c(v)$. Recall that these equations are an iterative version of the replica fixed point equation [66], [67].

Proof of Corollary 5.1. The first few steps are the same as in the proof for LDGM. First, we expand the logarithm in (5.36) and use Nishimori identities to obtain a series expansion like (4.25). Second, we notice that since the resulting series expansion is uniformly absolutely convergent it is again enough to show that

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} [\langle x_i \rangle_0^{2p} \mid N_d(i) \text{ tree}] = \lim_{d\to+\infty} \lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}} [\langle x_i \rangle_{0,N_d(i)}^{2p} | \mid N_d(i) \text{ tree}]. \tag{5.40}$$

Again, because of $|b^{2p}-a^{2p}| \le 2p|b-a|$ it is enough to show the last equation for $2p$ replaced by 1. Unfortunately, one cannot proceed as simply as in the LDGM case. We prove (5.40) as a consequence of the next two lemmas stated below. Let $\langle - \rangle^{\infty}_{0;N_d(i)}$ be the bracket defined on the subgraph $N_d(i)$ with $l_k = +\infty$ for $k \in \mathring{N}_d(i)$ and $l_i = 0$. This in fact is formally equivalent to fixing $x_k = +1$ boundary conditions on the leaves of the tree $k \in \mathring{N}_d(i)$. The first lemma says that the bit estimate can be computed locally.

Lemma 5.2. *Under the same conditions as in Corollary 5.1,*

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}}\big[\langle x_i\rangle_0 \mid N_d(i)\ tree\big] = \lim_{d\to+\infty}\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}}\big[\langle x_i\rangle^{\infty}_{0,N_d(i)} \mid N_d(i)\ tree\big]. \tag{5.41}$$

The second lemma says that at low enough noise, free and +1 boundary conditions are equivalent

Lemma 5.3. *Under the same conditions as in Corollary 5.1,*

$$\lim_{d\to+\infty}\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}}\big[\langle x_i\rangle^{\infty}_{0,N_d(i)} \mid N_d(i)\ tree\big] \tag{5.42}$$

$$= \lim_{d\to+\infty}\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},\underline{l}^{\sim i}}\big[\langle x_i\rangle_{0,N_d(i)} \mid N_d(i)\ tree\big]. \tag{5.43}$$

We prove the first lemma. It will then be clear that the proof of the second one is essentially the same except that the original full graph is replaced by $N_d(i)$, and thus it will be spared.

Proof of Lemma 5.2: In (5.41) (and (5.42)) the root node i has $l_i = 0$ which turns out to be technically cumbersome. For this reason we use the identities

$$\langle x_i\rangle = \frac{\langle x_i\rangle_0 + \tanh l_i}{1+\langle x_i\rangle_0 \tanh l_i}, \qquad \langle x_i\rangle^{\infty}_{N_d(i)} = \frac{\langle x_i\rangle^{\infty}_{0,N_d(i)} + \tanh l_i}{1+\langle x_i\rangle^{\infty}_{0,N_d(i)} \tanh l_i}, \tag{5.44}$$

to deduce

$$\langle x_i\rangle_0 - \langle x_i\rangle^{\infty}_{0,N_d(i)} = \frac{(1-(\tanh l_i)^2)(\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)})}{(1-\langle x_i\rangle \tanh l_i)(1-\langle x_i\rangle^{\infty}_{N_d(i)} \tanh l_i)}. \tag{5.45}$$

This implies

$$|\langle x_i\rangle_0 - \langle x_i\rangle^{\infty}_{0;N_d(i)}| \le \frac{1+|\tanh l_i|}{1-|\tanh l_i|}\,|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}|. \tag{5.46}$$

Averaging over the noise and using Cauchy-Schwartz,

$$\begin{aligned}\mathbb{E}_{\underline{l}^{\sim i}}\big[|\langle x_i\rangle_0 - \langle x_i\rangle^{\infty}_{0;N_d(i)}|\big] &\le 2\mathbb{E}[e^{4|l|}]^{1/2}\mathbb{E}_{\underline{l}}\big[|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}|^2\big]^{1/2} \\ &\le 2\sqrt{2}\,\mathbb{E}[e^{4|l|}]^{1/2}\,\mathbb{E}_{\underline{l}}\big[|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}|\big]^{1/2}. \end{aligned} \tag{5.47}$$

Notice that we have removed the square in the second expectation above and have a multiplicative factor of 2 in front. This is easy to see because $|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}| \leq 2$. Let us now prove

$$\lim_{d\to+\infty}\lim_{n\to+\infty}\mathbb{E}_{\mathcal{C},\underline{l}}\big[|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}| \mid N_d(i)\,\text{tree}\big] = 0. \tag{5.48}$$

We order the variable nodes at the boundary $\mathring{N}_d(i)$. Let $\mathring{l}_d$ denote the vector of LLR with components $\in \mathring{N}_d(i)$. We write $\langle - \rangle^{\infty}_{\leq k-1}$ to denote the average when the first $k-1$ components of $\mathring{l}_d$ are set as $l_1 = ... = l_{k-1} = +\infty$, the k-th component is l'_k, and the remaining ones are i.i.d. and are distributed as $c(l)$ (in other words they are "free"). From the fundamental theorem of calculus, it is not difficult to see that

$$\begin{aligned}\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)} &= -\sum_{k\in\mathring{N}_d(i)}\int_{l_k}^{+\infty} dl'_k\,\frac{d\langle x_i\rangle^{\infty}_{\leq k-1}}{dl'_k}\\ &= -\sum_{k\in\mathring{N}_d(i)}\int_{l_k}^{+\infty} dl'_k\,(\langle x_i x_k\rangle^{\infty}_{\leq k-1} - \langle x_i\rangle^{\infty}_{\leq k-1}\langle x_k\rangle^{\infty}_{\leq k-1}).\end{aligned} \tag{5.49}$$

Using $|a| \leq |a|^s$, for any $0 < s < 1$ and $|a| \leq 1$, for $a = \frac{1}{2}(|\langle x_i x_k\rangle^{\infty}_{\leq k-1} - \langle x_i\rangle^{\infty}_{\leq k-1}\langle x_k\rangle^{\infty}_{\leq k-1}|)$ we get

$$\begin{aligned}&|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}|\\ &\qquad\leq 2^{1-s}\sum_{k\in\mathring{N}_d(i)}\int_{l_k}^{+\infty} dl'_k\,|\langle x_i x_k\rangle^{\infty}_{\leq k-1} - \langle x_i\rangle^{\infty}_{\leq k-1}\langle x_k\rangle^{\infty}_{\leq k-1}|^s.\end{aligned} \tag{5.50}$$

Let $\langle - \rangle^{\infty,\perp}_{\leq k-1}$ be the Gibbs bracket corresponding to the dual code (with the first k components of $\mathring{N}_d(i)$ set as $l_1 = ... = l_{k-1} = +\infty$ and the k-th component equal to l'_k). Because of (5.16) we have

$$\begin{aligned}&|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}|\\ &\qquad\leq 2^{1-s}\sum_{k\in\mathring{N}_d(i)}\int_{l_k}^{+\infty} dl'_k\,\frac{|\langle \tau_i\tau_k\rangle^{\infty,\perp}_{\leq k-1} - \langle \tau_i\rangle^{\infty,\perp}_{\leq k-1}\langle \tau_k\rangle^{\infty,\perp}_{\leq k-1}|^s}{(|\sinh 2l_i||\sinh 2l'_k|)^s}.\end{aligned} \tag{5.51}$$

Note that the denominator in the integral is important to make the integral convergent for $l'_k \to \infty$. Moreover at l_i and $l'_k = 0$ the hyperbolic sin terms are harmless as long as for $s < 1$. The next step is to use the cluster expansion to estimate

$$\mathbb{E}_{\underline{l}}\left[\int_{l_k}^{+\infty} dl'_k\,\frac{|\langle \tau_i\tau_k\rangle^{\infty,\perp}_{\leq k-1} - \langle \tau_i\rangle^{\infty,\perp}_{\leq k-1}\langle \tau_k\rangle^{\infty,\perp}_{\leq k-1}|^s}{(|\sinh 2l_i||\sinh 2l'_k|)^s}\right]. \tag{5.52}$$

By following similar steps than in the proof of Theorem 5.1 one obtains an upper bound similar to (5.33) except that the likelihoods of the end points are

weighted differently. More precisely, we get

$$\mathbb{E}_{\underline{l}}\left[\int_{l_k}^{+\infty} dl'_k \frac{|\langle \tau_i\tau_k\rangle^{\infty,\perp}_{\leq k-1} - \langle\tau_i\rangle^{\infty,\perp}_{\leq k-1}\langle\tau_k\rangle^{\infty,\perp}_{\leq k-1}|^s}{(|\sinh 2l_i||\sinh 2l'_k|)^s}\right]$$
$$\leq \mathbb{E}_{\underline{l}}\left[\int_{l_k}^{+\infty} dl'_k \frac{\sum_{\hat{X}} 4^{|\hat{X}|+1}\sum_{\Gamma \text{ compatible with}\hat{X}}\prod_{j\in\Gamma}(2^s e^{-2sl_j}+e^{-4sl_j})}{(|\sinh 2l_i||\sinh 2l'_k|)^s}\right].$$

Now we separate out the terms corresponding to l_i and l_k (using their independence) to further upper bound the last expression above by

$$T_iT_k\,\mathbb{E}_{\underline{l}^{\sim ik}}\left[\sum_{\hat{X}} 4^{|\hat{x}|+1}\sum_{\Gamma \text{ compatible with}\hat{X}}\prod_{j\in\Gamma\setminus(l_i,l_k)}(2^s e^{-2sl_j}+e^{-4sl_j})\right], \tag{5.53}$$

where

$$T_i = \max\left\{\mathbb{E}_{l_i}\left[\frac{2^s e^{-2sl_i}+e^{-4sl_i}}{|\sinh 2l_i|^s}\right], \mathbb{E}_{l_i}\left[\frac{1}{|\sinh 2l_i|^s}\right]\right\},$$
$$T_k = \max\left\{\mathbb{E}_{l_k}\left[\int_{l_k}^{+\infty} dl'_k \frac{(2^s e^{-2sl_k}+e^{-4sl_k})}{|\sinh 2l'_k|^s}\right], \mathbb{E}_{l_k}\left[\int_{l_k}^{+\infty} dl'_k \frac{1}{|\sinh 2l'_k|^s}\right]\right\}.$$

Recall that $\mathbb{E}_{\underline{l}^{\sim ik}}$ denotes the expectation w.r.t. all the LLRs except l_i and l_k. The terms T_i, T_k arise because any particular Γ may or may not include the nodes i and/or k.

Since the i-nodes in Γ are distinct, we can again take the expectation inside the product for the third term in (5.53). Thus we proceed along the same lines as in (5.33) and (5.34) to finally get an upper bound for this term similar to (5.35). Again, all the previous manipulations are possible because our channel belongs to class $\mathcal{K}$.

Also for our class $\mathcal{K}$ we can show

$$\mathbb{E}\left[\frac{2^s e^{-2sl}+e^{-4sl}}{|\sinh 2l|^s}\right] < \infty \qquad \text{and} \qquad \mathbb{E}\left[\int_l^{+\infty} dl' \frac{2^s e^{-2sl'}+e^{-4sl'}}{|\sinh 2l'|^s}\right] < \infty. \tag{5.54}$$

which implies that the terms T_i and T_k are finite and depend only on the value of ϵ. Putting everything together we find

$$\mathbb{E}_{\underline{l}}\big[|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}|\big] \leq 2^{1-s}\sum_{k\in\mathring{N}_d(i)} c_1 e^{-\frac{d}{\xi(\epsilon)}} \leq c_1 2^{1-s}K^d e^{-\frac{d}{\xi(\epsilon)}}. \tag{5.55}$$

We obtain the final inequality by upper bounding the number of i-nodes in the boundary $\mathring{N}_d(i)$ by K^d. But for our class of channels $\mathcal{K}$ we can choose ϵ small enough so that $Ke^{-\frac{1}{\xi(\epsilon)}} < 1$. This immediately implies (5.48) and the Corollary 5.1 follows. □

5.7 Discussion

In this chapter we use duality to convert the LDPC coded system to a system which has the flavor of an LDGM (or equivalently "high temperature") coded system. A natural question would be whether one could use similar techniques to show decay of correlations for LDGM coded system at low noise. Unfortunately, using duality in this case, will give us an LDPC coded system, which is inherently at "low temperature". This would render the cluster expansion technique useless.

Of course, the biggest issue here would be to widen the regime of noise for which the correlations decay. We do not attempt to optimize the estimates/counting arguments. This would clearly be one way to increase the noise regime. We know in few instances, like regular LDPC code ensembles, that below ϵ_{BP}, both the asymptotic average MAP and BP-GEXIT curves are zero. This leads us to believe that the correlations should decay exponentially (in fact faster than the local graph expansion) till ϵ_{BP}.

We saw in the previous chapter, that in the special case of LDGM codes transmitted over the BSC, one could exchange the limits $d \to +\infty, n \to +\infty$, while computing the average BP-GEXIT function. It would nice to have such a statement for the case of LDPC codes. A final aim would be to provide such an exchange of limit result for the bit error rate of the BP decoder.

5.A Poisson Summation Formula and its Application

We re-state the Poisson summation rule here. Let C be a linear code of block-length n and let $C^{\perp}$ denote the dual code. We use $x_i = 1 - 2u_i$ ($\tau_i = 1 - 2v_i$) to map codewords $\underline{u}$ ($\underline{v}$) of C ($\mathcal{C}^{\perp}$) belonging to the space $\{0,1\}^n$ to the space $\{+1,-1\}^n$. Abusing notation, we let C also denote the set of codewords after being mapped to the space $\{+1,-1\}^n$. For any $f : C \to \mathbb{R}$ we have

$$\sum_{\underline{x}\in C} f(\underline{x}) = \frac{1}{|C^{\perp}|} \sum_{\underline{\tau}\in C^{\perp}} \widehat{f}(\underline{\tau}), \tag{5.56}$$

where the Fourier transform, $\widehat{f}(\cdot)$, is given by,

$$\widehat{f}(\underline{\tau}) = \sum_{\underline{x}\in\{-1,+1\}^n} f(\underline{x}) e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)}. \tag{5.57}$$

Note that $\mathrm{i}^2 = -1$. Let us start with the r.h.s. of (5.56). Using (5.57) we get

$$\begin{aligned}\frac{1}{|C^{\perp}|}\sum_{\underline{\tau}\in C^{\perp}} \widehat{f}(\underline{\tau}) &= \frac{1}{|C^{\perp}|}\sum_{\underline{\tau}\in C^{\perp}}\sum_{\underline{x}\in\{-1,+1\}^n} f(\underline{x}) e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)} \\ &= \frac{1}{|C^{\perp}|}\sum_{\underline{x}\in\{-1,+1\}^n} f(\underline{x}) \sum_{\underline{\tau}\in C^{\perp}} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)}. \end{aligned} \tag{5.58}$$

We claim that

$$\sum_{\underline{\tau}\in C^{\perp}} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)} = \begin{cases} |C^{\perp}|, & \underline{x}\in C, \\ 0, & \underline{x}\notin C. \end{cases}$$

If we assume the above claim to be true then we immediately obtain the Poisson summation rule. Let us now prove the claim. First assume that $\underline{x} \in C$. From the definition of duality we have

$$\underline{u}\cdot\underline{v}^{\top} = 0 \Rightarrow \quad \sum_{i=1}^{n} \frac{(1-x_i)}{2}\frac{(1-\tau_i)}{2} = \text{even}.$$

The sum in the expression above is carried over $\mathbb{R}$. Consequently we must have $e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)} = 1$ for each $\underline{\tau} \in C^{\perp}$. Thus we get the contribution of $|C^{\perp}|$ when $\underline{x} \in C$. Now assume that $\underline{x} \notin C$. This implies there must exist a dual codeword $\underline{\tau}'$ such that $\sum_{i=1}^{n} \frac{(1-x_i)}{2}\frac{(1-\tau_i)}{2} = \text{odd}$. We claim that

$$\sum_{\underline{\tau}\in C^{\perp}} e^{\mathrm{i}\frac{\pi}{4}\left(\sum_{j=1}^{n}(1-\tau_j)(1-x_j)+(1-\tau'_j)(1-x_j)\right)} = \sum_{\underline{\tau}\in C^{\perp}} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)}. \tag{5.59}$$

If we assume for the moment that this claim is true then using, $e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau'_j)(1-x_j)} = e^{\mathrm{i}\pi \text{ (odd number)}} = -1$, we obtain

$$-\sum_{\underline{\tau}\in C^\perp} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)} = \sum_{\underline{\tau}\in C^\perp} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)},$$

which implies that $\sum_{\underline{\tau}\in C^\perp} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-\tau_j)(1-x_j)} = 0$ for any $\underline{x} \notin C$.

It remains now to prove (5.59). Notice that the term $\frac{1-\tau_j}{2} + \frac{1-\tau'_j}{2}$ equals 0 or 2 when $\tau_j = \tau'_j$ and equals 1 when $\tau_j \neq \tau'_j$. Whenever $\frac{1-\tau_j}{2} + \frac{1-\tau'_j}{2} = 2$, we can replace it by a 0 (since both 2 and 0 are even) and it would not make a difference to the sum $\sum_{\underline{\tau}\in C^\perp} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-x_j)[(1-\tau_j)+(1-\tau'_j)]}$. Thus we can write

$$\begin{aligned}\sum_{\underline{\tau}\in C^\perp} e^{\mathrm{i}\frac{\pi}{4}\left(\sum_{j=1}^{n}(1-\tau_j)(1-x_j)+(1-\tau'_j)(1-x_j)\right)} &= \sum_{\underline{\tau}\in C^\perp} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-x_j)[(1-\tau_j)+(1-\tau'_j)]} \\ &= \sum_{\underline{\tau}\in C^\perp} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-x_j)[(1-\tau_j\tau'_j)]} \\ &= \sum_{\underline{\tau}''\in C^\perp} e^{\mathrm{i}\frac{\pi}{4}\sum_{j=1}^{n}(1-x_j)[(1-\tau''_j)]}.\end{aligned}$$

The second equality above follows because $\left(\frac{1-\tau_j}{2}+\frac{1-\tau'_j}{2}\right) \mod 2 = \frac{1-\tau_j\tau'_j}{2}$. The last equality follows because $\tau''_j \triangleq \tau_j\tau'_j = \prod_{a\in\partial j} u_a u'_a$ implies that $\underline{\tau}'' \in C^\perp$. Hence the claim. □

Is is easy to see that the partition function Z, of an LDPC code C, can be written as $\sum_{\underline{x}\in C} f(\underline{x})$ with $f(\underline{x}) = \prod_{i=1}^{n} e^{l_i x_i}$. Thus applying the Poisson summation rule we get

$$\begin{aligned}Z &= \frac{1}{C^\perp}\sum_{\underline{\tau}\in C^\perp}\sum_{\underline{x}\in\{+1,-1\}^n}\prod_{j=1}^{n} e^{l_j x_j}\prod_{j=1}^{n} e^{\mathrm{i}\frac{\pi}{4}(1-\tau_j)(1-x_j)} \\ &= \frac{1}{C^\perp}\sum_{\underline{\tau}\in C^\perp}\sum_{\underline{x}\in\{+1,-1\}^n}\prod_{j=1}^{n} e^{l_j x_j + \mathrm{i}\frac{\pi}{4}(1-\tau_j)(1-x_j)}.\end{aligned}$$

Exchanging the sum over $\underline{x} \in \{+1,-1\}^n$ and the product over j and using $e^{\mathrm{i}\frac{\pi}{2}(1-\tau_j)} = \tau_j$ we obtain

$$\begin{aligned}Z &= \frac{1}{C^\perp}\sum_{\underline{\tau}\in C^\perp}\prod_{j=1}^{n}\left(e^{l_j} + e^{-l_j+\mathrm{i}\frac{\pi}{2}(1-\tau_j)}\right) \\ &= \frac{1}{C^\perp}e^{\sum_{j=1}^{n} l_j}\sum_{\underline{\tau}\in C^\perp}\prod_{j=1}^{n}(1+e^{-2l_j}\tau_j) \\ &= \frac{1}{C^\perp}e^{\sum_{j=1}^{n} l_j}\sum_{\underline{u}\in\{+1,-1\}^m}\prod_{j=1}^{n}(1+e^{-2l_j}\prod_{a\in j} u_a),\end{aligned}$$

where $\underline{u}$ is the vector of m information bits. □

5.B Second Derivative Computation

Here we provide a quick derivation of (5.16). Taking the derivative of $\langle \tau_i \rangle_\perp$ with respect to l_j we find

$$\frac{\partial \langle \tau_i \rangle_\perp}{\partial l_j} = \frac{1}{Z_\perp} \frac{\partial}{\partial l_j} \left(\sum_{\underline{\tau}} \tau_i \prod_{k=1}^{n} (1 + \tau_k e^{-2l_k}) \right) + \left(\sum_{\underline{\tau}} \tau_i \prod_{k=1}^{n} (1 + \tau_k e^{-2l_k}) \right) \frac{\partial}{\partial l_j} \left(\frac{1}{Z_\perp} \right). \tag{5.60}$$

For the first term on the r.h.s. of (5.60), we proceed as (5.14) to get

$$\frac{\partial}{\partial l_j} \left(\sum_{\underline{\tau}} \tau_i \prod_{k=1}^{n} (1 + \tau_k e^{-2l_k}) \right) = \sum_{\underline{\tau}} \frac{-2\tau_j e^{-2l_j} (1 - \tau_j e^{-2l_j})}{1 - e^{-4l_j}} \tau_i \prod_{k=1}^{n} (1 + \tau_k e^{-2l_k}).$$

Combining with $1/Z_\perp$, we find the first term to be equal to

$$-\frac{\langle \tau_i \tau_j \rangle_\perp}{\sinh 2l_j} + \frac{2e^{-4l_j} \langle \tau_i \rangle_\perp}{1 - e^{-4l_j}}. \tag{5.61}$$

The second term in (5.60) is equal to

$$\begin{aligned} \frac{-1}{Z_\perp^2} \left(\frac{\partial Z_\perp}{\partial l_j} \right) &= \frac{-1}{Z_\perp^2} \sum_{\underline{\tau}} -2e^{-2l_j} \tau_j \prod_{k \neq j} (1 + \tau_k e^{-2l_k}) \\ &= \frac{-1}{Z_\perp^2} \sum_{\underline{\tau}} \frac{-2\tau_j e^{-2l_j} (1 - \tau_j e^{-2l_j})}{1 - e^{-4l_j}} \prod_{k=1}^{n} (1 + \tau_k e^{-2l_k}) \\ &= \frac{1}{Z_\perp} \left(\frac{\langle \tau_j \rangle_\perp}{\sinh 2l_j} \right) + \frac{-1}{Z_\perp} \left(\frac{2e^{-4l_j}}{1 - e^{-4l_j}} \right). \end{aligned}$$

Combining with $\left(\sum_{\underline{\tau}} \tau_i \prod_{k=1}^{n} (1 + \tau_k e^{-2l_k}) \right)$, the second term is equal to

$$\frac{\langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp}{\sinh 2l_j} - \frac{2e^{-4l_j} \langle \tau_i \rangle_\perp}{1 - e^{-4l_j}}. \tag{5.62}$$

Thus putting (5.61) and (5.62) together we obtain (5.16). □

5.C Proof of Lemma 5.1

Before we commence, let us recall that a-nodes and i-nodes are the check nodes and variable nodes, respectively, of the Tanner graph of the LDPC code. Recall that $\mathtt{l}_{\max}$ denotes the maximum degree of i-nodes and $\mathtt{r}_{\max}$ denotes the maximum degree of a-nodes.

Let L be a positive integer. Our aim is to count the number of $\hat{X}$ such that $|\hat{X}| = L$ and the conditions (a), (b) and (c) given in Section 5.5 are satisfied.

Instead, we will count the number of $\hat{X}$ such that $|\hat{X}| = L$ and which satisfy conditions (a) and (b). But we only require that it contains ∂i. It may or may not contain ∂j. Clearly, this will be an upper bound to our required number.

We know that the required $\hat{X}$ corresponds to some set of i-nodes X such that $\partial X = \hat{X}$ (condition (a) in Section 5.5). Also X is connected, i.e., there is a walk between any two i-nodes contained in X. Note that X may or may not contain the i-node i. Let us modify this X to include the i-node (if it is not already present). Let us denote this set by $\mathring{X}$. We point out two properties of the set $\mathring{X}$: (i) It is connected. Indeed, any two i-nodes, neither of which is i, are connected (follows from previous connectivity in X). Let $\partial^2 i$ denote the neighborhood of the set ∂i. It is clear that $\partial^2 i$ is a set of i-nodes. Since $\hat{X}$ includes ∂i, it must be that X contains at least one i-node which belongs to $\partial^2 i$. But i is already present in $\partial^2 i$. Consequently, i is connected to any other i-node in $\mathring{X}$ and (ii) The size of $\mathring{X}$ is bounded as, $\frac{L}{\mathtt{l}_{\max}} \leq |\mathring{X}| \leq L\mathtt{r}_{\max}$.

Thus for each $\hat{X}$ we have constructed an $\mathring{X}$, such that $\partial\mathring{X} = \hat{X}$, $i \in \mathring{X}$, $\mathring{X}$ is connected and $|\mathring{X}| \in \left[\frac{L}{\mathtt{l}_{\max}}\ ,\ L\mathtt{r}_{\max}\right]$. Suppose that $\hat{X}_1 \neq \hat{X}_2$ are two sets with size equal to L and both satisfy our requirements. Then it is clear that the corresponding $\mathring{X}_1$ and $\mathring{X}_2$ are also different. As a result, counting such $\mathring{X}$ will provide us with an upper bound on the number of required $\hat{X}$ of given size L.

Without loss of generality we consider the following graph, $\mathcal{G}$, to count the number of $\mathring{X}$. The nodes in $\mathcal{G}$ consists of the i-nodes in the original graph. We attach an edge between two nodes in $\mathcal{G}$, if there is a-node containing the corresponding i-nodes in the original graph.

We simplify our analysis as follows. Instead of counting on $\mathcal{G}$, we perform our counting on the *computation tree*, denoted by $\mathcal{C}_G$ rooted at the node i. It is not hard to see that by counting $\mathring{X}$ on $\mathcal{C}_G$ we will over-estimate the same on $\mathcal{G}$. We further simplify by considering the degree of the root node in $\mathcal{C}_G$ to be K and the degree of every other node to be $K+1$. Denote this last graph by $\mathcal{T}_G$.

With the above simplifications, let us now count the number of sets $\mathring{X}$ on the graph $\mathcal{T}_G$. We first provide a crude upper bound which is sufficient for the purpose.

We will now provide a *finger-print* for any set $\mathring{X}$. Given a set $\mathring{X}$, generate the following sequence of numbers. Start from the root node and number every edge coming out of the root node as follows. Mark an edge e coming out of the root node by 0, if the node (other than the root node) is not contained in the set $\mathring{X}$. If the node is contained in $\mathring{X}$ but not further extended (by extended we mean we include any of the node below the current node in the set $\mathring{X}$) then mark the edge by 1. If the node is further extended below, then mark the edge by 2.

Consider the following generation of the finger-print of the set $\mathring{X}$. Perform Breadth-First-Search (BFS) type labelling of edges according to previous rules. First consider the K-tuple ($\in \{0,1,2\}$) corresponding to the edges emanating

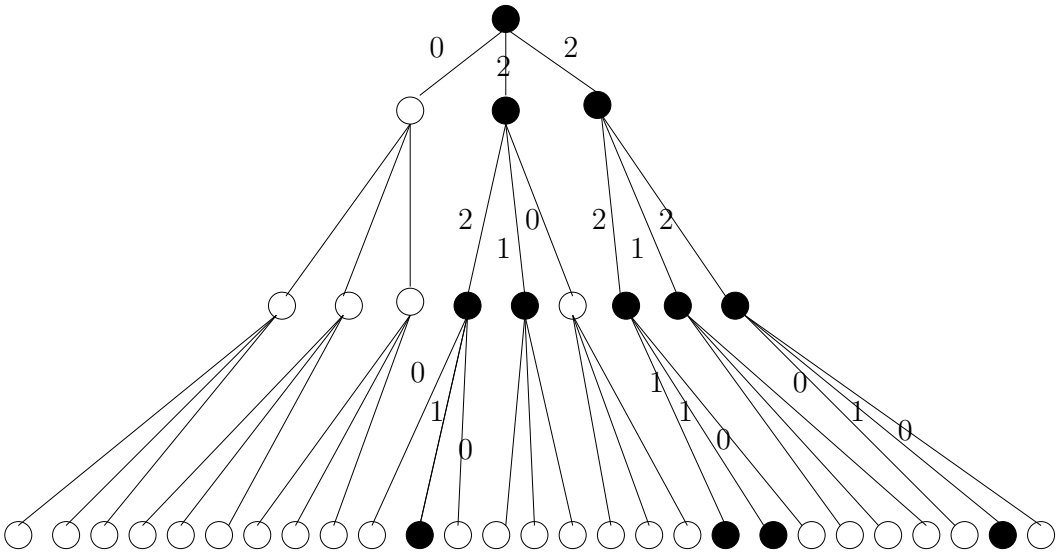

Figure 5.2: The finger-print corresponding to the set $\mathring{X}$ shown by dark circles is given by $F = \{(0,2,2)(2,1,0)(2,1,2)(0,1,0)(1,1,0)(0,1,0)\}$. It is easy to see that given the finger-print F we can reconstruct the set $\mathring{X}$ uniquely in BFS manner.

from the root node. Observe that the set $\mathring{X}$ grows only along edges labeled as 2. Thus write down the K-tuple corresponding to the end-node of all edges marked as 2. Proceed to build up the sequence of numbers in a BFS manner. At the end we have a sequence of K-tuples each containing either 0, 1, or 2. It is easy to see that there is one-to-one correspondence between the finger-print and the set $\mathring{X}$. Indeed, from a given finger-print sequence one can uniquely generate a set $\mathring{X}$ in a BFS manner. Note that we have written down a K-tuple only when we have added a node to the set $\mathring{X}$. Clearly the number of K-tuples in the finger-print is upper bounded by L. As a result, the number of possible sets, $\mathring{X}$, is upper bounded by $(3^K)^{|\mathring{X}|}$.

Putting everything together we get that the number of sets $\hat{X}$, with a given size L is upper bounded by

$$\sum_{|\mathring{X}|=\frac{L}{\mathtt{l}_{\max}}}^{\mathtt{r}_{\max}L} (3^{K|\mathring{X}|}) \leq L\mathtt{r}_{\max}(3^{K\mathtt{r}_{\max}L}),$$

proving the lemma. Figure 5.2 illustrates a BFS labeling. □

5.D Cluster Expansion for LDPC Codes

Here we adapt the Berretti cluster expansion to our setting of Tanner graphs. For more details we refer to [116], [118]. Consider the *replicated* partition

function

$$Z_\perp^2 = \sum_{\underline{u}^{(1)},\underline{u}^{(2)} \in \{-1,+1\}^m} \prod_{k=1}^{n} (1 + \tau_k^{(1)} e^{-2l_k})(1 + \tau_k^{(2)} e^{-2l_k}). \tag{5.63}$$

Recall that $Z_\perp$ corresponds to the dual code. To avoid confusion we will always have in mind the Tanner graph corresponding to the original LDPC code. Furthermore, we will refer to the variable nodes of the Tanner graph (of the LDPC code) as i-nodes and refer to the check nodes as a-nodes.

Above $\underline{u}^{(1)} = u_1^{(1)}, \ldots, u_m^{(1)}$ and $\underline{u}^{(2)} = u_1^{(2)}, \ldots, u_m^{(2)}$ are two replicas of the a-nodes. Also we have $\tau_k^{(1)} = \prod_{a \in k} u_a^{(1)}$ and $\tau_k^{(2)} = \prod_{a \in k} u_a^{(2)}$. Arguing on the same lines as in Appendix 4.A of the previous chapter, we can write the correlation function as

$$\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp = \frac{1}{2} \langle (\tau_i^{(1)} - \tau_i^{(2)})(\tau_j^{(1)} - \tau_j^{(2)}) \rangle_{\perp,12}, \tag{5.64}$$

where $\langle - \rangle_{\perp,12}$ corresponds to the replicated system. In the above language, i and j are i-nodes. Note that in obtaining a similar expression in Appendix 4.A, we only relied on the linearity of the Gibbs bracket and the fact the replicated system is a product of two copies. To lighten the notation we use $f_i = \tau_i^{(1)} - \tau_i^{(2)}$, $f_j = \tau_j^{(1)} - \tau_j^{(2)}$. Then we have

$$\langle f_i f_j \rangle_{\perp,12} = \frac{1}{Z_\perp^2} \sum_{\underline{u}^{(1)},\underline{u}^{(2)}} f_i f_j \prod_{k=1}^{n} (1 + E_k). \tag{5.65}$$

Recall that E_k (c.f. (5.22)) is given by

$$E_k = \tau_k^{(1)} e^{-2l_k} + \tau_k^{(2)} e^{-2l_k} + \tau_k^{(1)} \tau_k^{(2)} e^{-4l_k}. \tag{5.66}$$

Expanding the product $\prod_{k=1}^{n} (1 + E_k)$ we get,

$$\begin{aligned} \langle f_i f_j \rangle_{\perp,12} &= \frac{1}{Z_\perp^2} \sum_{\underline{u}^{(1)},\underline{u}^{(2)}} f_i f_j \sum_{V \subset \mathfrak{V}} \prod_{k \in V} E_k \\ &= \frac{1}{Z_\perp^2} \sum_{V \subset \mathfrak{V}} \sum_{\underline{u}^{(1)},\underline{u}^{(2)}} f_i f_j \prod_{k \in V} E_k, \end{aligned} \tag{5.67}$$

where $\mathfrak{V}$ denotes the set of all variable nodes of the original Tanner graph for the LDPC code and V is any subset of distinct variable nodes.

We will now reason that those V (in the sum in (5.67)) which do not "connect" i and j do not contribute to the above sum. Let us argue more formally. Suppose $V \subset \mathfrak{V}$ is such that *one cannot create a walk* (i.e., a sequence of alternating i-nodes and a-nodes) connecting any a-node in ∂i, to any a-node in ∂j such that all its i-nodes are contained entirely in V. Then we can partition V into three mutually disjoint sets of variable nodes, V_1, V_2, V_3 such

that $i \in V_1$, $j \in V_2$ and $V_3 = V \setminus (V_1 \cup V_2)$. Note also that $\partial V_1, \partial V_2, \partial V_3$ are also mutually disjoint otherwise we can create a walk between ∂i and ∂j. Thus we can write

$$\sum_{\underline{u}^{(1)},\underline{u}^{(2)}} f_i f_j \prod_{k\in V} E_k = \Bigg(\sum_{\substack{u_a^{(1)},u_a^{(2)} \\ \text{all } a\notin\partial V}} 1 \Bigg) \Bigg(\sum_{\substack{u_a^{(1)},u_a^{(2)} \\ \text{all } a\in\partial V_1}} f_i \prod_{k\in V_1} E_k \Bigg) \Bigg(\sum_{\substack{u_a^{(1)},u_a^{(2)} \\ \text{all } a\in\partial V_2}} f_j \prod_{k\in V_2} E_k \Bigg) \times \Bigg(\sum_{\substack{\underline{u}^{(1)},\underline{u}^{(2)} \\ u_a^{(1)},u_a^{(2)}\in\partial V_3}} \prod_{k\in V_3} E_k \Bigg). \tag{5.68}$$

We point out that the pre-factor in the r.h.s. above comes from the sum over all a-nodes which do not belong to ∂V. We observe that under the exchange $(1) \leftrightarrow (2)$, f_i (or f_j) is "antisymmetric" and E_k is "symmetric". More precisely, we have

$$\sum_{\substack{u_a^{(1)},u_a^{(2)} \\ \text{all } a\in\partial V_1}} (\tau_i^{(1)} - \tau_i^{(2)}) \prod_{k\in V_1} \left(\tau_k^{(1)} e^{-2l_k} + \tau_k^{(2)} e^{-2l_k} + \tau_k^{(1)}\tau_k^{(2)} e^{-4l_k} \right) = \sum_{\substack{u_a^{(1)},u_a^{(2)} \\ \text{all } a\in\partial V_1}} (\tau_i^{(2)} - \tau_i^{(1)}) \prod_{k\in V_1} \left(\tau_k^{(2)} e^{-2l_k} + \tau_k^{(1)} e^{-2l_k} + \tau_k^{(2)}\tau_k^{(1)} e^{-4l_k} \right).$$

where we exchanged the replica indices (1) and (2). Thus we obtain that (5.68) vanished. Thus only those V which contain a walk with all its i-nodes in V and which intersects both ∂i and ∂j contributes to the sum in (5.67).

For a given V (contributing to the sum (5.67)), we construct the set of i-nodes Γ_V as follows. Γ_V is the union of *all maximal connected clusters of distinct i-nodes* in V, such that each of those connected clusters intersects the union $\partial i \cup \partial j$. Let $\Gamma_V^c = V \setminus \Gamma_V$. From the definition of Γ_V, it is clear that no i-node $\in \Gamma_V^c$ shares an a-node in common with any node $\in \Gamma_V$. It is not hard to see that there exists such a set, Γ_V. Indeed, the walk which connects ∂i and ∂j, is a candidate set. It is clear that the set Γ_V is unique.

Let $\hat{X}_V = \partial\Gamma_V \cup \partial i \cup \partial j$ be a set of a-nodes. Since Γ_V is unique, it implies that $\hat{X}_V$ is also unique. It is not difficult to see that $\hat{X}_V$ satisfies all the requirements of the set $\hat{X}$ in the sum (5.20). Indeed, consider $X_V = \Gamma_V \cup i \cup j$. By construction, $\partial X_V = \hat{X}_V$; any two i-nodes in X_V are connected by a walk with all its i-nodes in X_V; $\hat{X}_V$ contains both ∂i and ∂j. Also note that Γ_V is compatible with $\hat{X}_V$ as is required in the sum (5.21). Indeed, by construction $\partial\Gamma_V \cup \partial i \cup \partial j = \hat{X}_V$; $\partial\Gamma_V \cap \partial i \neq \phi$ and $\partial\Gamma_V \cap \partial j \neq \phi$; there exists a walk between ∂i and ∂j with all its i-nodes in Γ_V. An important point to keep in mind is that $\partial\Gamma_V^c$ does not intersect $\hat{X}_V$. Because if it did, then it must intersect either $\partial i \cup \partial j$ or $\partial\Gamma_V$, which would be impossible.

With this we can write (5.67) as

$$\begin{aligned}\langle f_i f_j\rangle_{\perp,12} &= \frac{1}{Z_\perp^2}\sum_{V\subset\mathfrak{V}}\left\{\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ a\in\partial\Gamma_V\cup\partial i\cup\partial j=\hat{X}_V}} f_i f_j\prod_{k\in\Gamma_V}E_k\right\}\left\{\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ \text{remaining } a}}\prod_{k\in\Gamma_V^c}E_k\right\}\\ &= \frac{1}{Z_\perp^2}\sum_{\hat{X}}\sum_{\substack{V\subset\mathfrak{V}:\\ \hat{X}_V=\hat{X}}}\left\{\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ a\in\hat{X}}} f_i f_j\prod_{k\in\Gamma_V}E_k\right\}\left\{\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ \text{remaining } a}}\prod_{k\in\Gamma_V^c}E_k\right\}.\end{aligned} \tag{5.69}$$

In the second equality we first sum over all $\hat{X}$ which satisfies all the three conditions (a), (b) and (c) of Section 5.5. Then we sum over all the sets of i-nodes, V, such that $\hat{X}_V = \hat{X}$. We are able to perform the above exchange of sums because, for each V there is a unique $\hat{X}_V$. Note that there can be many V which have the same $\hat{X}_V$ as seen from the Figure 5.1 in Section 5.5. Also, in the second sum in (5.69), by "remaining a" we mean sum over all a-nodes which are not contained in V.

Recall that $\partial\Gamma_V^c$ does not intersect $\hat{X}$. And Γ_V is compatible with $\hat{X}$ and is unique for a given V. As a result we re-sum over the sets V. This re-summation consist of Γ compatible with $\hat{X}$ and the rest $\mathcal{G}$ which does not intersect $\hat{X}$. Thus we have

$$\begin{aligned}\langle f_i f_j\rangle_{\perp,12} &= \frac{1}{Z_\perp^2}\sum_{\hat{X}}\left\{\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ a\in\hat{X}}}\sum_{\substack{\Gamma\text{compatible}\\ \text{with}\hat{X}}} f_i f_j\prod_{k\in\Gamma}E_k\right\}\left\{\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ a\in\hat{X}^c}}\sum_{\substack{\mathcal{G}\subseteq\mathfrak{V}\\ \partial\mathcal{G}\cap\hat{X}=\phi}}\prod_{k\in\mathcal{G}}E_k\right\}\\ &= \frac{1}{Z_\perp^2}\sum_{\hat{X}}\left\{\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ a\in\hat{X}}}\sum_{\substack{\Gamma\text{compatible}\\ \text{with}\hat{X}}} f_i f_j\prod_{k\in\Gamma}E_k\right\}\left\{\sum_{\substack{u_a^{(1)},u_a^{(2)}\\ a\in\hat{X}^c}}\prod_{\substack{\text{all } k \text{ s.t.}\\ \partial k\cap\hat{X}=\phi}}(1+E_k)\right\}.\end{aligned} \tag{5.70}$$

The last bracket is equal to (5.23) and we recognize Berretti's expansion. Figure 5.3 shows a sample set V and Γ_V which give a non-vanishing contribution.

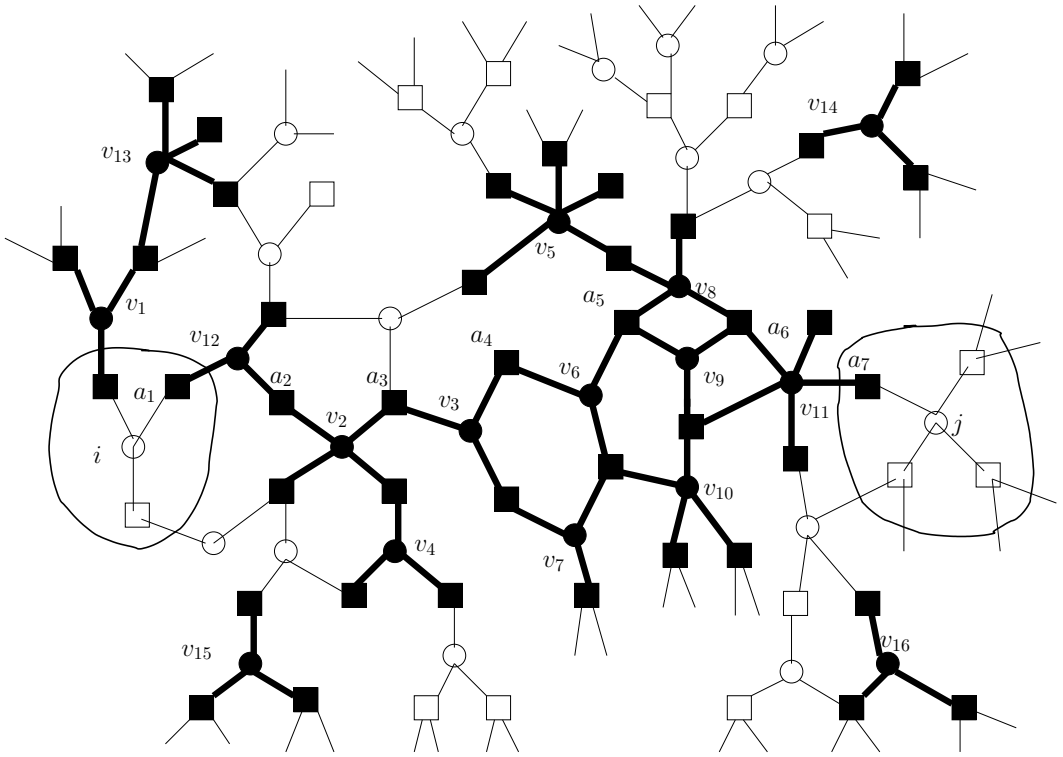

Figure 5.3: The *dark* circles (i-nodes) form the set $V = \{v_1, \ldots, v_{16}\}$. The walk $a_1 v_{12} a_2 v_2 a_3 v_3 a_4 v_6 a_5 v_9 a_6 v_{11} a_7$ connects ∂i to ∂j and hence V has a non-vanishing contribution to the sum (5.67). The set $\Gamma_V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}\}$, is union of the two maximal connected clusters: $\{v_1, v_{13}\}$ and $\{v_{12}, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}\}$. Both of these clusters have an intersection with $\partial i \cup \partial j$. Also, $\Gamma_V^c = \{v_{14}, v_{15}, v_{16}\}$.

Bibliography

[1] S.-Y. Chung, G. D. Forney, Jr., T. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, no. 2, pp. 58–60, Feb. 2001.

[2] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[3] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory*, vol. 8, pp. 21–28, Jan. 1962.

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding," in *Proc. of ICC*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[5] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 32, no. 18, pp. 1645–1646, Aug. 1996.

[6] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Info. Theory*, vol. 45, no. 2, pp. 399–431, 1999. [Online]. Available: http://www.inference.phy.cam.ac.uk/mackay/abstracts/mncN.html

[7] D. A. Spielman, "Linear-time encodeable and decodable error-correcting codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1723–1731, Nov. 1996.

[8] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 585–598, Feb. 2001.

[9] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[10] N. Wiberg, "Codes and decoding on general graphs," Ph.D. dissertation, Linköping University, S-581 83, Linköping, Sweden, 1996.

[11] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533–547, Sept. 1981.

[12] D. J. Costello and G. D. Forney, Jr., "Channel coding: The road to channel capacity," *Proc. IEEE*, vol. 95, p. 1150, June 2007.

[13] J. H. Kim and J. Pearl, "A computational model for causal and diagnostic reasoning in inference systems," in *IJCAI*, 1983, pp. 190–193.

[14] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA, USA: Morgan Kaufmann Publ., 1988.

[15] H. A. Bethe, "Statistical theory of superlattices," *Proc. Roy. Soc.*, vol. A150, pp. 552–75, 1935.

[16] D. Burshtein and G. Miller, "Bounds on the maximum-likelihood decoding error probability of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2696–2710, Nov. 2001.

[17] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.

[18] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. A. Spielman, and V. Stemann, "Practical loss-resilient codes," in *Proc. of the 29th annual ACM Symposium on Theory of Computing*, 1997, pp. 150–159.

[19] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.

[20] C. Di, D. Proietti, T. Richardson, E. Telatar, and R. Urbanke, "Finite length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1570–1579, June 2002.

[21] A. Shokrollahi, "Capacity-achieving sequences," in *Codes, Systems, and Graphical Models*, ser. IMA, USA Volumes in Math. and its Applications, B. Marcus and J. Rosenthal, Eds., vol. 123. Springer-Verlag, 2000, pp. 153–166.

[22] H. D. Pfister, I. Sason, and R. Urbanke, "Capacity-achieving ensembles for the binary erasure channel with bounded complexity," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2352–2379, July 2005.

[23] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[24] C. Méasson, A. Montanari, and R. Urbanke, "Maxwell's construction: The hidden bridge between maximum-likelihood and iterative decoding," 2005, submitted to IEEE Transactions on Inform. Theory.

[25] N. Sourlas, "Spin-glass models as error-correcting codes," *Nature*, vol. 339, no. 29, pp. 693–695, June 1989.

[26] I. Kanter and D. Saad, "Error-correcting codes that nearly saturate Shannon's bound," *Phys. Rev. Lett.*, vol. 83, no. 13, pp. 2660–2663, 1999.

[27] T. Murayama, Y. Kabashima, D. Saad, and R. Vicente, "Statistical physics of regular low-density parity-check error-correcting codes," *Phys. Rev. E*, vol. 62, no. 026705, pp. 1577–1591, 2000.

[28] S. Franz, M. Leone, A. Montanari, and F. Ricci-Tersenghi, "The dynamic phase transition for decoding algorithms," *Phys. Rev. E*, vol. 66, no. 046120, 2002, e-print: cond-mat/0205051.

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

[30] F. Guerra, "Sum rules for the free energy in the mean field spin-glass model," *Fields Institute Communications*, vol. 30, p. 161, 2001.

[31] F. Guerra and F. L. Toninelli, "Quadratic replica coupling in the Sherrington-Kirkpatrick mean field spin glass model," *J. Math.Phys.*, vol. 43, pp. 3704–3716, 2002.

[32] M. Talagrand, "On the high temperature region of the sherrigton-kirkpatrick model," *Annals of Probability*, vol. 30, pp. 364–381, 2002.

[33] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[34] C. Measson, "Conservation laws for coding," Ph.D. dissertation, EPFL, Lausanne, VD, Switzerland, Mar. 2006.

[35] R. McEliece and S. M. Aji, "The generalized distributive law," in *Int. Symp. Comm. Th. Appl.*, Ambleside, UK, July 1997.

[36] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, "The generalized area theorem and some of its consequences," 2005, submitted to IEEE Transactions on Inform. Theory.

[37] S. Korada and R. Urbanke, "Exchange of limits: Why iterative decoding works," 2008, submitted to IEEE Transactions on Inform. Theory.

[38] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel property," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[39] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," tMO Progress Report 42-139.

[40] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[41] D. Burshtein, M. Krivelevich, S. L. Litsyn, and G. Miller, "Upper bounds on the rate of LDPC codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2437–2449, Sept. 2002.

[42] A. Shokrollahi, "New sequences of linear time erasure codes approaching the channel capacity," in *Proc. of AAECC-13, Lect. Notes Comput. Sci. 1719*, ser. Lect. Notes Comput. Sci., no. 1719. Springer Verlag, 1999, pp. 65–76.

[43] N. C. Wormald, "Differential equations for random processes and random graphs," *Ann. Appl. Probab.*, vol. 5, pp. 1217–1235, 1995.

[44] C. Méasson, A. Montanari, and R. Urbanke, "Why we can not surpass capacity: the matching condition," in *Proc. of the Allerton Conf. on Commun., Control and Computing*, Monticello, IL, USA, Oct. 2005.

[45] T. Richardson and R. Urbanke, "Efficient encoding of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 638–656, Feb. 2001.

[46] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2003.

[47] C. Méasson and R. Urbanke, "An upper-bound on the ML thresholds of LDPC ensembles over the BEC," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2003, pp. 478–487.

[48] G. Miller and G. Cohen, "The rate of regular LDPC codes," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2989–2992, Mar. 2003.

[49] S. Sanghavi, D. Shah, and A. Willsky, "Message-passing for maximum weight independent set," *IEEE Transactions on Information Theory*, submitted.

[50] M. Bayati, D. Shah, and M. Sharma, "Max-product for maximum weight matching: Convergence, Correctness, and LP duality," *IEEE Transactions on Information Theory*, vol. 54, pp. 1241–1251, Mar. 2008.

[51] M. Chertkov, "Exactness of belief propagation for some graphical models with loops," *J. Stat. Mech.*, 2008.

[52] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.

[53] J. S. Yedidia, Y. Weiss, and W. T. Freeman, "Understanding belief propagation and its generalizations," *Exploring Artificial Intelligence in the New Millennium*, no. ISBN 1558608117, pp. 239–236, Jan. 2003, chap. 8.

[54] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, Dec. 2008.

[55] M. Chertkov and V. Chernyak, "Loop series for discrete statistical models on graphs," *JSTAT*, 2006.

[56] ——, "Loop calculus in statistical physics and information science," *Phys. Rev. E*, vol. 73, 2006.

[57] J. Salez and D. Shah, "Belief propagation: an asymptotically optimal algorithm for the random assignment problem," *Mathematics of Operations Research*, submitted.

[58] A. Montanari and D. Shah, "Counting good truth assignment in random k-sat," *Proceedings of ACM-SIAM SODA*, p. 9, 2007.

[59] A. Bandyopadhyay and D. Gamarnik, "Counting without sampling. new algorithms for enumeration problems using statistical physics," *Random Structures and Algorithms*, 2008.

[60] D. Weitz, "Counting independent sets up to the tree threshold," in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, Seattle, USA, 2006, pp. 140–149.

[61] W. T. Freeman and Y. Weiss, "Correctness of belief propagation in gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, pp. 2173–2200, 2001.

[62] S. Tatikonda and M. I. Jordan, "Loopy belief propagation and gibbs measures," in *18th Conference on Uncertainty in Artificial Intelligence*, Edmonton, Canada, Aug. 2002.

[63] N. Sourlas, "Spin glasses, error-correcting codes and finite-temperature decoding," *Europhys. Lett.*, vol. 25, pp. 159–164, 1994.

[64] A. Montanari and N. Sourlas, "The statistical mechanics and turbo codes," in *Proc. of the Int. Conf. on Turbo Codes and Related Topics*, Brest, France, Sept. 2000, pp. 63–66.

[65] A. Montanari, "The glassy phase of Gallager codes," *Eur. Phys. J. B*, vol. 23, pp. 121–136, 2001, e-print: cond-mat/0104079.

[66] Y. Kabashima and D. Saad, "Statistical mechanics of low-density parity check codes," *J. Phys. A*, vol. 37, pp. R1–R43, 2004, invited paper.

[67] R. Vicente, D. Saad, and Y. Kabashima, *Low Density Parity Check Codes–A Statistical Physics Perspective*, ser. Advances in Imaging and Electron Phys. Elsevier Science, 2002, vol. 125, in press.

[68] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction.* Oxford Science Publ., 2001.

[69] Y. Kabashima, N. Sazuka, K. Nakamura, and D. Saad, "Evaluating zero error noise thresholds by the replica method for Gallager code ensembles," in *Proc. of the IEEE Int. Symposium on Inform. Theory.* Lausanne, Switzerland: IEEE, June 2002, p. 255.

[70] J. van Mourik, D. Saad, and Y. Kabashima, "Magnetization enumerator for LDPC codes – A statistical physics approach," in *Proc. of the IEEE Int. Symposium on Inform. Theory.* Lausanne, Switzerland: IEEE, June 2002, p. 256.

[71] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass Theory and Beyond.* World Scientific Publ., 1987.

[72] M. Mézard and G. Parisi, "The cavity method at zero temperature," *J. Stat. Phys.*, vol. 111, pp. 1–34, 2003.

[73] B. Derrida, "Random-energy model: An exactly solvable model of disordered systems," *Phys. Rev. B*, vol. 24, pp. 2613–2626, 1981.

[74] M. Mézard and A. Montanari, *Information, Physics, and Computation.* Clarendon Press, Oxford, 2007.

[75] M. Talagrand, "The parisi formula," *Annals of Mathematics*, vol. 163, pp. 221–263, 2006.

[76] A. Braunstein, M. Mézard, and R. Zecchina, "Survey propagation: algorithm for satisfiability," e-print: cs.CC//0212002.

[77] M. Mézard and R. Zecchina, "Random k-satisfiability problem: From an analytic solution to an efficient algorithm," *Phys. Rev. E*, vol. 66, 2002.

[78] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina, "Alternative solutions to diluted p-spin models and XORSAT problems," *J. Stat. Phys.*, vol. 111, p. 505, 2003, e-print: cond-mat/0207140.

[79] S. Franz, M. Leone, and F. L. Toninelli, "Replica bounds for optimization and diluted spin glasses," *J. Stat. Phys.*, vol. 111, pp. 535–564, 2003.

[80] S. Mertens, "Phase transition in the number partitioning problem," *Phys. Rev. Lett.*, vol. 81, pp. 4281–4284, 1998.

[81] M. Mézard and G. Parisi, "Mean field equations for the matching and the travelling salesman problems," *Europhys. Lett.*, vol. 2, p. 913, 1986.

[82] E. Maneva, E. Mossel, and M. J. Wainwright, "A new look at survey propagation and its generalizations," in *SODA*, Vancouver, Canada, 2005, conference.

[83] R. Monasson and R. Zecchina, "Statistical mechanics of the random k-satisfiability model," *Phys. Rev. E*, vol. 56, pp. 1357–1370, 1997.

[84] N. Macris, "Griffiths-Kelly-Sherman correlation inequalities: a useful tool in the theory of LDPC codes," *IEEE Trans. Inform. Theory*, vol. IT–53, no. 2, pp. 664–683, 2007.

[85] M. Talagrand, "The high temperature case of the k-sat problem," *Probability Theory and Related Fields*, vol. 119, pp. 187–212, 2001.

[86] F. Guerra, "Replica broken bounds in the mean field spin glass model," *Comm. Math. Phys.*, vol. 233, pp. 1–12, 2003.

[87] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models*. New York, NY, USA: Springer, 2003.

[88] D. Pachenko and M. Talagrand, "Bounds for diluted mean-fields spin glass models," *Prob. Theory. Relat. Fields*, vol. 130, pp. 319–336, 2004.

[89] A. Montanari, "Estimating random variables from random sparse observations," *Eur. Trans. Telecom.*, 2007.

[90] M. S. Pinsker, "Information and information stability of random variables and processes," 1964.

[91] N. Macris, "Sharp bounds on generalized exit functions," *IEEE Trans. Inform. Theory*, vol. IT–53, no. 2, pp. 2365–2375, 2007.

[92] S. Franz, M. Leone, and F. L. Toninelli, "Replica bounds for diluted non-Poissonian spin systems," *J. Phys. A*, vol. 36, pp. 10 967–10 985, 2003.

[93] J. Lebowitz, "Ghs and other inequalities," *Comm. Math. Phys.*, vol. 35, no. 2, pp. 87–92, 1974.

[94] R. B. Griffiths, *Phase Transitions and Critical Phenomena*. New York: Academic, 1972, vol. 1.

[95] S. Morita, H. Nishimori, and P. Contucci, "Griffiths inequalities for gaussian spin glasses," *J. Phys. A*, vol. 37, p. 203, 2004.

[96] S. Korada and N. Macris, "Exact solution of a p-spin model and its relationship to error correcting codes," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Seattle, USA, July 2006, submitted to Journal of Statistical Physics (2009).

[97] S. Korada, S. Kudekar, and N. Macris, "Concentration of magnetization for linear block codes," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Toronto, Canada, Sept. 2008.

[98] A. Bovier, *Statistical Mechanics of Disordered Systems: A Mathematical Perspective*. Cambridge University Press, 2006.

[99] H. O. Georgii, "de gruyter studies in mathematics," in *Gibbs Measures and Phase Transitions*. Walter de Gruyter & Co., Berlin, 1988.

[100] D. Ruelle, *Statistical Mechanics: Rigorous Results*. New York, NY, USA: Imperial College Press, 2003.

[101] H. Dreifus, A. Klein, and J. Perez, "Taming griffiths' singularities: Infinite differentiability of quenched correlation functions," *Communications in Mathematical Physics*, vol. 170, pp. 21–39.

[102] A. Klein, "Who is afraid of griffith's singularities?" *On Three Levels. Micro, Meso and Macroscopic Approaches in Physics*, pp. 253–258, 1994.

[103] G. Gielis and C. Maes, "The uniqueness regime of gibbs fields with unbounded disorder," *JSTAT*, vol. 81, 1995.

[104] R. L. Dobrushin, "Description of a random field by means of its conditional probabilities and conditions of its regularity," *Theory Prob. Appl.*, vol. 13, pp. 197–224, 1968.

[105] D. Brydges, *A Short Course on Cluster Expansions*. Les Houches session XLIII, 1984.

[106] C. Méasson, A. Montanari, and R. Urbanke, "Maxwell's construction: The hidden bridge between maximum-likelihood and iterative decoding," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Chicago, IL, USA, 2004, p. 225.

[107] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, "Life above threshold: From list decoding to area theorem and MSE," in *Proc. of the IEEE Inform. Theory Workshop*, San Antonio, TX, USA, Oct. 2004, e-print: cs.IT/0410028.

[108] D. Weitz, "Combinatorial criteria for uniqueness of gibbs measures," *Random Structures and Algorithms*, vol. 27, pp. 445–475, 2005.

[109] L. A. Bassalygo and R. L. Dobrushin, "Uniqueness of a gibbs field with random potential – an elementary approach," *Theory Prob. Appl.*, vol. 31, pp. 572–589, 1986.

[110] P. Erdos and J. Spencer, *The Probabilistic Method*. Wiley Interscience Series in Discrete Mathematics and Optimization, 1992.

[111] B. Bollobás, *Random Graphs*. Cambridge Univ. Press, 2001.

[112] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge, MA, USA: Cambridge Univ. Press, 1995.

[113] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. A. Spielman, "Analysis of low density codes and improved designs using irregular graphs," in *Proc. of the 30th Annual ACM Symposium on Theory of Computing*, 1998, pp. 249–258.

[114] T. Richardson, A. Shokrollahi, and R. Urbanke, "Finite-length analysis of various low-density parity-check ensembles for the binary erasure channel," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Lausanne, Switzerland, June 2002, p. 1.

[115] C. Measson, A. Montanari, and R. Urbanke, "Asymptotic rate versus design rate," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Nice, France, July 2007, pp. 1541–1545.

[116] A. Berretti, "Some properties of random ising models," *Journal of Statistical Physics*, vol. 38, pp. 483–496, 1985.

[117] G. D. Forney, Jr., "Codes on graphs: Generalized state realizations," submitted to IEEE Trans. Inform. Theory.

[118] J. Frohlich, "Mathematical aspects of disordered systems," *Les Houches session XLIII*, p. 1984.

Curriculum Vitae – Shrinivas Kudekar

Education:

- **Swiss Federal Institute of Technology at Lausanne (EPFL)**
  Ph.D. in Communication Systems
  Thesis Title: Statistical Physics Methods for Sparse Graph Codes
  September 2004 - June 2009.

- **Indian Institute of Technology Bombay (IIT Bombay)**
  Bachelor of Technology in Electrical Engineering
  July 1999 - May 2003.

Journal Publications:

- [J1] *Sharp Bounds for MAP Decoding of General Irregular Low-Density Parity Check Codes.* Shrinivas Kudekar, Nicolas Macris. Accepted to IEEE Transactions on Information Theory.

- [J2] *Decay of Correlations for Sparse Graph Error Correcting Codes.* Shrinivas Kudekar, Nicolas Macris. Submitted to SIAM Journal on Discrete Mathematics.

Conference Publications:

- [C1] *Sharp Bounds for MAP Decoding of General Irregular Low-Density Parity Check Codes.* Shrinivas Kudekar, Nicolas Macris. In the Proceedings of IEEE International Symposium on Information Theory (ISIT), Seattle, 2006.

- [C2] *Exact solution for the Conditional Entropy of Poissonian Low-Density Parity Check Codes over the Binary Erasure Channel.* Satish Babu Korada, Shrinivas Kudekar, Nicolas Macris. In the Proceedings of IEEE International Symposium on Information Theory (ISIT), Nice, 2007.

- [C3] *Concentration of Magnetization for Linear Block Codes.* Satish Babu Korada, Shrinivas Kudekar, Nicolas Macris. In the proceedings of IEEE International Symposium on Information Theory (ISIT), Toronto, 2008.

- [C4] *Decay of Correlations: An Application to Low-Density Parity Check Codes.* Shrinivas Kudekar and Nicolas Macris. In the proceedings of 5th International Symposium on Turbo Codes and Related Topics.

- [C5] *Lower Bounds on the Rate-Distortion Function of Individual Low-Density Generator Matrix Codes* . Shrinivas Kudekar and Rüdiger Urbanke. In 5th International Symposium on Turbo Codes and Related Topics.

- [C6] *Proof of replica formulas in the high noise regime for communication using Low-Density Generator Matrix Codes.* Shrinivas Kudekar and Nicolas Macris. In the proceedings of Information Theory Workshop (ITW), Porto, 2007.

- [C7] *Decay of Correlations in Low-Density Parity Check Codes: Low Noise Regime.* Shrinivas Kudekar, Nicolas Macris. Submitted to IEEE International Symposium on Information Theory (ISIT), Seoul, 2009.

- [C8] *Learning TCP–Adaptive Congestion Detection for Heterogeneous Networks.* Ajay Kumar Singh, Abhishek Jain, Abhay Karandikar, Shrinivas Kudekar and Sachin Katti. In National Conference on Communications 2004, India.