High-Dimensional Inference on Dense Graphs with Applications to Coding Theory and Machine Learning

Mohamad Dia

EPFL - Ecole Polytechnique Fédérale de Lausanne

Thesis No. 8954 (September 2018)

Thesis presented to the faculty of computer and communication sciences for obtaining the degree of Docteur ès Sciences

Accepted by the jury:

Nicolas Macris
Thesis director

Olivier Lévêque
Expert

Marc Lelarge
Expert

Ramji Venkataramanan
Expert

Patrick Thiran
President of the jury

Ecole Polytechnique Fédérale de Lausanne, 2018

Abstract

We are living in the era of "Big Data", an era characterized by a voluminous amount of available data. Such amount is mainly due to the continuing advances in the computational capabilities for capturing, storing, transmitting and processing data. However, it is not always the volume of data that matters, but rather the "relevant" information that resides in it.

Exactly 70 years ago, Claude Shannon, the father of information theory, was able to quantify the amount of information in a communication scenario based on a probabilistic model of the data. It turns out that Shannon's theory can be adapted to various probability-based information processing fields, ranging from coding theory to machine learning. The computation of some information theoretic quantities, such as the *mutual information*, can help in setting fundamental limits and devising more efficient algorithms for many inference problems.

This thesis deals with two different, yet intimately related, inference problems in the fields of coding theory and machine learning. We use Bayesian probabilistic formulations for both problems, and we analyse them in the asymptotic high-dimensional regime. The goal of our analysis is to assess the algorithmic performance on the first hand and to predict the Bayes-optimal performance on the second hand, using an information theoretic approach. To this end, we employ powerful analytical tools from statistical physics.

The first problem is a recent forward-error-correction code called *sparse superposition code*. We consider the extension of such code to a large class of noisy channels by exploiting the similarity with the compressed sensing paradigm. Moreover, we show the amenability of sparse superposition codes to perform joint *distribution matching* and channel coding.

In the second problem, we study *symmetric rank-one matrix factorization*, a prominent model in machine learning and statistics with many applications ranging from community detection to sparse principal component analysis. We provide an explicit expression for the normalized mutual information and the minimum mean-square error of this model in the asymptotic limit. This allows us to prove the optimality of a certain iterative algorithm on a large set of parameters.

A common feature of the two problems stems from the fact that both

of them are represented on *dense graphical models*. Hence, similar *message-passing* algorithms and analysis tools can be adopted. Furthermore, *spatial coupling*, a new technique introduced in the context of low-density parity-check (LDPC) codes, can be applied to both problems. Spatial coupling is used in this thesis as a "construction technique" to boost the algorithmic performance and as a "proof technique" to compute some information theoretic quantities.

Moreover, both of our problems retain close connections with *spin glass* models studied in statistical mechanics of disordered systems. This allows us to use sophisticated techniques developed in statistical physics. In this thesis, we use the *potential function* predicted by the *replica method* in order to prove the *threshold saturation* phenomenon associated with spatially coupled models. Moreover, one of the main contributions of this thesis is proving that the predictions given by the "heuristic" replica method are exact. Hence, our results could be of great interest for the statistical physics community as well, as they help to set a rigorous mathematical foundation of the replica predictions for a wide range of Bayesian inference problems.

Keywords: Bayesian inference, channel coding, sparse superposition codes, machine learning, matrix factorization, dense graphical models, message-passing, spatial coupling, compressed sensing, distribution matching, statistical physics, replica method.

Résumé

Nous vivons à l'ère de "Big Data", une ère caractérisée par une énorme quantité de données disponibles. Cette quantité est principalement due aux progrès continus des capacités de calcul permettant de capturer, stocker, transmettre et traiter les données. Cependant, ce n'est pas toujours le volume de données qui compte, mais plutôt l'information "utile" qui y est contenue.

Il y a exactement 70 ans, Claude Shannon, le père de la théorie de l'information, a réussi de quantifier la quantité d'information dans un contexte de communication basé sur un modèle probabiliste des données. Il apparaît que la théorie de Shannon peut être adaptée à divers domaines de traitement probabiliste de l'information, allant de la théorie du codage au "machine learning". Le calcul de certaines quantités de la théorie de l'information, telles que l'information mutuelle, peut aider à déterminer des limites fondamentales et à développer des algorithmes plus efficaces pour plusieurs probèmes d'inférence.

Cette thèse traite deux problèmes d'inférence différents, mais intimement liés, dans les domaines de la théorie du codage et du machine learning. Nous utilisons des formulations probabilistes Bayésiennes pour les deux problèmes et nous les analysons dans le régime asymptotique de grande taille. Le but de notre analyse est d'évaluer la performance algorithmique d'un côté, et de prévoir la performance Bayes-optimale de l'autre côté, en utilisant une approche théorique d'informations. Pour cela, nous employons des outils analytiques puissants issus de la physique statistique.

Le premier problème est un code récent de correction d'erreurs anticipée appelé "*sparse superposition code*". Nous considérons l'extension d'un tel code à une vaste classe de canaux bruyants en exploitant la similitude avec le paradigme de "compressed sensing". De plus, nous montrons que les codes de sparse superposition permettent d'effectuer conjointement le "*distribution matching*" et le codage de canal.

Dans le deuxième problème, nous étudions la *factorisation matricielle symétrique de rang-un*, un modèle important dans machine learning et statistiques avec de nombreuses applications allant de la détection des communautés à l'analyse en composantes principales éparses. Nous présentons une expression explicite pour l'information mutuelle normalisée et l'erreur quadratique moyenne minimale de ce modèle dans la limite asymptotique. Cela nous per-

met de vérifier l'optimalité d'un certain algorithme itératif pour un large choix de paramètres.

Une caractéristique commune aux deux problèmes est due au fait qu'ils sont tous les deux représentés par des *modèles graphiques denses*. Ainsi, on peut adopter des algorithmes de *passage de messages* et des outils d'analyse similaires. De plus, le *couplage spatial*, une nouvelle technique introduite dans le contexte des codes "low-density parity-check" (LDPC), peut être appliqué aux deux problèmes. Le couplage spatial est utilisé dans cette thèse comme un " technique de construction" pour améliorer la performance algorithmique et comme un "technique de preuve" pour déterminer certaines quantités de la théorie de l'information.

En outre, nos deux problèmes maintiennent des liens très proches avec les modèles des *verres de spins* étudiés dans la mécanique statistique des systèmes désordonnés. Cela nous permet d'utiliser des techniques sophistiquées développées en physique statistique. Dans cette thèse, nous utilisons la *fonction potentielle* prédite par la *méthode des répliques* afin de démontrer le phénomène de *saturation du seuil* associé aux modèles spatialement couplés. De plus, l'une des principales contributions de cette thèse est la preuve que les prédictions "heuristiques" obtenues par la méthode des répliques sont exactes. Par conséquent, nos résultats pourraient être d'un grand intérêt aussi pour la communauté de la physique statistique, car ils aident à établir une base mathématique rigoureuse pour les prédictions de la méthode des répliques en vue d'un large spectre de problèmes d'inférence Bayésienne.

Mots clés: Inférence Bayésienne, codage de canal, codes de "sparse superposition", "machine learning", factorisation matricielle, modèles graphiques denses, passage de messages, couplage spatial, "compressed sensing", "distribution matching", physique statistique, méthode des répliques.

Acknowledgements

I would like first to express my sincere gratitude to Nicolas Macris for being my supervisor during the Ph.D. journey. I had the honor and fortune to work with Nicolas for almost four years, during which I have learned much on how to approach research and organize random ideas. I am deeply indebted to him for his continuous support, guidance, persistence, patience, and most importantly for introducing me to the field of statistical physics.

I would like also to thank Rüdiger Urbanke for giving me the opportunity to join his lab at the first place, for the advice he gave, and for the short, though extremely useful, discussions we had. Rüdiger was always a source of excitement, support, and fun.

I am very grateful to Jean Barbier, the postdoc and the "dynamo" of our lab. In fact, the completion of this thesis would not have been possible without his collaboration, hard-work, and valuable inputs. I will always remember the enjoyable moments we spent together inside and outside the office. I would like to thank Jean as well for introducing me to Florent Krzakala, Thibault Lesieur and Lenka Zdeborová, the very passionate researchers in Paris with whom we had collaborated on various problems. I would like to take this opportunity to thank Erdem Bıyık for spending his summer internship with us and for the valuable contribution he made.

I would like to thank Vahid Aref and Laurent Schmalen for hosting me for three months at Nokia Bell Labs. Special thanks go to Vahid for being a collaborator, a friend, and most importantly the father of the heavenly creatures Dorsa and Diana. I am also thankful to Rana Salem for all the advice and support. I truly believe that the interaction I had with the people at Nokia Bell Labs helped me develop as a researcher.

I am grateful to the members of the thesis committee Olivier Lévêque, Marc Lelarge, Patrick Thiran, and Ramji Venkataramanan for the careful reading of the thesis manuscript and the insightful feedback.

During my Ph.D. studies, I was fortunate to be a member of the Information Processing Group (IPG), the big family of four labs at EPFL. I am extremely grateful to all IPG members for the unforgettable memories and for making this place a great work environment. First, I would like to thank the senior members: Bixio, Emre, and Michael. Also, a big thanks to Damir,

France, Françoise, and Muriel for their help and support in different administrative and IT tasks. Special thanks to Muriel for making my stay very smooth and enjoyable, and for insisting on talking with me in French. I am also grateful to all the current and previous members: Adrian, Akshat, Amedeo, Andrei, Clément, Eric, Erixhen, Giel, Hamed, Karol, Kirill, Mani, Marc, Marco, Nicolae, Pinar, Saied, Sepand, Shrinivas, Su, Wei, and Young. In particular, I would like to thank Eric and Young for being my office mates during different periods of my stay. I am deeply grateful to the Lebanese gang at IPG: Ibrahim, Elie, Rafah, Rajai, and Serj for making this work-place a second home.

My stay in Lausanne wouldn't have been that fun without the Lebanese friends living around and making Lausanne *great again*. Thank you so much Abbas B., Abbas H., Abbas S., Abdallah, Akram, Ali A., Ali B., Ali & Maud, Amer, David, Elio, Farah, Grace, Hamza & Walaa, Hani & Carole, Hassan, Hiba, Hussein K., Hussein N., Mahdi, Mahmoud, Marwa, Mohamad, Raed, Rida, Serj & Sahar, and Taha & Mira. Special thanks to Abdallah, Ayoub and "Haboush" for bearing me as an annoying friend. I would like also to extend my gratitude to the non-Lebanese friends in Lausanne, including Brunella, Baran, Farnood, Mohamad, Mohsen, Pedram, and many others.

I am very grateful to all my friends living in Lebanon and abroad, who were a source of support during this journey. Special thanks to Abbas, Ahmad, Ali, Haidar, Hassan, Mahmoud, Mohamad A., Mohamad H., Sadek, Safi, Saleh, and Youssef.

No words or actions can express my gratitude and appreciation towards my extended family and family-in-law. This work would have never been possible without your support, affection, and prayers. I am profoundly grateful to all of you and I shall forever remain indebted. Thank you Mom, Roukaya, for your kindness, continuous encouragement, and unconditional sacrifices. Thank you Fatima and Mahdi for your genuine love and support. These lines are not enough to thank my wife, Ghofran, for her devotion, unwavering support, and endless love.

Finally, I am thankful to God for all successes, and for whatever I am and I have. This thesis is dedicated to my beloved wife, my mother, my siblings, and to the soul of my late father, Youssef.

Lausanne, September 27, 2018 M.D.

Contents

1 Introduction

Seventy years before this thesis was written, Claude Shannon founded the field of information theory in his 1948 celebrated work "A Mathematical Theory of Communication". In his landmark paper [1], Shannon established fundamental limits for both data compression and transmission. Namely, Shannon was able to quantify the maximal amount of information that can be reliably transmitted through a given noisy channel. This quantity was termed *channel capacity*. More importantly, Shannon proved the existence of reliable transmission schemes that achieve the channel capacity in the high-dimensional regime. Although Shannon did not give a practical recipe on how to construct such schemes, his notion of channel capacity paved the way for generations of coding theorists to come up with practical transmission schemes, or *codes*, that approach the channel capacity limit [2].

The notion of channel capacity, or *mutual information*, was established with communication systems in mind. However, this notion is transversal to many probability-based information processing fields. One of these fields is machine learning that we are witnessing its "revolution" nowadays. Indeed, the performance of various algorithms in machine learning and data science is dictated by the amount of relevant information contained in the data. Hence, being able to quantify this information and setting up information theoretic limits can give many insights on designing more efficient machine learning systems in the future, the same way Shannon's capacity inspired coding theorists for the last 70 years.

This thesis addresses two problems in the fields of coding theory and machine learning. The two problems are intimately related through the adoption of similar analysis tools, graphical model representations, iterative algorithms and the *spatial coupling* technique. Moreover, powerful tools developed in statistical mechanics of disordered systems are adapted to both problems in order

to perform the analysis on a rigorous mathematical basis.

In the first part of this thesis, we focus on a channel coding problem by studying a recent forward-error-correction code called *sparse superposition* (SS) code. The decoding task of SS code can be represented via a *dense* graphical model, where iterative *message-passing* algorithms can be applied. We demonstrate how spatial coupling is employed to boost the algorithmic performance of SS code. Furthermore, we show that SS code "universally" achieves Shannon's capacity over a large class of noisy channels in a proper high-dimensional regime. Moreover, we illustrate the amenability of SS code to perform joint source and channel coding.

In the second part, we focus on symmetric rank-one matrix factorization, a fundamental problem in machine learning with many applications ranging from community detection to sparse principal component analysis (PCA) and spiked Wigner model. We provide a probabilistic model for the problem and we tackle it in a Bayesian inference approach. Our central result is an explicit expression for the mutual information in the high-dimensional regime. Consequently, we are able to establish an information theoretic limit that governs the performance of any algorithm on this problem. Moreover, we prove the optimality of message passing algorithm in a large region of system parameters. The spatial coupling technique is exploited here as well but for a dual-purpose. Besides its practical implication in boosting the algorithmic performance, spatial coupling is used as a proof technique to compute some information theoretic quantities of the original "uncoupled" version of the problem.

A common feature in both problems stems from the existence of a large number of degrees of freedom interacting in a random environment. In fact, this is reminiscent of *spin glass* models studied in statistical physics over the last century. The central aim of statistical physics is to describe the macroscopic behavior of such models in the *thermodynamic limit*. This includes the understanding and prediction of some natural phenomena such as nucleation, clustering of solutions and *phase transitions*. One of the most powerful, yet non-rigorous, statistical physics techniques is the *replica method*, or its alternative more probabilistic form known as the *cavity method* [3, 4]. This thesis is not only concerned in using the *potential function* predicted by the "heuristic" replica method in order to perform the analysis, but also in proving the accuracy of this prediction on a rigorous mathematical basis. Hence, our approach could be of great interest for the statistical physics community as well.

In the remaining of this chapter, we give a brief history of coding theory in Section 1.1. We then formulate both of our problems in Section 1.2 and Section 1.3 respectively. We also describe the graphical model representation used for both problems along with variants of the message-passing algorithm and their associated analysis tools in Section 1.4. In Section 1.5, we introduce the spatial coupling technique. The connections to statistical physics is illustrated in Section 1.6. Finally, we outline the main contributions of this thesis in Section 1.7.

1.1 Coding Theory

In his seminal work [1], Shannon formalized the data transmission problem. His fundamental result is the existence of a non-vanishing transmission rate for reliable communication over noisy channels.[1] Furthermore, Shannon was able to quantify the maximal possible rate for reliable communication, over a given probabilistic model of the channel, using the notion of mutual information between two random variables. This maximal rate represents the channel capacity. Moreover, Shannon proved the existence of transmission schemes that achieve channel capacity. However, Shannon's proof was rather probabilistic than constructive. His *random coding* argument ensures achievability of the capacity with codes that are far from being practical.

Shannon's 1948 theory was thenceforth the source of inspiration for generations of coding theorists. The central objective of coding theory is to meet Shannon's challenge using practical codes. This objective can be attained by first constructing "structured" codes that are potentially capacity-achieving.[2] Second, coding theory is concerned in devising low-complexity algorithms that operate very close to capacity on the constructed codes.

The first family of codes that dominated the early days of coding theory is the family of *algebraic codes*. Algebraic coding follows a deterministic approach to construct block codes. This coding paradigm aims to maximize the minimum distance between *codewords*, and hence it is more suitable for worst-case analysis. This family includes Hamming codes [5], Golay codes [6], Reed-Muller codes [7, 8], BCH codes [9, 10], and Reed-Solomon codes [11].

A more probabilistic approach for coding theory started with Elias' convolutional codes [12]. However, the real breakthrough in coding theory was in 1993 after the introduction of turbo codes [13] that initiated a new era of *iterative coding*. The low complexity of iterative decoding in turbo codes led to the "rediscovery" of low-density parity-check (LDPC) codes [14, 15], which were first introduced in Gallager's 1963 thesis [16]. This new coding paradigm, which is based on *sparse* graphical models with iterative message-passing decoding, was then coined *modern coding theory* [17]. The capacity of the LDPC codes under message-passing decoder was determined in [18]. Moreover, it was shown that LDPC codes with an optimized degree distribution can operate very close to Shannon's capacity [19, 20].

An alternative approach to design capacity-achieving codes was introduced in Arikan's 2009 celebrated work [21]. Arikan's codes were then termed *polar codes* due to the channel polarization phenomenon induced by the construction. Polar codes are the first provably capacity-achieving codes under low-complexity decoding over a large class of channels, namely the binary-input memoryless symmetric (BMS) channels.

[1] Being able to communicate with a non-zero rate (bits/sec) while achieving a vanishing probability of error was itself an astonishing result for the communication society.

[2] Codes that achieve capacity under optimal algorithm (e.g. exhaustive search algorithm with exponential time complexity).

Recently, spatially coupled LDPC codes, originally introduced as convolutional LDPC codes [22, 23, 24], have reinforced the modern coding theory. Spatial coupling can boost the performance of the LDPC codes and make these codes capacity-achieving under low-complexity message-passing decoder [25, 26, 27].

1.2 Sparse Superposition Codes

Sparse superposition (SS) codes constitute a recent class of codes that can fall under the umbrella of iterative coding schemes. Starting from a structured sparse *message* with sparsity ratio $1/B$, the codewords of SS codes are constructed by the superposition of L columns from a given Gaussian coding matrix (or dictionary). The name of SS codes was inspired by the superposition principle in the multi-stage decoding [28], yet SS codes use single-stage decoding for point-to-point coding schemes. The decoding task of SS codes can be represented on a graphical model as in the LDPC case. However, the underlying graph for SS codes is dense, as opposed to the sparse[3] graphical models for LDPC codes.

SS codes, alternatively known as sparse regression codes, were first introduced by Barron and Joseph in 2010 for the additive white Gaussian noise (AWGN) channels [29]. Such codes were proven to be capacity-achieving under optimal, or maximum likelihood (ML), decoding [30, 31, 32]. Moreover, practical decoding schemes were introduced and proven to be capacity-achieving under proper power allocation. Such schemes include adaptive successive decoding and adaptive successive soft-decision decoding [33, 34, 35].

An iterative decoding approach for SS codes was introduced in [36, 37] by exploiting the similarity with the compressed sensing paradigm. The decoding (or inference) task in SS codes is to recover a sparse message, with a certain structure, based on noisy linear observations (or measurements); a task which is very similar to signal reconstruction in compressed sensing [38, 39]. Hence, the same iterative message-passing algorithms used in the compressed sensing literature can be adapted to decode SS codes. Moreover, the same way LDPC codes under message-passing require spatial coupling or irregular degree distribution to achieve capacity, SS codes also need spatial coupling or nonuniform power allocation, an alternative way to introduce irregularity in dense graphs, in order to achieve capacity.

An SS code is defined in terms of a coding matrix $\mathbf{F} \in \mathbb{R}^{M \times LB}$ made of i.i.d. Gaussian entries with zero mean and variance $1/L$. M represents the codeword length and we set $N = LB$. The coding matrix can be seen as L sub-matrices with B columns each. A codeword is generated by adding

[3]Not to confuse between the notion of sparsity in the graphical models and the notion of sparsity in the message. In the first context, it means that each node is connected to few other nodes in the graph. In the second context, it means that the message has few non-zero components.

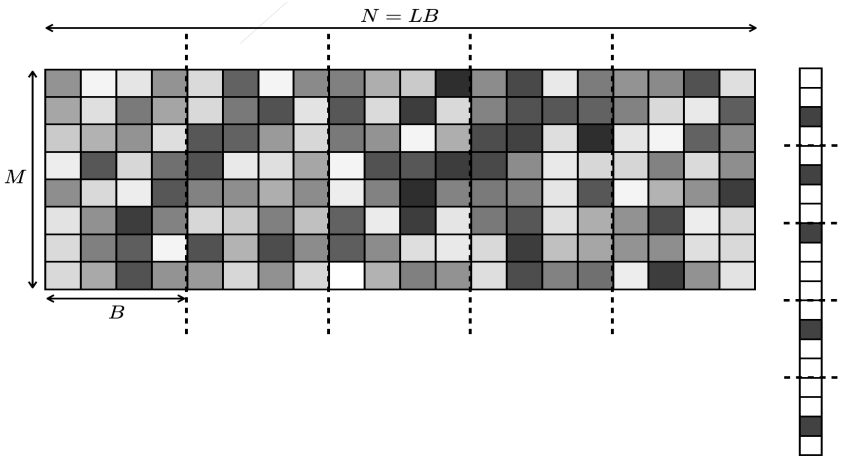

Figure 1.1: A Gaussian coding matrix $\mathbf{F} \in \mathbb{R}^{M \times LB}$ (left) with $B = 4$ and $L = 5$. A codeword of length M is constructed by choosing one column at random from each of the 5 sub-matrices of $\mathbf{F}$ then adding them. This corresponds to multiplying $\mathbf{F}$ by a sparse vector $\mathbf{s}$ (right).

uniformly at random L columns from $\mathbf{F}$, such that we have one and only one column coming from each sub-matrix (see Fig. 1.1). Therefore, a codeword can be represented as a linear projection of a sparse vector $\mathbf{s}$ using the coding matrix $\mathbf{F}$, where the sparsity in $\mathbf{s}$ is structured according to the size of the sub-matrices. Formally, the message $\mathbf{s}$ can be seen as a vector made of L *sections*, $\mathbf{s} = [\mathbf{s}_1, \dots, \mathbf{s}_L]$. Each section $\mathbf{s}_l$, $l \in \{1, \dots, L\}$, is a B-dimensional vector with a single non-zero component equal to 1. According to this construction, we end up with B^L possible codewords, and hence a rate of $L \log_2(B)/M$. The prior distribution of each message, assuming a uniform distribution over the codebook and independence between sections, reads

$$P_0(\mathbf{s}) = \prod_{l=1}^{L} p_0(\mathbf{s}_l) = \prod_{l=1}^{L} \frac{1}{B} \sum_{i=1}^{B} \delta_{s_{li},1} \prod_{j \neq i}^{B-1} \delta_{s_{lj},0}, \tag{1.1}$$

where s_{li} is the i^{th} component of the l^{th} section (here $i \in \{1, \dots, B\}$ and $l \in \{1, \dots, L\}$).

For a given memoryless channel $P_{\text{out}}(\mathbf{y}|\mathbf{Fs})$, the posterior distribution of an N-dimensional message $\mathbf{s}$ given the coding matrix $\mathbf{F}$ and the M-dimensional noisy observation $\mathbf{y}$ reads[4]

$$P(\mathbf{s}|\mathbf{y}, \mathbf{F}) = \frac{\prod_{l=1}^{L} p_0(\mathbf{s}_l) \prod_{\mu=1}^{M} P_{\text{out}}(y_\mu|[\mathbf{Fs}]_\mu)}{\int d\mathbf{s} \prod_{l=1}^{L} p_0(\mathbf{s}_l) \prod_{\mu=1}^{M} P_{\text{out}}(y_\mu|[\mathbf{Fs}]_\mu)}. \tag{1.2}$$

[4]The integral in the denominator boils down to a sum in our discrete case. However, we keep it here in the most general form.

One of the main contributions of this thesis is to extend the analysis of SS codes beyond the scope of AWGN channel and account for a large class of memoryless channels $P_{\text{out}}(\mathbf{y}|\mathbf{Fs})$.

In order to perform bit-decoding, either in the minimum mean-square error (MMSE) sense or in the maximum a posteriori (MAP) sense, one needs to find the marginals of the posterior distribution. The calculation of the marginals is computationally prohibitive, especially in the high-dimensional regime ($N, M \to \infty$ with a fixed rate), as it involves an exponential sum. We will see in Section 1.4 how to alleviate this computational burden by approximating the marginals using variants of the message-passing algorithm.

The performance of the asymptotic optimal decoder, i.e. using an exhaustive algorithm to compute the exact marginals when $N \to \infty$, is characterized by the point of non-analyticity[5] of the average normalized mutual information[6] in the asymptotic limit. This object is given by

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}[I(\mathbf{S};\mathbf{Y})], \tag{1.3}$$

where $I(\cdot;\cdot)$ is the well-known mutual information measure evaluated here on two random vectors. The expectation $\mathbb{E}[\cdot]$ is over the *ensemble* of coding matrices $\mathbf{F}$. Once again, this is an intractable quantity in our range of interest where $N \to \infty$.

Note that it is equivalent to look for the non-analyticity in the asymptotic conditional entropy $H(\mathbf{S}|\mathbf{Y})/N$ as the prior distribution is independent of the channel parameter. For the special case of LDPC codes, the non-analyticity point is defined by the point where the conditional entropy becomes strictly positive. This is because LDPC codes with proper degrees do not exhibit an *error-floor* in the "decodable region", a property that does not necessarily apply to all coding schemes.[7] In this thesis, we will keep the definition of the optimal performance in the most general form (i.e. in terms of non-analyticity), which can be used in both coding and estimation problems. In Section 1.6, we will see how such intractable quantities can be "guessed" using statistical physics techniques.

1.3 Symmetric Rank-One Matrix Factorization

Rank-one matrix factorization, or rank rank-one matrix estimation, is another inference problem with many applications in machine learning. These include community detection [40, 41, 42], sparse PCA [43], Spiked Wigner model [44] and matrix completion [45, 46]. In this thesis, we provide a probabilistic

[5]The non-analyticity is w.r.t. the channel parameter.

[6]Note that this quantity is not the same as Shannon's capacity. The latter is a generic quantity that computes the mutual information independent of any specific code construction. This mutual information is a model specific one for the SS codes.

[7]Low-density generator-matrix (LDGM) codes, for example, suffer from an error-floor.

formulation for the symmetric case of the problem and we analyze it in a Bayesian approach. Hence, similar graphical representations and analysis tools used for SS codes can be used here.

In the symmetric rank-one matrix factorization, one has access to a noisy observation $\mathbf{w} \in \mathbb{R}^{N\times N}$. As the name suggests, these observations are coming from a symmetric rank-one matrix $\mathbf{s}\mathbf{s}^\intercal$, with $\mathbf{s} \in \mathbb{R}^N$, subject to a certain noise. The entries of $\mathbf{s}$ are i.i.d. with a prior distribution P_0. The task is to recover the vector $\mathbf{s}$ from the noisy matrix $\mathbf{w}$. The entries w_{ij}, with $i, j = 1, \dots, N$, are observed through a probabilistic noisy channel $P_{\text{out}}(w_{ij}|s_i s_j)$.

The simplest example in the context of community detection is to have a binary vector $\mathbf{s}$ with $s_i \in \{-1, 1\}$. This vector represents the "membership" of each entry to one of two possible communities. The interconnection between the two communities can be represented by the symmetric rank-one matrix $\mathbf{s}\mathbf{s}^\intercal$ (the value of the ij^{th} entry of this matrix dictates whether s_i and s_j belong to the same community or not). Indeed, from observing this matrix one can recover the membership vector $\mathbf{s}$ up to a global flip of sign. However, one has access to a noisy version of $\mathbf{s}\mathbf{s}^\intercal$, which makes the problem more challenging [40, 41].

The previous example is perhaps a very simple one but it helps to motivate the set-up. What is important about symmetric rank-one matrix factorization is that many interesting problems in machine learning, such as asymmetric community detection in stochastic block model [42], can be formulated as such. Yet, more sophisticated prior distributions are used compared to the previous binary case (see Chapter 4).

The probabilistic channel $P_{\text{out}}(w_{ij}|s_i s_j)$ can be any complicated (possibly non-linear) noise model. However, a recent "channel universality" result [42, 47] yielded an equivalent Gaussian model that completely characterizes a large set of noise models (more details in Chapter 4). Formally, we define the model as follows

$$w_{ij} = \frac{s_i s_j}{\sqrt{N}} + \sqrt{\Delta} z_{ij}, \tag{1.4}$$

where $\mathbf{z} = (z_{ij})_{i,j=1}^N$ is a symmetric matrix with $Z_{ij} \sim \mathcal{N}(0, 1)$, $1 \le i \le j \le N$, and $\mathbf{s} = (s_i)_{i=1}^N$ has i.i.d components $S_i \sim P_0$. Hence, the posterior distribution of $\mathbf{s}$ given $\mathbf{w}$ reads

$$P(\mathbf{s}|\mathbf{w}) = \frac{e^{-\frac{1}{2\Delta}\sum_{i\le j}\left(\frac{s_i s_j}{\sqrt{N}} - w_{ij}\right)^2} \prod_{i=1}^N P_0(s_i)}{\int d\mathbf{s} e^{-\frac{1}{2\Delta}\sum_{i\le j}\left(\frac{s_i s_j}{\sqrt{N}} - w_{ij}\right)^2} \prod_{i=1}^N P_0(s_i)}. \tag{1.5}$$

Generally, one is interested in estimating $\mathbf{s}$ in the MMSE sense. The algorithmic difficulty in solving this inference problem resides in the computation of the posterior expectation, an intractable quantity in the high-dimensional regime. A more fundamental difficulty is to know what is information theoretically possible to do in such inference problem, regardless of the algorithm being

used. This requires first to detect the noise region in which the estimation is possible up to a "low" estimation error.[8] Second, it requires knowing the value of the MMSE, which is a priori unknown.[9] Knowing this will help in assessing the optimality of any algorithm in terms of both its range of operation and its estimation error.

Once again, the fundamental quantity that we are looking after is the asymptotic normalized mutual information

$$\lim_{N\to\infty} \frac{1}{N} I(\mathbf{S};\mathbf{W}). \tag{1.6}$$

Note that the only randomness here is in $\mathbf{S}$ and $\mathbf{W}$, unlike (1.3) where the graph itself is random and the ensemble average is needed. Note also that we are interested in $\mathbf{S}$ up to a global flip of sign, and hence it is equivalent to look for $I(\mathbf{S};\mathbf{W})$ or $I(\mathbf{S}\mathbf{S}^\intercal;\mathbf{W})$. Both of these quantities are computationally intractable in the high-dimensional regime ($N \to \infty$).

The non-analyticity point of (1.6) yields the detectability region for the problem. Moreover, due to the relation between the mutual information of a Gaussian model and the corresponding MMSE (known as the I-MMSE relation in the scalar case [48]), one can derive the value of the MMSE from (1.6). Thus, having a closed-form expression for (1.6) sets all our quantities of interest, a task that we opt to address in this thesis on a rigorous mathematical basis.

1.4 Factor Graph Representation and Message-Passing

Graphical models constitute a powerful framework to represent the statistical dependencies between a large number of random variables interacting in a complex domain. The relationships between these variables can be at a local or global level. Some of the most prominent graphical models are Markov random fields, Bayes nets and Factor graphs. Such models are heavily used in many disciplines such as coding theory, compressed sensing, machine learning, natural language processing and computer vision. In fact, graphical models are very effective not only in visualizing the statistical dependencies, but also in providing insights on how to devise low-complexity algorithms for the respective problems.

Both of our inference problems can be represented via a factor graph, also known as Tanner graph in some contexts. A factor graph is a *bipartite* graphical representation that illustrates the factorization of the posterior distribution

[8] In our Gaussian model (1.4), this corresponds to know the values of Δ where the estimation is possible. We expect that as Δ increases the estimation becomes harder. Moreover, we expect that there exists a critical value of Δ in the high-dimensional regime such that the low-error estimation becomes information theoretically impossible.

[9] The value of the MMSE is also computationally intractable in the high-dimensional regime.

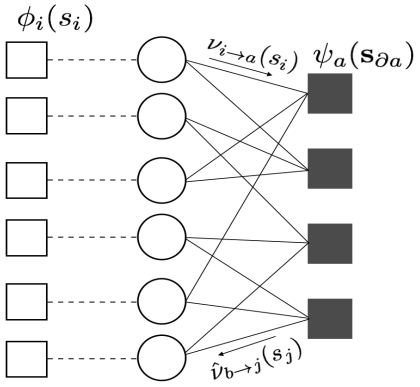

Figure 1.2: Factor graph representation for the posterior distribution (1.7). The s_i's are sitting on the variable nodes (circles). There are two types of factor nodes: factor nodes that correspond to the terms ϕ_i's in the posterior (plain squares), and factor nodes that correspond to the terms ψ_a's (colored squares). Each factor node a on the right is connected to the variable nodes in the set ∂a. The number of these connections is denoted by $|\partial a| := d_a$, which stands for the *degree* of check node a.

function, and hence enables efficient computation for many quantities of interest such as the marginal distributions. The factor graph consists of two interconnected[10] types of nodes: variable nodes which represent the variables of interest in the inference problem, and factor nodes which represent the "constraints" to be satisfied by the connected variable nodes according to the posterior distribution (see Fig. 1.2).

In general, consider a probability distribution function that takes the following form

$$p(\mathbf{s}) = \frac{1}{\mathcal{Z}} \prod_{a \in C} \psi_a\big(\{s_i, i \in \partial a\}\big) \prod_{i \in V} \phi_i(s_i), \tag{1.7}$$

where V denotes the set of variables, C the set of factors and ∂a the set of variables involved in the computation of given factor a (we usually use the shorthand notation $\mathbf{s}_{\partial a}$ in order to refer to the corresponding variables in this set). The term $\mathcal{Z}$ is a normalization constant. Such probability distribution can be represented via a factor graph as shown in Fig. 1.2. Note that both of our posterior distributions (1.2) and (1.5) follow this form.

A factor graph can be random or deterministic. The randomness appears in different forms: by drawing the edges between variable nodes and check nodes according to some probability distribution, by assigning a random degree to

[10] In a bipartite graph, an edge exists only between two nodes of different types.

each node, or by giving a certain random weight to each edge. Furthermore, a factor graph can be sparse or dense depending on the degree of the nodes (or the number of edges) relative to the overall number of variable nodes.

Both of our posterior distributions (1.2) and (1.5) are represented on dense factor graphs. Their respective factor graphs are bipartite complete in the sense that each factor node is connected to all variable nodes. While the factor graph of SS code (1.2) is random because of the Gaussian weight assignment through the coding matrix $\mathbf{F}$, that of symmetric rank-one matrix factorization (1.5) is deterministic.

1.4.1 Belief Propagation

The factor graph representation of the posterior distribution allows for efficient computation of the posterior marginals through iterative message-passing algorithms. A prominent algorithm in this family is the sum-product message -passing algorithm, also known as belief-propagation (BP) algorithm.

For a given posterior distribution of the form (1.7), the BP algorithm involves an exchange of local messages, or beliefs, along the edges between the variable nodes and the check nodes according to the following update rules

$$\nu_{i\to a}(s_i) = \frac{\phi_i(s_i)\prod_{b\in\partial i\setminus a}\hat{\nu}_{b\to i}(s_i)}{\sum_{s_i}\phi_i(s_i)\prod_{b\in\partial i\setminus a}\hat{\nu}_{b\to i}(s_i)} \tag{1.8}$$

$$\hat{\nu}_{a\to i}(s_i) = \frac{\sum_{\mathbf{s}_{\partial a\setminus i}}\psi_a(\mathbf{s}_{\partial a})\prod_{j\in\partial a\setminus i}\nu_{j\to a}(s_j)}{\sum_{\mathbf{s}_{\partial a}}\psi_a(\mathbf{s}_{\partial a})\prod_{j\in\partial a\setminus i}\nu_{j\to a}(s_j)}, \tag{1.9}$$

with the letters i, j, k used for variable nodes and a, b, c for check nodes. The set of directly connected nodes is similarly defined for a variable node i through ∂i and for a check node a through ∂a. The "\" operation is the difference operation defined over sets. We denote by $\nu_{i\to a}(s_i)$ the variable-to-check messages and by $\hat{\nu}_{a\to i}(s_i)$ the check-to-variable messages (see Fig. 1.2).

When there is only one relevant solution of the BP equations (1.8) and (1.9), the set of messages $\{\nu_{i\to a}, \hat{\nu}_{a\to i}\}$ can be obtained by an iterative method and the BP marginals are computed as follows

$$p_i^{\mathrm{BP}}(s_i) = \phi_i(s_i)\prod_{a\in\partial i}\hat{\nu}_{a\to i}(s_i). \tag{1.10}$$

Intuitively speaking, a variable-to-check message $\nu_{i\to a}(s_i)$ represents the belief variable node i has about its marginal probability of being s_i, and this is based on the information it receives from its neighborhood except from check node a. Similarly, a check-to-variable message $\hat{\nu}_{a\to i}(s_i)$ represents the belief check node a has about the marginal probability of variable i being s_i, and this is based on the information it receives from its neighborhood except from variable node i.

The BP update rules assume that the beliefs are conditionally independent, e.g. the beliefs advertised by variable node i to its neighborhood are conditionally independent given the value of this variable node. When the underlying factor graph is a tree, this assumption is satisfied and the BP rules are exact in computing the marginals [49]. Hence, the BP algorithm yields the optimal performance on a tree.

Clearly, the conditional independence assumption does not always hold for any factor graph. The more the graph is dense, the more likely this assumption is violated due to the loops, or cycles, that appear in the factor graph.[11] However, BP remains a very efficient algorithm in approximating the marginal distributions and it retains a strong connection with the optimal performance even in the presence of loops [50].

There exists a myriad of problems in coding theory [17], artificial intelligence [51], computer science [52] and statistical physics [53] where BP algorithm has been successfully applied. It helps here to note that the BP equations have two independent origins: in the field of statistical physics where it appeared in the 1935 Bethe-Peierls equations [54], and in the field of artificial intelligence where it was first introduced in the current form in Pearl's 1982 work [55].[12]

One of the most successful applications of BP algorithm in the last three decades is the decoding of forward-error-correction codes defined on sparse factor graphs. In particular, the iterative decoding algorithms for both turbo codes and LDPC codes can be seen as instances of BP algorithm. Besides its empirical success, the popularity of BP in coding theory stems from the fact that its behavior can be accurately analyzed and predicted in the asymptotic block-length limit. This is due to the notion of *density evolution* introduced by Richardson and Urbanke [18].

Density evolution is a simple recursion that tracks the performance of BP at every iteration instant t. The justification for density evolution is that the sparse factor graph boils down to a locally tree-like graph, also known as the computation graph, in the asymptotic limit.[13] Hence, the independence assumption can be reclaimed in order to evaluate the distribution of the BP messages. It turns out that predicting the performance of BP algorithm accounts for finding the fixed point solution of the density evolution recursion, a very important result that allows for rigorous analysis of coding problems on sparse graphs.

[11] Of course loops can exist in very sparse graphs, but this happens with low probability if we use random construction.

[12] Note that the similarity of the initials "BP" in belief-propagation and Bethe-Peierls is a pure coincidence.

[13] Using density evolution one can predict the BP performance when $N \to \infty$ then $t \to \infty$, which is not the realistic order of limits. However, it was shown that it is possible to swap this order of limits in some cases and have the same performance [56].

1.4.2 Approximate Message-Passing

The success story of message-passing algorithms, mainly the BP algorithm, in sparse graphical models is due to their accuracy in computing the marginal distributions in addition to their low-complexity implementation. For example, for a factor graph of size N and average node degree d: the overall number of BP messages is $\mathcal{O}(Nd)$ while the computational complexity of each message is exponential in d. Sparse graphs are typically characterized by a small node degree $d = \mathcal{O}(1)$, and hence BP represents a viable solution.

However, for a dense graphical model with $d = \mathcal{O}(N)$, the direct application of BP is computationally prohibitive for two reasons: *i*) The computation of each message, namely the check-to-variable message, requires an exhaustive sum. Moreover, for real-valued random variables, it is not a priori clear how to parameterize the messages; *ii*) The number of messages scales quadratically in the size of the problem.[14]

In the recent years, Donoho, Maleki and Montanari proposed a new iterative message-passing algorithm inspired by BP and suited for dense graphical models. This algorithm was coined *approximate message-passing* (AMP) [60, 61, 62]. AMP was first introduced for the compressed sensing problem as an alternative of standard convex optimization solutions [63, 64], which are not scalable, and fast *iterative soft thresholding* (IST) solutions [65], which suffer from weak performance. The authors showed that, in the asymptotic limit and with a proper scaling of the "measurement matrix", the two computational difficulties mentioned above can be overcome. This is possible via Gaussian approximations[15] and after adding a proper correction term to account for the correlation between the messages.

In fact, the curse of high density in the graph turned into a blessing! Due to the large number of incoming messages, the sum in the check-to-variable messages can be approximated by a Gaussian random variable via central limit theorem. Indeed, Gaussian messages can be parameterized by two terms, the mean and the variance. Hence, the first computational difficulty is alleviated since it is enough to keep track of only two parameters during the exchange of messages (See Fig. 1.3). Moreover, the high density ensures that all the ongoing messages from a given node are almost the same (the contribution of a single incoming message under a proper scaling of the weights is negligible). Therefore, the second computational difficulty is also alleviated and the number of messages is brought down to linear in the number of variables.

The main challenge in the derivation of the AMP algorithm is in the application of central limit theorem, which is not straightforward. Of course, the cycles in the dense graph make the messages correlated. Hence, the appli-

[14] Besides the computational burden, it is not a priori clear whether BP accurately computes the marginal distributions in the presence of many cycles. Yet, BP was proven to be asymptotically Bayes optimal for cases of our interest which involve dense graphs [57, 58, 59].

[15] Not to confuse with Gaussian BP. The Gaussian assumption here refers to the density of the graph and the distribution of the matrix entries, not to the input distribution.

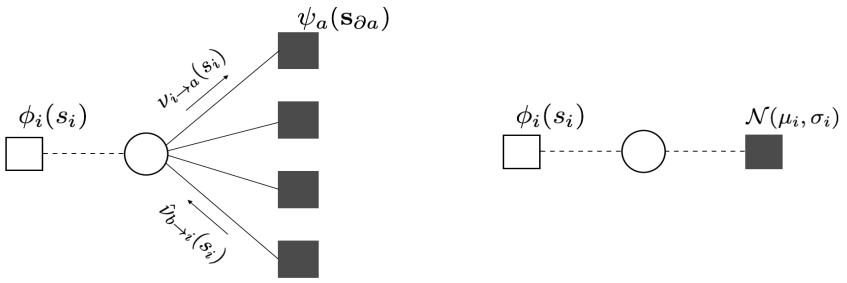

Figure 1.3: Gaussian approximation in the AMP algorithm. The high density in the graph simplifies the BP update rules where each node is subject to an effective Gaussian field. Instead of exchanging messages, the nodes keep track of the Gaussian parameters μ_i's and σ_i's.

cation of central limit theorem is not direct. The latter can be applied only after adding a correction term which, somehow miraculously, decorrelates the messages in the sum. This term is called the "Onsager" reaction term and it was first introduced by Onsager in 1936 [66].

It is worth noting that the AMP simplifications starting from the BP update rules are similar to the "TAP equations" introduced in 1977 by Thouless, Anderson and Palmer [67]. The authors originally derived these equations for a spin glass model defined on a complete graph, namely the celebrated Sherrington-Kirkpatrick (SK) model in statistical physics. The so-called Onsager term was also added to correct for the correlation without a complete mathematical justification. A rigorous mathematical justification for the use of Onsager term in the derivation of the AMP equations appeared recently by Bolthausen [68].

Note that in the context of compressed sensing, the Onsager term is the only difference between the AMP equations [60] and the IST equations [65]. It turns out that the missing Onsager term is the reason for performance degradation in IST.

The same way BP algorithm can be tracked on sparse graphs through density evolution, the performance of AMP algorithm on dense graphs can be tracked through an analogous tool called *state evolution*. At the same time AMP was introduced, extensive empirical simulations showed that state evolution tracks the performance of AMP in the context of compressed sensing [60]. Soon after that, Bayati and Montanari provided a rigorous justification for state evolution on a general basis that can be extended beyond the scope of compressed sensing [69].

Despite the analogy between density evolution and state evolution, their mathematical justifications are fundamentally different. While the density evolution relies on the locally tree-like structure of the sparse graph [18], the state evolution relies instead on the conditioning techniques introduced by Bolthausen [70]

1.5 Spatial Coupling

Spatial coupling was originally developed as an engineering tool to construct a new class of LDPC codes with better performance. It first appeared under the name of convolutional LDPC codes [22, 23, 24]. It was observed that the local coupling of several LDPC ensembles with a proper termination considerably improves the performance under BP decoding.[16] The density evolution analysis was then adapted to track the performance of spatially coupled codes. Moreover, it was proven that the BP algorithmic performance, or threshold, can be improved up to the optimal performance.[17] This phenomenon was termed *threshold saturation* [25, 26, 27].

Spatial coupling was then applied to various graphical models, both sparse and dense, where it was shown to boost the performance under iterative algorithms. Furthermore, the threshold saturation was shown to be a universal phenomenon. The problems where spatial coupling was successfully applied include code division multiple access (CDMA) [71, 72], satisfiability [73], compressed sensing [74, 75, 76], SS codes [77, 78], Curie-Weiss model [79, 80] and more generally any coupled graphical model tracked by a coupled scalar recursion [81, 82].

Spatial coupling can be represented via a graphical model starting from the original factor graph. Assume that we have a factor graph of size N. We take several instances of this factor graph and we place them next to each other on a chain of length Γ . Then, we locally couple the underlying factor graphs with a coupling window w to obtain a bigger factor graph of size $\Gamma \times N$ (see Fig. 1.4). In the resulting factor graph, each variable node is connected to the corresponding check nodes of the same underlying factor graph and to the check nodes of the neighboring factor graphs. This construction creates a spatial dimension, along the positions of the chain, that will help the algorithm. The second step in constructing efficient spatially coupled graphs is to introduce a *seed* at a certain position of the chain. This seed can be introduced as a side information which helps the algorithm at the boundaries and initiates a "wave" that propagates inwards and boosts the performance.

Note that in order to have an efficient spatially coupled scheme, both of the steps mentioned above should be applied. Having a spatial construction without a seed makes the resulting spatially coupled model of no advantage over the underlying "uncoupled" model. Likewise, applying a seed on the underlying model without constructing a spatial dimension to carry the wave is of no advantage.

There are several ways to construct the spatially coupled graphs. In the context of SS codes, for example, a spatially coupled code corresponds to a band-diagonal coding matrix as explained in Chapter 2. Moreover, the seed

[16] Spatial coupling modifies the construction of the code, or the graphical model in general, not the algorithm. The new spatially coupled construction is more robust against noise and it performs better under the same BP algorithm

[17] The performance under optimal algorithm, e.g. ML or exhaustive search algorithm.

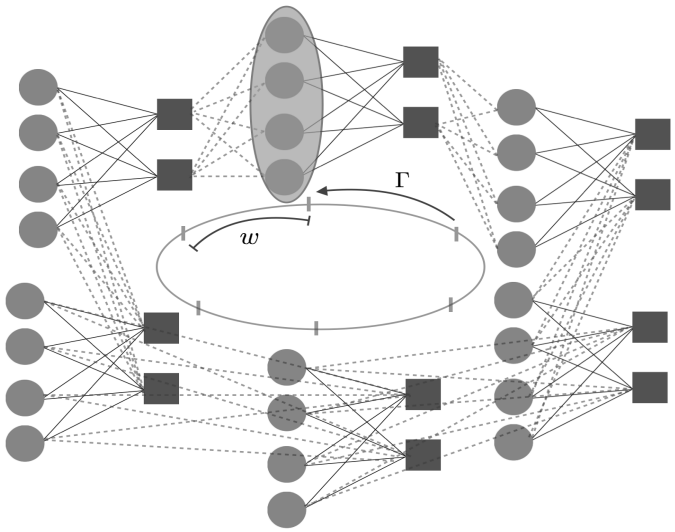

Figure 1.4: Spatial coupling of a dense graphical model. A coupling chain of length $\Gamma = 6$ is taken where one factor graph is placed at each position. The coupling window is $w = 1$. At each of the 6 positions, the variable nodes (circles) are connected to the corresponding factor nodes (squares) via solid lines and to the w-neighbor factor nodes via dotted lines. The seed is illustrated by the elliptical shape.

can be introduced in different forms. One way is to fix the values of the variable nodes at a certain position in order to "help" the algorithm (i.e. provide perfect side information). Based on the application scenario, it is not always possible to provide such perfect information. Hence, the seed is sometimes introduced by providing partial side information as illustrated in Chapter 3. The desired size of the seed and its effect on the speed of the propagation wave was recently analyzed in [83, 84]. Note that the seed induces a small loss in terms of degrees of freedom. However, this loss can be amortized by taking a proper asymptotic regime.

Interestingly, spatial coupling turns out to be a versatile tool. Besides its practical advantage as an engineering tweak that boosts the performance, spatial coupling can be used as a proof technique to compute many quantities of interest which are a priori intractable [85]. More precisely, both the underlying model and the spatially coupled one share the same information theoretic quantities, such as the normalized mutual information and the optimal performance. Hence, one can gain some insights about the underlying problem by studying a spatially coupled version of it.

Therefore, even if the problem at hand does not provide the freedom of constructing a spatially coupled model in practice,[18] one can still use spatial

[18]In some problems we can not control the design of the factor graph, unlike the coding problems. Community detection via symmetric rank-one matrix factorization is one example where the factor graph is dictated by the model.

coupling for an auxiliary model. Intuitively speaking, since the low-complexity algorithm on the auxiliary model is optimal by the threshold saturation phenomenon, it is easier to compute the information theoretic quantities on that model and then apply them to the underlying model. In this thesis, we are going to benefit from both aspects of spatial coupling, i.e. the engineering and the theoretical aspects.

1.6 Connection to Statistical Physics

Over the last century or so, statistical physics techniques have developed with the aim to predict and describe the macroscopic behavior of systems involving an interaction of a large number of degrees of freedom (possibly in a random environment). A very common phenomenon that occurs in these systems is the *phase transition* phenomenon, which is characterized by an abrupt change of behavior. Such systems are ubiquitous in nature and they have been the subject of study in statistical physics of *spin glasses* [3, 86].

Interestingly, numerous problems in different fields of science and engineering can be formulated as spin glass models. These include LDPC codes [87], turbo codes [88], CDMA systems [89], compressed sensing [75], satisfiability [90] and neural network models in machine learning [91, 92].

Many powerful techniques used in statistical physics have rigorous mathematical justifications such as correlation inequalities [93], decay of correlations [94] and interpolation methods [95]. However, other techniques are used as prediction tools to "guess" some quantities of interest which are notoriously hard to compute. Such heuristic techniques are based on mathematically unjustified assumptions (or ansatz). One of these techniques is the celebrated replica method [3]. Despite the lack of rigor, the predictions of the replica method, which are highly nontrivial, have been extensively applied to many probability-based information processing fields.

In a certain class of problems coined as the *mean-field* models,[19] the replica predictions have been proven to be exact on some paradigmatic examples (e.g. the SK model by Talagrand [96]). Moreover, the replica method has been able to successfully reproduce many results for problems with already known solutions [97, 98]. Hence, it is believed that the predictions of the replica method are accurate for mean-field models.

Over the last decade or so, a plethora of rigorous work has been conducted in order to prove the exactness of the replica predictions. In many instances, it was shown that the replica predictions provide tight bounds on many fundamental quantities[20] [99, 100, 101]. Recently, it was proven that the replica predictions are rather exact, i.e. met with equality, for many interesting problems [85, 102, 103, 104].

[19] Complex models that can be studied by looking at the behavior of a simpler model.

[20] Namely the asymptotic mutual information or conditional entropy.

In this thesis, the use of replica method is two-folded: *i*) We use the potential function, which is computed by the heuristic replica method, as an analysis tool for SS codes in order to prove the threshold saturation phenomenon. Our analysis in this part does not rest on the rigor of the replica predictions, but rather on some of the potential function's properties that we prove.[21] *ii*) We prove that the replica predictions for the symmetric rank-one matrix factorization problem are exact. This yields an explicit expression for the asymptotic mutual information along with major practical implications.

1.6.1 Spin Glass Models

A spin glass system consists of N degrees of freedom, or spins, denoted by $s_i, i = 1, \dots N$. Every configuration $\mathbf{s} = [s_1, \dots, s_N]$, or state, of the spins is associated with a *Hamiltonian* $\mathcal{H}(\mathbf{s})$. This Hamiltonian represents the cost function of a given state $\mathbf{s}$ and reflects the interaction between the spins. When the system is in equilibrium, statistical physics postulates that the probability of a given state is related to the Hamiltonian through the Boltzmann distribution[22]

$$P(\mathbf{s}) = \frac{e^{-\beta \mathcal{H}(\mathbf{s})}}{\mathcal{Z}}, \tag{1.11}$$

where $\mathcal{Z}$ is the normalization function, known as the partition function in statistical physics. The parameter β is defined to be the inverse temperature in statistical physics language. The precise value of β is not relevant to our discussion, hence we assume that $\beta = 1$ in this thesis. Note that both of our posterior distributions (1.2) and (1.5) can be formulated as a Boltzmann distribution with a certain Hamiltonian.

The randomness of the spins w.r.t. the Boltzmann distribution is called *annealed* randomness in the statistical physics literature. Moreover, given the value of the spins, the Hamiltonian function itself in (1.11) can be random due to the random interaction between the spins. For example, the Hamiltonian of SS codes induced from the posterior distribution (1.2) is random w.r.t. the noisy observations $\mathbf{y}$ and the code ensemble $\mathbf{F}$. This external randomness is called *quenched* randomness.

A central object in statistical physics is the *free energy* of the system defined as

$$f_N = -\frac{1}{N}\mathbb{E}[\log(\mathcal{Z})], \tag{1.12}$$

where $\mathcal{Z}$ is the partition function in (1.11) and $\mathbb{E}$ denotes the expectation over the quenched random variables. Note that the negative sign here is a matter

[21] One can prove that the replica formulation is intimately related to the algorithmic performance of message-passing.

[22] Also called Gibbs distribution.

of convention. In this thesis, we use the negative sign to denote the free energy while the positive counterpart refers to the free entropy.

The free energy is a fundamental quantity in statistical physics. In fact, the behavior of the system and many physical properties can be extracted from the free energy. In the limit $N \to \infty$, the non-analyticity of (1.12) w.r.t. some model parameters[23] indicates a phase transition in the system. This is very reminiscent of the behavior of the normalized mutual information discussed earlier. Actually, one can show that computing the normalized mutual information for both of our problems is equivalent to compute the free energy (the two quantities are equal up to a trivial term, see Chapter 4.). More generally, when the Boltzmann distribution (1.11) represents a posterior measure for an inference problem in the Bayesian setting, the free energy in statistical physics is closely related to the normalized mutual information (or conditional entropy) in information theory. Hence, it is equivalent to study either of them.

The regime of our interest is the *thermodynamic limit*, i.e. the limit of large number of spins $N \to \infty$. In this limit, statistical physics postulates that the free energy is self-averaging, in the sense that the free energy concentrates around its expectation. Again, the computation of such quantity in the limit $N \to \infty$ is computationally prohibitive. The replica method provides a "trick" to compute the free energy in the thermodynamic limit as we will see in the following.

1.6.2 Replica Method

The replica method is a heuristic analytical tool developed in statistical physics in order to compute the free energy, or equivalently the normalized mutual information, in the thermodynamic limit. The replica method is based on mathematically unjustified manipulations and it also involves some assumptions on the structure of the solution. A typical assumption is the *replica symmetry* (RS) assumption.[24] Nevertheless, the predictions provided by the non-rigorous replica method are believed to be accurate for mean-field models. Recently, the replica predictions have been proven to be exact on various nontrivial problems. Note that the exactness of the replica predictions means that the final solutions it provides, e.g. the expression for the free energy, are exact. This does not necessarily mean that the replica method itself is rigorous.[25]

The replica method consists of "replicating" the partition function in (1.12) n times. This helps in transforming the expectation of the logarithm into a logarithm of an expectation, which involves the n^{th} moment of the partition function (see Eq. (1.14) below). The replica method then involves a couple of

[23] This can be a noise parameter for example. The non-analyticity can also occur as a function of the inverse temperature β.

[24] It assumes that the "replicas" of the system are symmetric under permutation of their labels.

[25] "At the present time, it is difficult to see in the physicist's replica method more than a way to guess the correct formula" — Talagrand, 2003 [105].

mathematically unjustified steps. This includes an exchange of limits and an analytic continuation argument that assumes $n \to 0$.

In order to compute the free energy (1.12) in the thermodynamic limit, the replica method, which has many incarnations, computes the following

$$\lim_{N\to\infty} \frac{-\mathbb{E}[\log(\mathcal{Z})]}{N} = \lim_{N\to\infty} \lim_{n\to 0} \frac{\partial}{\partial n} \frac{-\log(\mathbb{E}[\mathcal{Z}^n])}{N}. \tag{1.13}$$

This former step can be justified by a simple differentiation where one can show that

$$\frac{\partial}{\partial n} \log(\mathbb{E}[\mathcal{Z}^n]) = \frac{\mathbb{E}[\mathcal{Z}^n \log(\mathcal{Z})]}{\mathbb{E}[\mathcal{Z}^n]}. \tag{1.14}$$

The replica method then assumes that we can exchange the order of $\lim_{n\to 0}\partial/\partial n$ and $\lim_{N\to\infty}$ [106]. The next trick is to assume that n is an integer despite the fact that we take the limit $n \to 0$ afterwards. The reason for this is that computing the n^{th} moment might be more feasible than averaging the partition function itself. The expectation and the limit $N \to \infty$ are then computed using the saddle-point integration scheme.

Using these manipulations, the replica prediction for the free energy (1.12) in the thermodynamic limit is given in a variational form

$$\lim_{N\to\infty} f_N \approx \min_{m\in\mathcal{C}} f_{\mathrm{RS}}(m), \tag{1.15}$$

where m is a trial parameter defined on a certain set $\mathcal{C}$ and f_{RS} is the RS *potential* function. The "$\approx$" symbol is used to stress on the lack of rigor in the replica method. It turns out that the RS potential represents a free energy of a simple problem that we can easily compute. Besides the trial parameter m, the RS potential also depends on some other parameters of the original model (e.g. the inverse temperature β or the noise parameter denoted by Δ, see Fig. 1.5). Here, we show explicitly the dependency on m while all other dependencies are implicit in f_{RS}.

Interestingly, the replica formulation does not only predict the asymptotic free energy, but it also retains a very close connection with the algorithmic message-passing performance when solving an inference problem on graphical model. Indeed by computing the free energy, the replica solution can predict the optimal performance as the non-analyticity point (i.e. the optimal phase transition). Moreover, the replica formulation can predict the performance of the iterative message-passing algorithm by looking at the stationary points of the RS potential (see Fig. 1.5). In fact, one can show that the stationary points of the RS potential correspond to the fixed point solutions of the density evolution, or state evolution, tracking the performance of the message-passing algorithm. Hence, by looking at the RS potential one can predict the algorithmic phase transition as well.[26] Furthermore, the trial parameter m involved

[26]The algorithmic phase transition corresponds to the point where the behavior under a low-complexity algorithm exhibits an abrupt change. The optimal phase transition corresponds to the abrupt change under the optimal algorithm (e.g. exhaustive search).

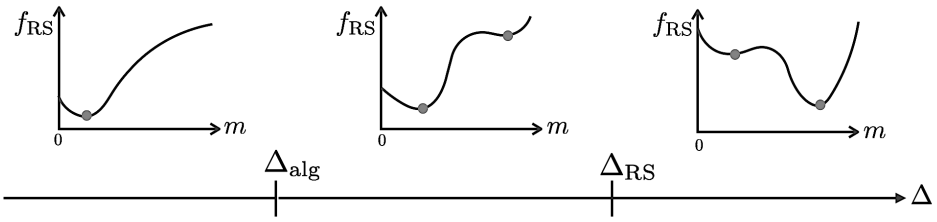

Figure 1.5: The RS potential as a function of the trial parameter m and the external noise parameter Δ. The potential is plotted for three different values of Δ. The minima of the potential as a function of m are marked by dots. The potential has a unique minimum below Δ_{alg}. The local minimum appears as Δ increases beyond Δ_{alg}. The two minima change roles, i.e. from global to local and vice versa, after Δ_{RS}.

in the replica formulation can have some practical interpretations in terms of the probability of error or MMSE.

For illustration purposes, the RS potential as a function of the trial parameter m and an external noise parameter Δ is shown in Fig. 1.5. For each fixed value of Δ, the potential behaves in a certain fashion as a function of m. For small enough Δ, the potential has a unique global minimum.[27] As Δ increases, the potential starts to develop a local minimum. The value of Δ where the local minimum starts to appear is called the *algorithmic threshold* denoted by Δ_{alg}. This is because the procedure tracking the message-passing algorithm can be interpreted as a descent algorithm on the potential function: as long as the potential has a unique minimum the algorithm succeeds; as soon as the potential develops a local minimum the algorithmic behavior changes discontinuously since the algorithm gets stuck at the new local minimum. As Δ increases further, the local minimum becomes a global minimum. The value of Δ where the two minima are global minima, i.e. have the same f_{RS} value, is called the *optimal threshold* (or the *potential threshold*) denoted by Δ_{RS}. One can see that at Δ_{RS}, the replica solution of the free energy, defined by $\min_m f_{\mathrm{RS}}$, develops a point of non-analyticity.

The RS potential is used in Chapter 2 as an analysis tool for spatially coupled SS codes. We prove that spatial coupling improves the performance of the AMP algorithm and moves the algorithmic threshold up to the potential threshold. In Chapter 4, we address the validity of the replica prediction. In particular, we show that (1.15) holds with equality for symmetric rank-one matrix factorization.

Note that the potential shown in Fig. 1.5 is a special case where a *first-order*

[27] For inference problems with zero error-floor, this minimum occurs at $m = 0$ as in the LDPC case. However, we present here the analysis in the most general form where an error-floor can exist.

phase transition occurs, i.e. the potential has a maximum of three stationary points (two minima and one maximum). Other types of phase transitions, or no phase transition, are possible in spin glass models. It turns out that the case of first-order phase transition is the most relevant to our discussion when dealing with our problems at hand.

1.6.3 Interpolation Method

The heuristic replica method was applied by Giorgio Parisi in 1980 to compute the free energy of the SK model [107], an archetypal mean-field model in statistical physics named after Sherrington and Kirkpatrick [108]. Parisi's free energy remained a conjecture for more than two decades. A rigorous proof showing that Parisi's formula for the SK model is exact appeared by Talagrand in 2006 [96]. The proof idea relies heavily on the interpolation method developed by Guerra and Toninelli [109].

The interpolation method is a mathematical rigorous tool that we will use extensively in Chapter 4. The idea is that we "interpolate" between two models in order to compare their respective free energies. Typically, one model is the original model and the other is the simple model guessed from the replica solution. In many instances, the interpolation method is able to prove that the replica prediction provides an upper bound for the free energy.[28] In this thesis, we will use the interpolation method on different occasions in order to prove the validity of the replica prediction.

Formally, assume that we have two spin glass models labeled by the letters a and b where $\left(\mathcal{H}^a(\mathbf{s}), \mathcal{Z}^a, f_N^a\right)$ and $\left(\mathcal{H}^b(\mathbf{s}), \mathcal{Z}^b, f_N^b\right)$ represent their respective Hamiltonians, partition functions and free energies. We then define a "linear" interpolated Hamiltonian as follows

$$\mathcal{H}_t(\mathbf{s}) = t\mathcal{H}^a(\mathbf{s}) + (1-t)\mathcal{H}^b(\mathbf{s}), \tag{1.16}$$

with $t \in [0,1]$. The resulting Hamiltonian interpolates between the two models as a function of t. It yields model a at $t = 1$ and model b at $t = 0$. Note that for illustration purposes, the interpolation path is assumed to be linear here. In practice, the proper interpolation is more complicated and needs to be carefully chosen (see Chapter 4). However, the interpolation path has to satisfy the following: $\mathcal{H}_0(\mathbf{s}) = \mathcal{H}^a(\mathbf{s})$ and $\mathcal{H}_1(\mathbf{s}) = \mathcal{H}^b(\mathbf{s})$.

Let $\mathcal{Z}_t$ and $f_{N,t}$ represent the partition function and the free energy of the interpolated model. Hence, the fundamental theorem of algebra gives

$$\int_0^1 dt \frac{d}{dt} f_{N,t} = f_{N,1} - f_{N,0} := f_N^a - f_N^b. \tag{1.17}$$

Therefore, one can compare the free energies of the two models by looking at the sign of the derivative in the left-hand side term of (1.17). The beauty of

[28] This is only possible when the interpolation method yields a non-negative remainder term. In many other instances, it is hard to argue about the sign of the remainder and different proof techniques are needed.

the interpolation method is that even if this term is hard to compute, one can often argue about its sign by choosing a proper interpolation path.

One of the main challenges in applying the interpolation method, other than choosing the proper interpolation path, is to choose the two models a and b. In some special instances when deriving an upper bound on the free energy, the second model can be guessed from the replica method. However, in order to have a complete proof of the replica solution, the choice of the models becomes much more challenging and the interpolation method has to be applied more than once as we will see in Chapter 4.

1.7 Organization and Main Contributions of the Thesis

This thesis addresses two different, yet very related, topics. The first topic is a channel coding problem that spans Chapter 2 and 3. In these chapters, we study SS codes which constitute a recent class of forward-error-correction codes. The second topic, studied in Chapter 4, is the symmetric rank-one matrix factorization which can be used to model many interesting problems in machine learning and statistics. In both of these topics, we formulate the problems at hand as Bayesian inference problems that we study in the asymptotic high-dimensional regime. It turns out that both formulations can be represented on dense factor graphs. This common feature suggests that similar message-passing algorithms can be used to solve our inference problems. Furthermore, the graphical representation spurs the employment of the spatial coupling technique.

In addition to that, both of our problems exhibit several forms of phase transitions in the high-dimensional regime. This behavior is very reminiscent of the behavior of spin glass models studied in statistical physics. Therefore, the use of statistical physics techniques, such as the replica method or the RS potential function, is a very natural choice for analysis.

In **Chapter 2**, we introduce the SS codes. A main contribution of this chapter is the extension of the application and analysis of SS codes beyond the scope of the AWGN channel, the channel for which these codes were first introduced and used so far. This extension is possible via the introduction of a mapping function that maps the real-valued Gaussian entries to the channel input alphabet of a general memoryless channel. The concatenation of both the deterministic mapping function and the physical probabilistic channel can be then seen, from the algorithmic perspective, as a new effective channel. Hence, the generalized approximate message-passing (GAMP) algorithm, which is a generalization of the AMP algorithm [110], is adapted for the decoding task of SS codes. Another important contribution of this chapter is the rigorous analysis of spatial coupled instances of SS codes. This allows us to show that spatially coupled SS codes "universally" achieve capacity on general channel

under low-complexity GAMP decoding.

The adoption of the GAMP algorithm for a channel coding problem is an unprecedented approach that we present in Chapter 2. A central result in this chapter is the proof of threshold saturation for spatially coupled SS codes, a result which ensures that SS codes universally achieve capacity on general memoryless channels.[29] The proof of the threshold saturation is based on the state evolution analysis and the RS potential predicted by the statistical physics' replica method. Our threshold saturation proof follows the lines of [82]. However, we consider a more general coupling construction. More specifically, we assume a general coupling strength which is not necessarily uniform or symmetric as in [82]. This relaxation could significantly improve the performance in practice [111].

The proof requires an extension of the scalar state evolution analysis to the vector case of the spatially coupled system. The main strategy is to assume a "bad" fixed point solution of the spatially coupled state evolution and to calculate the change in the RS potential due to a small *shift* in two different ways: *i*) by second-order Taylor expansion, *ii*) by direct evaluation. We then show by contradiction that for large coupling parameters and as long as the rate is below the potential rate, the state evolution converges to the "good" fixed point. Hence, by taking the proper limit one can show that the algorithmic rate of the spatially coupled system saturates the potential rate (i.e. the former is lower bounded by the latter). Moreover, we show by analytical calculation that the potential rate tends to capacity and the error floor (when it exists e.g., in the AWGN case) vanishes in the proper limit. Furthermore, we provide a closed-form formula for the algorithmic rate of the uncoupled code ensemble in terms of a Fisher information.

It is worth noting that carrying out this program presents some technical difficulties while bounding the second-order Taylor expansion of the coupled state evolution which do not appear in [82]. This is due to the special form of the state evolution tracking the performance of the GAMP algorithm for SS codes over general channels.

Note also that the analysis in this chapter does not rely on the exactness of the replica predictions, but rather on the intimate connection between the RS potential and the state evolution as already explained in Section 1.6.

Chapter 3 presents a novel application of SS codes and GAMP algorithm for a source coding problem, namely the distribution matching problem. Distribution matching is the building block for probabilistic shaping, a problem that has recently attracted lots of attention in long-haul fiber optical communications.

We present an exact matcher based on a position modulation, that introduces sparsity in the source, followed by a simple quantization of a Gaussian signal. This yields any discrete target distribution. The quantizer can be seen

[29] Using low-complexity GAMP algorithm and under the assumption that state evolution tracks the performance of the spatially coupled system.

as a deterministic nonlinear channel. At the receiver, the dematcher exploits the sparsity in the source and performs low-complexity dematching based on the GAMP algorithm as done for the SS codes.

We show that GAMP algorithm and spatial coupling lead to an asymptotically optimal performance, in the sense that the rate tends to the entropy of the target distribution with vanishing reconstruction error in a proper limit. Furthermore, we assess the performance of the GAMP algorithm on practical Hadamard-based operators. A remarkable feature of our approach is the ability to perform joint channel coding and distribution matching at the symbol level, a promising solution for probabilistically-shaped coded modulation schemes [112, 113, 114]. Note that the analysis done in this chapter is mostly numerical, whereas all the proofs and the theoretical guarantees can be derived from Chapter 2.

In **Chapter 4**, we introduce symmetric rank-one matrix factorization, a prominent problem with many applications in machine learning and high-dimensional statistics. The central result is an explicit expression for the normalized mutual information in the high-dimensional regime. Indeed, the expression for the normalized mutual information is given by the replica prediction. Our main contribution is to show that this expression is exact. Our proof strategy uses spatial coupling as a proof technique.

We first show that the replica solution provides an upper bound for the asymptotic normalized mutual information. This is done by applying the interpolation method between the original model and a simple denoising model guessed from the replica solution. The proof that the replica solution also yields a lower bound is quite more involved. Roughly speaking, the lower bound can be established as long as the RS potential has a unique minimum. This is possible via the I-MMSE relation and the suboptimality of the AMP algorithm. However, in the most interesting regime where the RS potential develops a second local minimum (i.e. the presence of first-order phase transition), the proof requires the use of an "auxiliary" spatially coupled model. It turns out that the spatially coupled model is easier to analyze. This is because of the threshold saturation phenomenon where the algorithmic threshold coincides with the optimal threshold. Being able to deduce the information theoretic optimal threshold from the algorithmic threshold is the crux of proving the lower bound, and hence the exactness of the replica prediction.

More specifically, we show, using the interpolation method, that the normalized mutual information is the same for both the spatially coupled model and the underlying (uncoupled) model. The interpolation here, although similar in spirit, is different than the one used to prove the upper bound. We apply the interpolation method twice: *i*) interpolation between the spatially coupled model and a sequence of "independent" underlying models, *ii*) interpolation between the spatially coupled model and a sequence of "fully connected" underlying models. This allows to sandwich the spatially coupled model between two models that are asymptotically equivalent to a single underlying model.

This result together with the threshold saturation result[30] allow to extend the proof of the lower bound to the whole region of parameters. Note that in the process of showing the invariance of the mutual information under spatial coupling, we derive the existence of the thermodynamic limit using the interpolation method, once again, and a standard argument for *superadditive* sequences.

The closed-form expression for the mutual information helps in detecting the optimal phase transition and computing the MMSE. Hence, we are able to assess the optimality of any algorithm in terms of both its range of operation and its estimation error. In particular, we are able to provide an exact formula for both the "vector" MMSE and the "matrix" MMSE from the RS potential.

A remarkable practical implication of Chapter 4 is the proof that the AMP algorithm is optimal in a large region of parameters. Nonetheless, we show that, in a specific region that we refer to it as the "computational gap", the currently known polynomial algorithms (in particular AMP and spectral methods) fail to reach the information theoretic optimal performance. Note that the spatially coupled construction is mainly used in this chapter for the purposes of the proof. However, the spatially coupled matrix factorization is an interesting model in itself, specially in view of the fact that the computational gap disappears for such model. Hence, one can imagine many possible applications where spatial coupling is involved.

In **Chapter 5**, we summarize the findings of this thesis and we discuss some open challenges and potential research directions.

We would like to end this introductory chapter by pointing out that this thesis heavily relies on the state evolution analysis for the AMP algorithm. However, the scope of this work does not cover the derivation of the AMP and the state evolution. Interested readers can refer to [60, 69, 115]. Moreover, one of our central results is the proof that the predictions given by the heuristic replica method of statistical physics are accurate. Nonetheless, we do not intend to give a recipe on how to perform the replica trick. Interested readers can refer to the statistical physics literature [3, 105].

[30]The threshold saturation result for symmetric rank-one matrix factorization is established following the same proof techniques used for SS codes in Chapter 2.

Universal Sparse Superposition Codes

2

Sparse superposition codes, or sparse regression codes, constitute a new class of codes which was first introduced for communication over the additive white Gaussian noise (AWGN) channel. It has been shown that such codes are capacity-achieving over the AWGN channel under optimal maximum-likelihood decoding as well as under various efficient iterative decoding schemes equipped with power allocation or spatially coupled constructions. In this chapter,[1] we generalize the analysis of these codes to a much broader setting that includes all memoryless channels. We show, for a large class of memoryless channels, that spatial coupling allows an efficient decoder, based on the generalized approximate message-passing (GAMP) algorithm, to reach the potential (or Bayes optimal) threshold of the underlying (or uncoupled) code ensemble. Moreover, we argue that spatially coupled sparse superposition codes universally achieve capacity under GAMP decoding by showing that both the error floor vanishes and the potential threshold tends to capacity as one of the code parameter goes to infinity. Furthermore, we provide a closed-form formula for the algorithmic threshold of the underlying code ensemble in terms of a Fisher information. Relating an algorithmic threshold to a Fisher information has theoretical as well as practical importance. Our proof relies on the state evolution analysis and uses the potential method developed in the theory of low-density parity-check (LDPC) codes and compressed sensing.

[1] The content of this chapter is based on a joint work with J. Barbier and N. Macris [116].

2.1 Introduction

Sparse superposition (SS) codes, or sparse regression codes, were first introduced by Barron and Joseph [29] for reliable communication over the additive white Gaussian noise (AWGN) channel. The SS codes were then proven to be capacity-achieving under adaptive successive decoding along with power allocation [33, 34]. Later on, the connection between SS codes and compressed sensing was made in [36]. The decoding of SS codes can be interpreted as an estimation of a sparse signal, with structured prior distribution, based on a relatively small number of noisy observations. Hence, the approximate message-passing (AMP) algorithm, originally developed for compressed sensing, was adapted in [36] to decode SS codes where it exhibited better finite-length performance than adaptive successive decoding. SS codes, with appropriate power allocation on the transmitted signal, were then proven to achieve capacity under AMP decoding [37]. Furthermore, the extension of the state evolution (SE) equations, originally developed to track the performance of AMP for compressed sensing [69], was proven to be exact for SS codes in [37].

The idea of *spatial coupling* was originaly introduced for low-density parity-check (LDPC) codes under the name of LDPC convolutional codes [22, 23]. Spatial coupling has been then successfully applied to various problems including error correcting codes [25], code division multiple access (CDMA) [71, 72], satisfiability [73], and compressed sensing [74, 75, 76]; where it has been shown to boost the performance under iterative algorithms. Recently, spatial coupling was applied to SS codes in [77, 78]. The construction of coding matrices for SS codes with local coupling and a proper termination was shown to considerably improve the performance. Moreover, practical Hadamard-based operators were used in [77] to encode SS codes, where they showed better finite-length performance than random operators under AMP decoding. The spatially coupled construction used in [77, 78] has many similarities with that introduced in the context of compressed sensing [117, 111, 115]. Empirical evidence shows that spatially coupled SS codes perform much better than power allocated ones and that they achieve capacity under AMP decoding without any need for power allocation. This motivated the initiation of their rigorous study [118] using the *potential method*, originally developed for the spatially coupled Curie-Weiss model [79, 80] and LDPC codes [81, 27, 82]. The phenomenon of *threshold saturation* for AWGN channels was shown in [118], i.e. the *potential threshold* that characterizes the performance of SS codes under the Bayes optimal minimum mean-square error (MMSE) decoder can be reached using spatial coupling and AMP decoding. Moreover, the potential threshold itself was shown to achieve capacity in the large input alphabet size limit.

Threshold saturation was first established in the context of spatially coupled LDPC codes for general binary input memoryless symmetric channels in [27, 26], and is recognized as the mechanism underpinning the excellent performance of such codes [24]. It is interesting that essentially the same phenomenon can be established for a coding system operating on a channel with

continuous inputs. This result was a stepping-stone towards establishing that spatially coupled SS codes achieve capacity on the AWGN channel under AMP decoding [118]. Note that a similar (but different) potential to the one used in [118] has been introduced in the context of scalar compressed sensing [69, 82]. It is interesting that the potential method goes through for the present system involving a dense coding matrix and a fairly wide class of spatial couplings. Related results on the optimality of spatial coupling in compressed sensing [76] and on the threshold saturation of systems characterized by a 1-dimensional state evolution [82, 119] have been obtained by different approaches.

In the classical noisy compressed sensing problem, the AMP algorithm and the SE recursion tracking the algorithmic performance were derived for the AWGN channel [69, 60]. The extension of AMP to general memoryless (possibly non-linear) channels with arbitrary input and output distributions was introduced in [110] via the generalized approximate message-passing (GAMP) algorithm. Moreover, an extension of SE describing the exact behavior of GAMP was also provided in [110]. Later on, a full rigorous analysis proving the tractability of GAMP via SE was given in [115]. These encouraging results naturally led to generalize the analysis of SS codes in [118] to a much broader setting that includes all memoryless channels and potentially any input signal model that factorizes over B-dimensional sections [120, 121, 116].

In this chapter we prove that threshold saturation is a universal phenomenon for SS codes; i.e. we show that, for any memoryless channel, spatial coupling allows GAMP decoding to reach the potential threshold of the code ensemble (Theorem 2.1 and Corollary 2.3). Moreover, we argue that spatially coupled SS codes universally achieve capacity under GAMP decoding by showing that the error floor vanishes and the potential threshold tends to capacity as one of the code's parameters goes to infinity. Indeed, a fully rigorous statement requires to proof that the state evolution tracks the performance of GAMP over general memoryless channels, which is beyond the scope of this work. Furthermore, we give a simple expression of the GAMP algorithmic threshold of the underlying code ensemble in terms of a Fisher information (Section 2.6). Although we focus on coding for the sake of coherence with the related literature, the framework and methods are very general and hold for a wide class of non-linear estimation problems with random linear mixing.

Our proof strategy uses a potential function, which is inspired from the statistical physics replica method. However, we stress that the proof *does not* rely on the replica method (which is not rigorous). Recently, it has been shown that the replica prediction is exact for random linear estimation problems including compressed sensing and SS codes on the AWGN channel [122, 104, 103, 123]. Hence, the potential threshold can be rigorously interpreted as the optimal threshold under MMSE decoding.

This chapter is organized as follows. The code construction of the underlying and coupled ensembles are described in Section 2.2. Section 2.3 reviews the GAMP algorithm, while Section 2.4 presents the SE equations and potential function adapted to the present context. The GAMP thresholds of the under-

lying and coupled ensembles as well as the potential threshold are then given precise definitions. The essential steps for the proof of threshold saturation are presented in Section 2.5. The connection between the potential threshold at infinite input alphabet size and Shannon's capacity, as well as the closed form expression of the algorithmic threshold in terms of a Fisher information, are given in Section 2.6. Four different channel models are then used to illustrate the results. Section 2.7 is dedicated to open challenges.

2.2 Code Ensembles

We first define the underlying and spatially coupled ensembles of SS codes for transmission over a generic memoryless channel. In the rest of this chapter a subscript "un" indicates a quantity related to the underlying ensemble and a subscript "co" a quantity related to the spatially coupled ensemble. The probability law of a Gaussian random variable X with mean m and variance σ^2 is denoted $X \sim \mathcal{N}(m, \sigma^2)$ and the corresponding probability distribution function as $\mathcal{N}(x|m, \sigma^2)$.

2.2.1 The Underlying Ensemble

In the framework of SS codes, the *information word* or *message* is a vector made of L *sections*, $\mathbf{s} = [\mathbf{s}_1, \dots, \mathbf{s}_L]$. Each section $\mathbf{s}_l$, $l \in \{1, \dots, L\}$, is a B-dimensional vector with a single component equal to 1 and $B-1$ components equal to 0. The non-zero component of each section can be set differently, especially when schemes with power allocation are considered [33, 34]. However, we will restrict ourselves to the binary case in this chapter where spatial coupling is used to achieve capacity instead of power allocation. We call B the *section size* (or alphabet size usually chosen to be a power of 2) and set $N = LB$. The message $\mathbf{s}$ can be seen as a one-to-one mapping from an original message $\mathbf{u} \in \{0,1\}^{L \log_2(B)}$, where the position of the non-zero component in $\mathbf{s}_l$ is specified by the binary representation of $\mathbf{u}_l$ (i.e. $\mathbf{s}$ is obtained from $\mathbf{u}$ using a simple position modulation (PM) scheme). For example if $B = 4$ and $L = 5$, a valid message is $\mathbf{s} = [0001, 0010, 1000, 0100, 0010]$ which corresponds to $\mathbf{u} = [00, 01, 11, 10, 01]$. One can think of the information words as being defined for a B-ary alphabet with a constant power allocation for each symbol.

We consider random codes generated by a fixed *coding matrix* $\mathbf{F} \in \mathbb{R}^{M \times N}$ drawn from the ensemble of random matrices with i.i.d real Gaussian entries distributed as $\mathcal{N}(0, 1/L)$. The variance of the coding matrix entries is such that the *codeword* $\mathbf{Fs} \in \mathbb{R}^M$ has a normalized average power $\mathbb{E}[||\mathbf{Fs}||_2^2]/M = 1$. Note that the cardinality of this code is B^L and the length of the codeword is M. Hence, the (design) rate is defined as

$$R = \frac{L \log_2 B}{M} = \frac{N \log_2 B}{MB}. \tag{2.1}$$

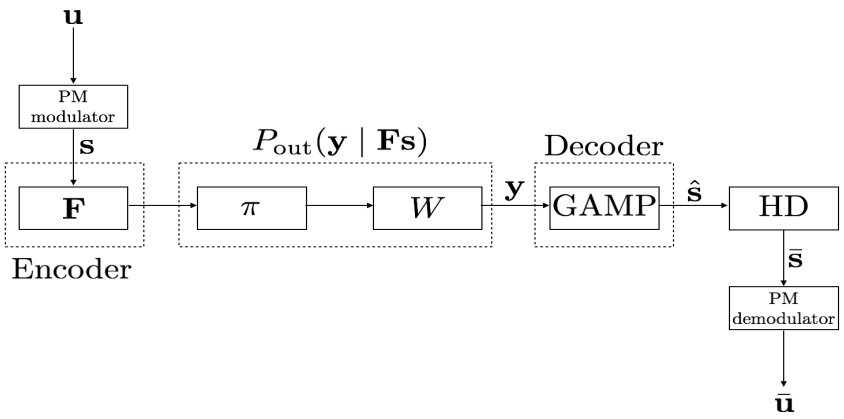

Figure 2.1: The encoder/decoder block diagram of the SS codes under GAMP decoding over any memoryless channel W. The map π is needed when the capacity achieving input distribution of W is not Gaussian. The GAMP algorithm provides soft valued estimate $\hat{\mathbf{s}}$ of $\mathbf{s}$ in the MMSE sense. A simple hard decision (HD) mechanism is used to provide the binary decoded message $\bar{\mathbf{s}}$ by setting the most biased component in each section of $\hat{\mathbf{s}}$ to 1 and the others to 0. The original message $\mathbf{u}$ and its decoded version $\bar{\mathbf{u}}$ can be easily recovered from $\mathbf{s}$ and $\bar{\mathbf{s}}$ respectively using PM modulator and demodulator as illustrated in Section 2.2.1.

The code is thus specified by (M, R, B) where R is the code rate, M the block length, B the section size.

Codewords are transmitted through a known memoryless channel W. This requires to map the codeword components $[\mathbf{Fs}]_\mu \in \mathbb{R}$, $\mu \in \{1, \ldots, M\}$, onto the input alphabet of W. We call π this map and refer to Section 2.6 for various examples. The concatenation of π and W can be seen as an *effective memoryless channel* P_{out}, such that

$$P_{\text{out}}(\mathbf{y}|\mathbf{Fs}) = \prod_{\mu=1}^{M} P_{\text{out}}(y_\mu|[\mathbf{Fs}]_\mu) := \prod_{\mu=1}^{M} W(y_\mu|\pi([\mathbf{Fs}]_\mu)). \tag{2.2}$$

Note that one can look equivalently at π as a part of the channel model or as a part of the encoder. In the present framework, it is more convenient to work with the effective memoryless channel from which the receiver obtains the noisy channel observation $\mathbf{y}$. However in the analysis of Section 2.6, the capacity of W is considered.

The decoding task is to recover $\mathbf{s}$ from channel observations $\mathbf{y}$ as depicted in Fig. 2.1. The decoding can be interpreted as a compressed sensing problem with structured sparsity—due to the sectionwise structure of $\mathbf{s}$—where $\mathbf{y}$ would be the compressed measurements. The rate R can be linked to the "measurement rate" α, used in the compressed sensing literature, by

$$\alpha = \frac{M}{N} = \frac{\log_2 B}{BR}. \tag{2.3}$$

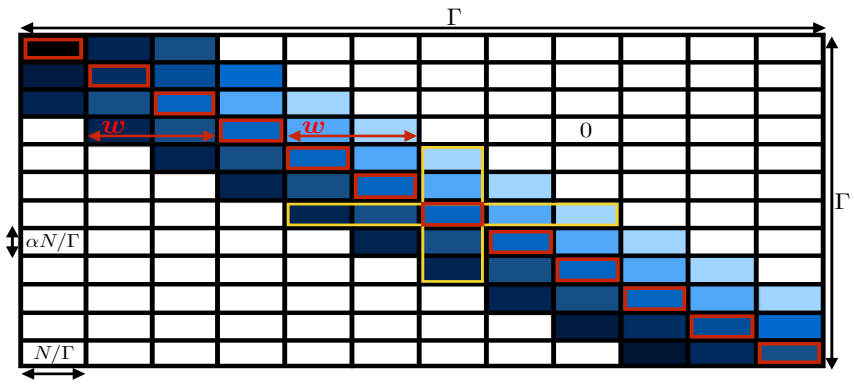

Figure 2.2: A spatially coupled coding matrix $\mathbf{F}^{\text{co}} \in \mathbb{R}^{M\times N}$ made of $\Gamma \times \Gamma$ blocks indexed by (r, c), each with N/Γ columns and $M/\Gamma = \alpha N/\Gamma$ rows where $\alpha = (\log_2 B)/BR$. The i.i.d elements in block (r, c) are distributed as $\mathcal{N}(0, J_{r,c}\Gamma/L)$. Away from the boundaries, in addition to the diagonal (in red), there are w forward and w backward coupling blocks. In this example, the design function g_w enforces a stronger backward coupling where the non-uniform variance across blocks is illustrated by the level of shading. Blocks are darker at the boundaries because the variances are larger so as to enforce the *variance normalization* $\sum_{c=1}^{\Gamma} J_{r,c} = 1 \; \forall \; r$. The yellow shape emphasizes *variance symmetry*.

Thus, the same algorithms and analysis used in compressed sensing theory like the GAMP algorithm and SE can be used in the present context. See [78] for more details on this interconnection.

2.2.2 The Spatially Coupled Ensemble

We consider spatially coupled codes based on coding matrices $\mathbf{F}^{\text{co}} \in \mathbb{R}^{M\times N}$ as depicted in Fig. 2.2. A spatially coupled coding matrix $\mathbf{F}^{\text{co}}$ is made of $\Gamma \times \Gamma$ *blocks* indexed by (r, c), each with N/Γ columns and $M/\Gamma = \alpha N/\Gamma$ rows. The structure of $\mathbf{F}^{\text{co}}$ induces a natural decomposition of the message into Γ blocks, $\mathbf{s} = [\mathbf{s}_1, \ldots, \mathbf{s}_\Gamma]$, where each block is made of L/Γ sections.[2] $\mathbf{F}^{\text{co}}$ is constructed such that each block is coupled (except at the boundaries) with w forward blocks and w backward blocks, where w is the *coupling window*. The strength of the coupling is specified by the variance $J_{r,c}$ of each block (r, c). The entries inside each block (r, c) of $\mathbf{F}^{\text{co}}$ are i.i.d. distributed as $\mathcal{N}(0, J_{r,c}\Gamma/L)$.[3] In order to impose homogeneous power over all the components of $\mathbf{F}^{\text{co}}\mathbf{s}$, we tune the (unscaled) block *variances* $J_{r,c}$ such that the following *variance normalization*

[2] Of course N, M, L, Γ can always be chosen s.t $N/\Gamma, M/\Gamma, L/\Gamma$ are integers.

[3] In the original uncoupled construction the variance scales as the inverse number of sections. In the coupled construction the variances within a block scales as the inverse number of sections within a block.

condition holds for all $r \in \{1, \dots, \Gamma\}$

$$\sum_{c=1}^{\Gamma} J_{r,c} = 1. \tag{2.4}$$

This normalization induces homogeneous average power over all codeword components, i.e. $||\mathbf{F}^{\text{co}}\mathbf{s}||_2^2/M = 1$. There are various ways to construct the variance matrix J of the spatially coupled matrix such that (2.4) holds. For instance, one can pick $J_{r,c}$'s such that the coupling strength is uniform over the window. However, we will consider a more general construction in this chapter by using a *design function* g_w. The design function satisfies

$$\begin{cases} g_w(x) = 0 & \text{if} \quad |x| > 1 \\ \underline{g} \leq g_w(x) \leq \bar{g} & \text{if} \quad |x| \leq 1, \end{cases} \tag{2.5}$$

where $\bar{g}$, $\underline{g}$ are strictly positive constants independent of w. Moreover, g_w is assumed to be Lipschitz continuous on $|x| < 1$ with Lipschitz constant g_* independent of w. In particular

$$\left|g_w\left(\frac{k}{w}\right) - g_w\left(\frac{k'}{w}\right)\right| \leq \frac{g_*}{w}|k - k'|, \tag{2.6}$$

for $k, k' \in \{-w, \dots, w\}$. Furtheremore, we impose the following normalization

$$\frac{1}{2w+1} \sum_{k=-w}^{w} g_w\left(\frac{k}{w}\right) = 1. \tag{2.7}$$

The design function is then used to construct the variances such that (2.4) and (2.7) are satisfied. Hence, we choose

$$J_{r,c} = \gamma_r \frac{g_w((c-r)/w)}{2w+1} = \frac{g_w((c-r)/w)/(2w+1)}{\sum_{c=1}^{\Gamma} g_w((c-r)/w)/(2w+1)}, \tag{2.8}$$

where γ_r is tuned to enforce (2.4). Note that, away from the boundaries, γ_r is a trivial term equal to 1. However, γ_r changes at the boundaries to compensate for the lower number of blocks being coupled (see Fig. 2.2 where darker colors were used at the boundaries to stress on this point). The following remarks will be used in the analysis. We always have $1 \leq \gamma_r \leq \underline{g}^{-1}$ and

$$J_{r,c} \leq (\bar{g}/\underline{g})(2w+1)^{-1}. \tag{2.9}$$

In the bulk (i.e. away from the boundaries), the following *variance symmetry* property holds for $k \in \{2w+1, \dots, \Gamma - 2w\}$

$$\sum_{r=1}^{\Gamma} J_{r,k} = \sum_{c=1}^{\Gamma} J_{k,c} = 1. \tag{2.10}$$

The ensemble of spatially coupled matrices is then parametrized by the parameters $(M, R, B, \Gamma, w, g_w)$. Note that the coupling induced by g_w is not necesssarily symmetric, hence the present construction generalizes the ones in [81, 82, 119] which all require $g_w(-x) = g_w(x)$, while we do not. This relaxation may strongly improve the perfomances in practice [111].

One key element of spatially coupled codes is the *seed* introduced at the boundaries. We assume the sections in the first $4w$ and last $4w$ blocks of the message $\mathbf{s}$ to be known by the decoder (the choice of $4w$ blocks is convenient for the proofs and will become clear in Section 2.5). This boundary condition can be interpreted as perfect side information that propagates inwards and boosts the performance. Note that one could also impose the seed differently by constructing a coding matrix with lower communication rate (higher measurement rate) at the boundaries [77, 78, 117, 111, 115]. The seed induces a rate loss in the *effective rate* of the code

$$R_{\text{eff}} = R\Big(1 - \frac{8w}{\Gamma}\Big). \tag{2.11}$$

However, this loss vanishes as $L \to \infty$ and then $\Gamma \to \infty$ for any fixed R. As already mentioned, in addition to lower decoding error, the main advantage of coupled SS codes w.r.t power allocated ones is that they allow communication at high rate with a small section size B, while power allocated codes require a much larger B, which prevents communication of messages of practically relevant sizes [78].

2.3 Generalized Approximate Message-Passing Algorithm

The posterior distribution describing the statistical relationships in the decoding task is given by (in the following discussion $\mathbf{F}$ denotes a generic coding matrix)

$$P(\mathbf{s}|\mathbf{y}, \mathbf{F}) = \frac{\prod_{l=1}^{L} p_0(\mathbf{s}_l) \prod_{\mu=1}^{M} P_{\text{out}}(y_\mu|[\mathbf{Fs}]_\mu)}{\int d\mathbf{s} \prod_{l=1}^{L} p_0(\mathbf{s}_l) \prod_{\mu=1}^{M} P_{\text{out}}(y_\mu|[\mathbf{Fs}]_\mu)}. \tag{2.12}$$

In the SS codes setting, the sections of the information word are uniformly distributed over all the possible B-dimensional vectors with a single non-zero component equal to 1. Hence, the prior of each section reads

$$p_0(\mathbf{s}_l) = \frac{1}{B} \sum_{i=1}^{B} \delta_{s_{li},1} \prod_{j \neq i}^{B-1} \delta_{s_{lj},0}, \tag{2.13}$$

where s_{li} is the i^{th} component of the l^{th} section (here $i \in \{1, \dots, B\}$ and $l \in \{1, \dots, L\}$). The posterior distribution (2.12) can be represented via a graphical model as shown in the l.h.s of Fig. 2.3. Therefore, it is natural to consider an iterative message-passing algorithm to perform the decoding. For a

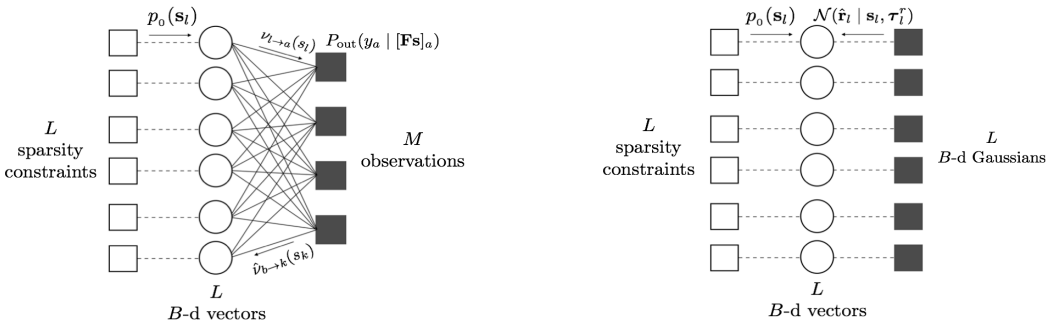

Figure 2.3: Left: Factor graph of the underlying ensemble showing the statistical relationships between the B-dimensional sections (circles) of the information word **s** given the known prior $p_0(\mathbf{s})$ (plain squares), the coding matrix **F** and the channel observation **y** (colored squares). The BP algorithm estimates **s** via iterative exchange of messages, along edges, between circle-nodes and square-nodes. Right: The GAMP algorithm simplifies the BP operations to a sequence of estimation problems from Gaussian noise. At the l^{th} section, $\hat{\mathbf{r}}_l$ is the output of an effective Gaussian channel of zero mean and covariance matrix $\text{diag}(\boldsymbol{\tau}_l^{\text{r}})$.

dense graphical model, Belief Propagation (BP) is computationally prohibitive but can be simplified down to the AMP algorithm which has been successfully used in many applications, mainly in compressed sensing [69, 60]. The AMP algorithm uses efficient Gaussian (or quadratic) approximations of BP that "decouple" the vector-valued estimation problem into a sequence of scalar estimation problems under an *effective Gaussian noise* (r.h.s of Fig. 2.3). The sum-product version of AMP (originally used to perform MMSE estimation in compressed sensing with AWGN channel) was adapted in [36, 78] to SS codes by incorporating the structured B-dimensional prior distribution (2.13). The GAMP algorithm extends the approximations made in AMP to any memoryless channel [110]. Interestingly, the same Gaussian approximations on a dense graph remain valid under GAMP, even for a non-Gaussian channel, and the only difference appears in the computation of the effective Gaussian noise parameters.

The GAMP algorithm was originally introduced to estimate signals with i.i.d components [110]. In the present context the message components are correlated through $p_0(\mathbf{s}_l)$, therefore we adapt GAMP to cover this vectorial setting. The steps of GAMP are shown in Algorithm 2.1 below. The "$\circ 2$" and "$\circ -1$" symbols mean that the square and inverse operations are taken componentwise: $(\mathbf{F}^{\circ 2})_{\mu i} = F_{\mu i}^2$ and $(\mathbf{F}^{\circ -1})_{\mu i} = F_{\mu i}^{-1}$. All the derivatives in Algorithm 2.1 are also taken componentwise. The function g_{in} depends on the input prior distibution and it is adapted from [110] to act on B-dimensional vectors. Due to the code construction, $g_{\text{in}}(\hat{\mathbf{r}}_l, \text{diag}(\boldsymbol{\tau}_l^{\text{r}}))$ can be interpreted as the MMSE estimator, or *denoiser*, of a given B-dimensional section $\mathbf{s}_l$ sent through an effective Gaussian channel of zero mean and covariance matrix

$\mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}})$ where

$$\hat{\mathbf{r}}_l = \mathbf{s}_l + \boldsymbol{\xi}, \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}})). \tag{2.14}$$

Definition 2.1 (Denoiser)**.** *Formally, we define the denoiser acting* section-wise *on each B-dimensional section of the message as follows*

$$g_{\mathrm{in}}(\hat{\boldsymbol{r}}_l, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}})) := \mathbb{E}[\boldsymbol{S}_{\mathrm{l}} \mid \hat{\boldsymbol{R}}_{\mathrm{l}} = \hat{\boldsymbol{r}}_{\mathrm{l}}] = \frac{\int \mathrm{d}\boldsymbol{s}_{\mathrm{l}}\, \mathrm{p}_0(\boldsymbol{s}_{\mathrm{l}}) \mathcal{N}(\hat{\boldsymbol{r}}_{\mathrm{l}} | \boldsymbol{s}_{\mathrm{l}}, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}}))\boldsymbol{s}_{\mathrm{l}}}{\int \mathrm{d}\boldsymbol{s}_{\mathrm{l}}\, \mathrm{p}_0(\boldsymbol{s}_{\mathrm{l}}) \mathcal{N}(\hat{\boldsymbol{r}}_{\mathrm{l}} | \boldsymbol{s}_{\mathrm{l}}, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}}))}, \tag{2.15}$$

where $\boldsymbol{S}_l \sim p_0(\boldsymbol{s}_l)$. Plugging (2.13) *yields the componentwise expression of the denoiser used in the GAMP algorithm for SS codes*

$$\begin{aligned}[g_{\mathrm{in}}(\hat{\boldsymbol{r}}_l, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}}))]_{\mathrm{i}} &= \frac{\exp((2\hat{r}_{li} - 1)/(2\tau_{li}^r))}{\sum_{j=1}^{B} \exp((2\hat{r}_{lj} - 1)/(2\tau_{lj}^r))} \\ &= \Big[1 + \sum_{j \neq i}^{B} \exp\Big((2\hat{r}_{lj} - 1)/(2\tau_{lj}^r) - (2\hat{r}_{li} - 1)/(2\tau_{li}^r)\Big)\Big]^{-1},\end{aligned}$$

where $i \in \{1, \ldots, B\}$.

Moreover, the componentwise product $\boldsymbol{\tau}_l^r \circ \frac{\partial}{\partial \hat{\mathbf{r}}_l} g_{\mathrm{in}}$, that appears in Algorithm 2.1, is the estimate of the posterior variance which quantifies how "confident" GAMP is in its current iteration. This is given by

$$\begin{aligned}\boldsymbol{\tau}_l^r \circ \frac{\partial}{\partial \hat{\mathbf{r}}_l} g_{\mathrm{in}}(\hat{\mathbf{r}}_l, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}})) &:= \mathrm{var}(\mathbf{S}_l \mid \hat{\mathbf{R}}_l = \hat{\mathbf{r}}_l) \\ &= \mathbb{E}[\mathbf{S}_l^{\circ 2} \mid \hat{\mathbf{R}}_l = \hat{\mathbf{r}}_l] - (\mathbb{E}[\mathbf{S}_l \mid \hat{\mathbf{R}}_l = \hat{\mathbf{r}}_l])^{\circ 2},\end{aligned} \tag{2.16}$$

where the expectation and the variance are induced from (2.14). As the message $\mathbf{s}$ in SS codes consists of only 0's and 1's, we have that $\mathbb{E}[\mathbf{S}_l^{\circ 2} \mid \hat{\mathbf{R}}_l = \hat{\mathbf{r}}_l] = \mathbb{E}[\mathbf{S}_l \mid \hat{\mathbf{R}}_l = \hat{\mathbf{r}}_l]$. Hence, the calculation of $\mathrm{var}(\mathbf{S}_l \mid \hat{\mathbf{R}}_l = \hat{\mathbf{r}}_l)$ is immediate using (2.13) which yields the following componentwise expression

$$[\boldsymbol{\tau}_l^r \circ \frac{\partial}{\partial \hat{\mathbf{r}}_l} g_{\mathrm{in}}(\hat{\mathbf{r}}_l, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}}))]_{\mathrm{i}} = [\mathrm{g}_{\mathrm{in}}(\hat{\mathbf{r}}_{\mathrm{l}}, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}}))]_{\mathrm{i}} - ([\mathrm{g}_{\mathrm{in}}(\hat{\mathbf{r}}_{\mathrm{l}}, \mathrm{diag}(\boldsymbol{\tau}_{\mathrm{l}}^{\mathrm{r}}))]_{\mathrm{i}})^2.$$

The function g_{out} (in Algorithm 2.1 below) is acting componentwise and depends solely on the physical channel P_{out}. The general expression of g_{out} is given in Appendix 2.8.1 as well as examples for different communication channels.

The computational complexity of GAMP is dominated by the $\mathcal{O}(MN) = \mathcal{O}(L^2 B \ln(B))$ matrix-vector multiplication. It can be reduced, for practical implementations, by using structured operators such as Fourier and Hadamard matrices [77, 121]. Fast Hadamard-based operators constructed as in [77], with random sub-sampled modes of the full Hadamard operator, allow to achieve a lower $\mathcal{O}(L \ln(B) \ln(BL))$ decoding complexity and strongly reduce the memory need [78, 124]. Besides practical advantages, using structured operators

Algorithm 2.1 GAMP $(\mathbf{y}, \mathbf{F}, B, \text{nIter})$

$\hat{\mathbf{s}}^{(0)} \leftarrow \mathbf{0}_{N,1}$
$\boldsymbol{\tau}^{s(0)} \leftarrow (1/B)\mathbf{1}_{N,1}$
$\hat{\mathbf{z}}^{(-1)} \leftarrow \mathbf{0}_{M,1}$
$t \leftarrow 0$
while $t <$ nIter **do**
  $\boldsymbol{\tau}^{p(t)} \leftarrow \mathbf{F}^{\circ 2}\boldsymbol{\tau}^{s(t)}$
  $\hat{\mathbf{p}}^{(t)} \leftarrow \mathbf{F}\hat{\mathbf{s}}^{(t)} - \boldsymbol{\tau}^{p(t)} \circ \hat{\mathbf{z}}^{(t-1)}$
  $\hat{\mathbf{z}}^{(t)} \leftarrow g_{\text{out}}(\hat{\mathbf{p}}^{(t)}, \mathbf{y}, \boldsymbol{\tau}^{p(t)})$
  $\boldsymbol{\tau}^{z(t)} \leftarrow -\frac{\partial}{\partial \hat{\mathbf{p}}^{(t)}} g_{\text{out}}(\hat{\mathbf{p}}^{(t)}, \mathbf{y}, \boldsymbol{\tau}^{p(t)})$
  $\boldsymbol{\tau}^{r(t)} \leftarrow \left(((\boldsymbol{\tau}^{z(t)})^{\intercal}\mathbf{F}^{\circ 2})^{\intercal}\right)^{\circ -1}$
  $\hat{\mathbf{r}}^{(t)} \leftarrow \hat{\mathbf{s}}^{(t)} + \boldsymbol{\tau}^{r(t)} \circ ((\hat{\mathbf{z}}^{(t)})^{\intercal}\mathbf{F})^{\intercal}$
  $\hat{\mathbf{s}}^{(t+1)} \leftarrow g_{\text{in}}(\hat{\mathbf{r}}^{(t)}, \text{diag}(\boldsymbol{\tau}^{\text{r(t)}}))$
  $\boldsymbol{\tau}^{s(t+1)} \leftarrow \boldsymbol{\tau}^{r(t)} \circ \frac{\partial}{\partial \hat{\mathbf{r}}^{(t)}} g_{\text{in}}(\hat{\mathbf{r}}^{(t)}, \text{diag}(\boldsymbol{\tau}^{\text{r(t)}}))$
  $t \leftarrow t + 1$

can lead to a more robust finite-length performance [77]. However, random operators are mathematically more tractable and easier to analyse. Hence, we restrict ourselves in this chapter to random operators.

Decoding SS codes using an iterative message-passing algorithm, such as GAMP, leads asymptotically in L to a sharp *phase transition* below Shannon's capacity. The decoder is therefore blocked at a certain threshold separating the "decodable" and "non-decodable" regions. Moreover, SS codes under message-passing decoding may exhibit, asymptotically in L and for any fixed alphabet size B, a non-negligible *error floor*[4] in the decodable region (similarly to low-density generator-matrix codes [27]). Whenever the error floor exists, it can be made arbitrarily small by increasing B [77, 78].

2.4 State Evolution and Potential Formulation

The asymptotic behavior of the AMP algorithm operating on dense graphs can be tracked by a simple recursion called state evolution (SE), similar to the density evolution (DE) for sparse graphs. The rigorous proof showing that SE tracks exactly the asymptotic performance of AMP and GAMP was given in [69, 115]. Moreover, the extension of the SE equation of AMP to SS code settings, with B-dimensional structured prior distribution and power allocation, was proven to be exact in [37]. We believe that the methods of [37] and [115] can be extended to the present setting of spatially coupled SS codes and GAMP algorithm. This would prove that SE correctly tracks GAMP, a conjecture which is firmly supported by numerical simulations [121].

[4] In fact, the existence of error floor depends on the communication channel being used. for example there is no error floor for the BEC and BSC but there is one for the AWGN channel.

2.4.1 State Evolution of the Underlying System

SE tracks the performance of GAMP by computing the average asymptotic mean-square error (MSE) of the GAMP estimate $\hat{\mathbf{s}}^{(t)}$ at each iteration t

$$\tilde{E}^{(t)} := \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} \|\hat{\mathbf{s}}_l^{(t)} - \mathbf{s}_l\|_2^2. \tag{2.17}$$

It turns out that tracking the GAMP algorithm is equivalent to running a simple recursion that iteratively computes the MMSE of a single section sent through an *equivalent AWGN channel*. This equivalent channel is induced by the code construction and has an *effective* noise variance that depends solely on the physical channel $P_{\text{out}}(y|x)$. In order to formalize this, we first need some definitions.

Definition 2.2 (Effective noise)**.** *The effective noise variance $\Sigma^2(E)$, parametrized by $E \in [0,1]$, is defined via the following relation*

$$\Sigma^{-2}(E) := \frac{\mathbb{E}_{p|E}[\mathcal{F}(p|E)]}{R},$$

where the expectation $\mathbb{E}_{p|E}$ is w.r.t $\mathcal{N}(p|0, 1-E)$ and

$$\mathcal{F}(p|E) := \int dy f(y|p,E)(\partial_p \ln f(y|p,E))^2$$

is the Fisher information of the parameter p associated with the probability distribution of the random variable Y with density

$$f(y|p,E) := \int dx P_{\text{out}}(y|x)\mathcal{N}(x|p,E).$$

See Appendix 2.8.1 for explicit expressions for various communication channels.

We will need some regularity properties for the function $\Sigma(E)$ which boils down to mild assumptions on the channel transition probability $P_{\text{out}}(y|x)$.

Assumption 2.1 (Continuity and boundedness of $\Sigma(E)$)**.** *The channel transition probability $P_{\text{out}}(y|x)$ is such that $\Sigma(E)$ is a continuous and twice differentiable function of $E \in [0,1]$.*

Assumption 2.2 (Scaling of $\Sigma^{-2}(E)$ as $E \to 0$)**.** *The channel transition probability $P_{\text{out}}(y|x)$ is such that $\Sigma^{-2}(E)$ and its first two derivatives are bounded by a polynomial in E^{-1}. Formally, for a given channel there exist two constants $C > 0$ and $\beta > 0$ such that*

$$\max\left(\Sigma^{-2}(E), \left|\frac{\partial \Sigma^{-2}(E)}{\partial E}\right|, \left|\frac{\partial^2 \Sigma^{-2}(E)}{\partial E^2}\right|\right) \leq \frac{C}{RE^{\beta}} \equiv \lambda(E) \tag{2.18}$$

for all $E \in [0,1]$.

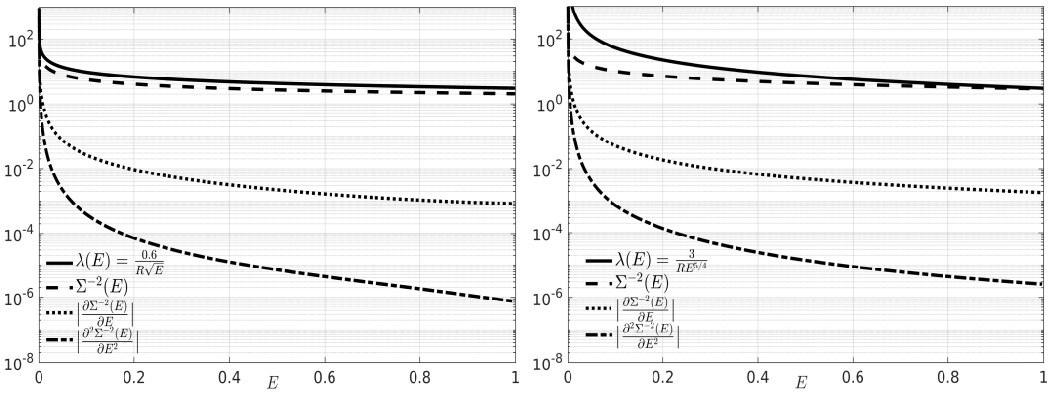

Figure 2.4: $\Sigma^{-2}(E)$ and its first two derivatives in a semi-log scale for the BSC (left) and the BEC (right) with flip and erasure probabilities $\epsilon = 0.1$ and $R = 0.2$. Assumption 2.2 is satisfied with exponents $\beta = 1/2$ and $5/4$. Furthermore, the effective noise variance of both channels is bounded with $\Sigma^2(E) < 1/2$. Note that the mapping $\pi([\mathbf{Fs}]_\mu) = \mathrm{sign}([\mathbf{Fs}]_\mu)$ was used here.

These assumptions will be needed in the proof of threshold saturation in Section 2.5. In practice they can be checked on a case by case basis for each channel at hand. For the AWGN channel, we have the analytic simple expression $\Sigma^2(E) = (\mathrm{snr}^{-1} + E)R$ so the assumptions are obviously satisfied. One can also check them for the binary symmetric channel (BSC), binary erasure channel (BEC) and Z channel (ZC), using the tedious expressions for the Fisher information given in Table 2.1 in Appendix 2.8.1. Fig. 2.4 illustrates $\Sigma^{-2}(E)$ and its derivatives for the BSC and BEC.

The following lemma (which is independent from the assumptions) will also be needed.

Lemma 2.1. *$\Sigma^2(E)$ is non-negative and increasing with E. In particular $\Sigma^2(E) \leq \Sigma^2(1) < +\infty$.*

Proof. Positivity of the Fisher information implies $\Sigma^2(E) \geq 0$. The proof that it is increasing is a straightforward application of the data processing inequality for Fisher information (e.g. Corollary 6 in [125]). □

From now on, $\mathbf{S} \sim p_0(\mathbf{s})$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_B)$ are B-dimensional random vectors with corresponding expectations denoted $\mathbb{E}_{\mathbf{S},\mathbf{Z}}$, and $Z \sim \mathcal{N}(0,1)$ with expectation denoted $\mathbb{E}_Z$.

Definition 2.3 (SE of the underlying system)**.** *The SE operator of the underlying system is the average MMSE of the equivalent channel*

$$\begin{aligned} T_{\mathrm{un}}(E) &:= \mathrm{mmse}\big(\Sigma(E)\big) \\ &= \mathbb{E}_{\boldsymbol{S},\boldsymbol{Z}}\Bigg[\sum_{i=1}^{B}\Big(\Big[g_{\mathrm{in}}\Big(\boldsymbol{S} + \frac{\mathbf{Z}\Sigma(E)}{\sqrt{\log_2 B}}, \frac{\boldsymbol{I}_B\,\Sigma^2(E)}{\log_2 B}\Big)\Big]_i - S_i\Big)^2\Bigg], \end{aligned}$$

where g_{in} is the denoiser given in Definition 2.1

$$\Big[g_{\text{in}}\Big(\boldsymbol{s}+\frac{\boldsymbol{z}\Sigma}{\sqrt{\log_2 B}},\frac{\boldsymbol{I}_B\,\Sigma^2}{\log_2 B}\Big)\Big]_i=\left[1+\sum_{k\neq i}^{B}e^{(s_k-s_i)\log_2 B/\Sigma^2+(z_k-z_i)\sqrt{\log_2 B}/\Sigma}\right]^{-1}. \tag{2.19}$$

The SE iteration tracking the performance of the GAMP decoder for the underlying system can be expressed as

$$\tilde{E}^{(t+1)}=T_{\text{un}}(\tilde{E}^{(t)}),\qquad t\geq 0,$$

with the initialization $\tilde{E}^{(0)}=1$.

Note for further use that (2.19) is a well defined continuous function of $\Sigma>0$ (all other arguments being fixed). At $\Sigma=0$ we define the function by its continuous extension which is obviously finite. Thus we will consider that g_{in} is continuous for $\Sigma\geq 0$.

After t iterations of the GAMP algorithm, the MSE tracked by SE is denoted by $T_{\text{un}}^{(t)}(\tilde{E}^{(0)})$. The monotonicity properties and the continuity of the SE operator, discussed in Section 2.5.1, ensure that eventually all initial conditions converge to a fixed point. More specifically, the following limit exists

$$\lim_{t\to\infty}T_{\text{un}}^{(t)}(\tilde{E}^{(0)}):=T_{\text{un}}^{(\infty)}(\tilde{E}^{(0)}), \tag{2.20}$$

for all $\tilde{E}^{(0)}\in[0,1]$ and satisfies

$$T_{\text{un}}(T_{\text{un}}^{(\infty)}(\tilde{E}^{(0)}))=T_{\text{un}}^{(\infty)}(\tilde{E}^{(0)}). \tag{2.21}$$

Having introduced the SE iteration, the following definitions can be properly stated.

Definition 2.4 (MSE Floor)**.** *The MSE floor E_{f} is the fixed point reached from the initial condition of zero error,*

$$E_{\text{f}}=T_{\text{un}}^{(\infty)}(0).$$

Note that for the channels where $E=0$ is not a trivial fixed point of the SE at a finite section size B, the MSE floor E_{f} is strictly positive. For example, this is the case for the AWGN channel [36, 78]. However, one can show that for certain channels W there exists a trivial fixed point $E=0$ of SE leading to vanishing MSE floor even at finite B. This is typically the case for binary input channels and has been proved explicitly for the BEC, BSC and Z channels [121]. For generality, we will always denote the MSE floor as E_{f} whether it is zero or not.

Definition 2.5 (Basin of attraction)**.** *The basin of attraction $\mathcal{V}_0$ to the MSE floor E_{f} is defined as*

$$\mathcal{V}_0:=\big\{E\in[0,1]\,|\,T_{\text{un}}^{(\infty)}(E)=E_{\text{f}}\big\}.$$

Definition 2.6 (Threshold of underlying ensemble)**.** *The GAMP threshold of the underlying ensemble is defined as*

$$R_{\text{un}} := \sup\{R > 0 \,|\, T_{\text{un}}^{(\infty)}(1) = E_{\text{f}}\}.$$

For the present system, one can show that the only two possible fixed points are $T_{\text{un}}^{(\infty)}(0)$ and $T_{\text{un}}^{(\infty)}(1)$. For $R < R_{\text{un}}$, there is only one fixed point, namely the "good" one $T_{\text{un}}^{(\infty)}(0) = E_{\text{f}}$. Whenever E_{f} is non-zero, it will vanish as the section size B increases (see Section 2.6). Instead if $R > R_{\text{un}}$, the GAMP decoder is blocked by the "bad" fixed point $T_{\text{un}}^{(\infty)}(1) > E_{\text{f}}$. The "bad" fixed point does not vanish as B increases.

The GAMP algorithm "tries" to minimize the MSE. Thus the natural quantity being tracked by SE is the MSE. But one can also assess the performance of GAMP by looking at the section error rate (SER) (which is more natural for coding problems) after applying a hard decision (HD) thresholding on the decoder's output. The analytical relationship between MSE and the SER has been discussed in [36, 78] and one verifies that an MSE going to zero implies a SER going to zero.

2.4.2 State Evolution of the Coupled System

For a spatially coupled system, the performance of GAMP at each iteration t is described by an average *MSE vector* $[\tilde{E}_c^{(t)} \mid c \in \{1, \dots, \Gamma\}]$ along the "spatial dimension" indexed by the blocks of the message with

$$\tilde{E}_c^{(t)} := \lim_{L\to\infty} \frac{\Gamma}{L} \sum_{l \in c} \|\hat{\mathbf{s}}_l^{(t)} - \mathbf{s}_l\|_2^2, \quad c \in \{4w+1, \dots, \Gamma - 4w\}, \tag{2.22}$$

where the sum $l \in c$ is over the set of indices of the L/Γ sections composing the c-th block of $\mathbf{s}$. To reflect the seeding at the boundaries, we enforce the following *pinning condition* for all $c \in \{1, \dots, 4w\} \cup \{\Gamma - 4w + 1, \dots, \Gamma\}$

$$\tilde{E}_c^{(t)} = 0, \qquad t \geq 0, \tag{2.23}$$

where the message at these positions is assumed to be known to the decoder at all times.

It turns out that the following change of variables

$$E_r^{(t)} := \sum_{c=1}^{\Gamma} J_{r,c} \tilde{E}_c^{(t)}, \tag{2.24}$$

where $\mathbf{E} = [E_r \,|\, r \in \{1, \dots, \Gamma\}]$ is called the *profile*, makes the problem mathematically more tractable for spatially coupled codes. The pinning condition implies

$$E_r^{(t)} = 0, \qquad t \geq 0, \tag{2.25}$$

for all $r \in \mathcal{R} := \{1, \dots, 3w\} \cup \{\Gamma - 3w + 1, \dots, \Gamma\}$.

An important concept is that of *degradation* because it allows to compare different profiles.

Definition 2.7 (Degradation)**.** *A profile $\boldsymbol{E}$ is degraded (resp. strictly degraded) with respect to another one $\boldsymbol{G}$, denoted as $\boldsymbol{E} \succeq \boldsymbol{G}$ (resp. $\boldsymbol{E} \succ \boldsymbol{G}$), if $E_r \geq G_r \ \forall \ r$ (resp. there exists some r such that the inequality is strict).*

In order to define the SE of the spatially coupled system, we need first the following definition.

Definition 2.8 (Per-block effective noise)**.** *The per-block effective noise variance $\Sigma_c^2(\boldsymbol{E})$ is defined, for all $c \in \{1, \dots, \Gamma\}$, by*

$$\Sigma_c^{-2}(\boldsymbol{E}) := \sum_{r=1}^{\Gamma} \frac{J_{r,c}}{\Sigma^2(E_r)} = \sum_{r=1}^{\Gamma} \frac{J_{r,c}}{R} \mathbb{E}_{p|E_r}[\mathcal{F}(p|E_r)].$$

Definition 2.9 (SE of the coupled system)**.** *The vector valued coupled SE operator is defined componentwise for $t \geq 0$ as*

$$\begin{aligned} E_r^{(t+1)} &= [T_{\text{co}}(\boldsymbol{E}^{(t)})]_r \\ &= \sum_{c=1}^{\Gamma} J_{r,c} \mathbb{E}_{\boldsymbol{S},\boldsymbol{Z}} \Big[\sum_{i=1}^{B} \Big(g_{\text{in},i}\Big(\boldsymbol{S} + \frac{\mathbf{Z}\Sigma_c(\boldsymbol{E}^{(t)})}{\sqrt{\log_2 B}}, \frac{\Sigma_c^2(\boldsymbol{E}^{(t)})}{\log_2 B} \Big) - S_i \Big)^2 \Big], \end{aligned}$$

for $r \notin \mathcal{R}$. Note that for $r \in \mathcal{R}$, the pinning condition $E_r^{(t)} = 0$ is enforced at all times. SE is initialized with $E_r^{(0)} = 1$ for $r \notin \mathcal{R}$.

Definition 2.10 (Threshold of coupled ensemble)**.** *The GAMP threshold of the spatially coupled system is defined as*

$$R_{\text{co}} := \liminf_{w \to \infty} \liminf_{\Gamma \to \infty} \sup\{R > 0 \,|\, T_{\text{co}}^{(\infty)}(\mathbf{1}) \preceq \boldsymbol{E}_{\text{f}}\}$$

where $\mathbf{1}$ is the all ones vector and $\boldsymbol{E}_{\text{f}} := [E_r = E_{\text{f}} \,|\, r \in \{1, \dots, \Gamma\}]$ is the MSE floor profile *(recall E_{f} in Definition 2.4). The existence of the limit $T_{\text{co}}^{(\infty)}(\mathbf{1})$ is verified in Section 2.5.1. Note that the degradation $\preceq$ holds with equality for the cases where $E_{\text{f}} = 0$.*

Assumption 2.3. *For the noisy compressed sensing problem, the rigorous proof that SE tracks the performance of GAMP, on both the underlying and spatially coupled models, was already done in [115] by generalizing the work of [69]. For the SS codes, we assume that the same results hold. The proof is beyond the scope of this work and would follow the same analysis of [37] to account for the B-dimensional prior of the SS codes. Our assumption is, however, supported by numerical simulations [121].*

2.4.3 Potential Formulation

The fixed point solutions of SE can be reformulated as stationary points of a *potential function*. This potential function can be obtained from the replica method [36] as shown in Appendix 2.8.4 or by directly integrating the SE fixed

point equations with the correct "integrating factor" as done in [82]. Our subsequent analysis does not depend on the means of obtaining the potential function which is here a mere mathematical tool.

Definition 2.11 (Potential function of underlying ensemble)**.** *The potential function of the underlying ensemble is given by*

$$F_{\mathrm{un}}(E) := U_{\mathrm{un}}(E) - S_{\mathrm{un}}(\Sigma(E)),$$

with

$$U_{\mathrm{un}}(E) := -\frac{E}{2\ln(2)\Sigma^2(E)} - \frac{1}{R}\mathbb{E}_Z\Big[\int dy\, \phi(y|Z,E)\log_2\phi(y|Z,E)\Big],$$

$$S_{\mathrm{un}}(\Sigma(E)) := \mathbb{E}_{\boldsymbol{S},\boldsymbol{Z}}\Big[\log_B \int d^B\boldsymbol{x}\, p_0(\boldsymbol{x})\theta(\boldsymbol{x},\boldsymbol{S},\mathbf{Z},\Sigma(E))\Big],$$

where

$$\phi(y|z,E) := \int dx P_{\mathrm{out}}(y|x)\mathcal{N}(x|z\sqrt{1-E},E),$$

$$\theta(\boldsymbol{x},\boldsymbol{s},\boldsymbol{z},\Sigma(E)) := \exp\left(-\frac{\|\boldsymbol{x}-(\boldsymbol{s}+\boldsymbol{z}\Sigma(E)/\sqrt{\log_2 B})\|_2^2}{2\Sigma^2(E)/\log_2 B}\right).$$

Replacing the prior distribution of SS codes (2.13) *in the definition of* S_{un}*, one gets*

$$S_{\mathrm{un}}(\Sigma(E)) := \mathbb{E}_{\boldsymbol{Z}}\Big[\log_B\Big(1+\sum_{i=2}^{B} e_i\big(\boldsymbol{Z},\frac{\Sigma(E)}{\sqrt{\log_2 B}}\big)\Big)\Big],$$

where

$$e_i(\boldsymbol{z},a) := \exp\big(\frac{z_i-z_1}{a}-\frac{1}{a^2}\big).$$

Definition 2.12 (Free energy gap)**.** *The free energy gap is*

$$\Delta F_{\mathrm{un}} := \inf_{E\notin\mathcal{V}_0}(F_{\mathrm{un}}(E)-F_{\mathrm{un}}(E_{\mathrm{f}})),$$

with the convention that the infimum over the empty set is ∞ *(i.e. when* $R<R_{\mathrm{un}}$*).*

Definition 2.13 (Potential threshold)**.** *The potential threshold is defined as*

$$R_{\mathrm{pot}} := \sup\{R>0 \,|\, \Delta F_{\mathrm{un}}>0\}.$$

We give examples of potential functions for the BEC and the AWGN channel in Fig. 2.5 for $B=2$. Because of Lemma 2.2 below, the minimum that is in the basin of attraction of $E=0$ corresponds to the error floor E_{f}. We observe that there is a non-vanishing error floor for the AWGN channel but a vanishing one for the BEC. The latter situation is also the case for the BSC and Z channel.

Similarly to the underlying ensemble, one can define the potential function of the spatially coupled ensemble that is applied on a vector indexed by the spatial dimension.

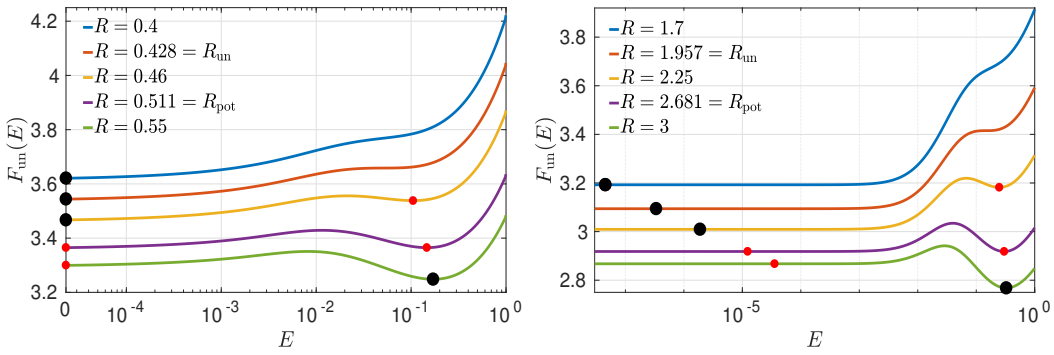

Figure 2.5: The potential functions for the BEC with $\epsilon = 0.1$ (left) and the AWGN channel with $\text{snr} = 100$ (right), in both cases with $B = 2$. The black dots correspond to the global minima while the red dots correspond to the local minima preventing GAMP to decode (e.g.yellow curve). The x-axis is given in the log scale to differentiate between the BEC where there is no error floor and the AWGN channel with non-negligible error floor.

Definition 2.14 (Potential function of spatially coupled ensemble)**.** *The potential function of the spatially coupled ensemble is given by*

$$F_{\text{co}}(\boldsymbol{E}) := U_{\text{co}}(\boldsymbol{E}) - S_{\text{co}}(\boldsymbol{E}) = \sum_{r=1}^{\Gamma} U_{\text{un}}(E_r) - \sum_{c=1}^{\Gamma} S_{\text{un}}(\Sigma_c(\boldsymbol{E})).$$

The following lemma links the potential and SE formulations.

Lemma 2.2. *If* $T_{\text{un}}(\mathring{E}) = \mathring{E}$*, then* $\frac{\partial F_{\text{un}}}{\partial E}|_{\mathring{E}} = 0$*. Similarly for the spatially coupled system, if* $[T_{\text{co}}(\mathring{\boldsymbol{E}})]_r = \mathring{E}_r \ \forall\ r \in \mathcal{R}^c = \{3w+1, \dots, \Gamma - 3w\}$ *then* $\frac{\partial F_{co}}{\partial E_r}|_{\mathring{\boldsymbol{E}}} = 0 \ \forall\ r \in \mathcal{R}^c$.

Proof. See Appendix 2.8.2. □

We end this section by pointing out that the terms composing the potentials have natural interpretations in terms of effective channels. The term $\mathbb{E}_Z[\int dy\, \phi \log_2(\phi)]$ in $U_{\text{un}}(E)$ is minus the conditional entropy $H(Y|Z)$ for the concatenation of the channels $\mathcal{N}(x|z\sqrt{1-E}, E)$ and $P_{\text{out}}(y|x)$ with a standardised input $Z \sim \mathcal{N}(0,1)$. The term $S_{\text{un}}(\Sigma(E))$ is equal to minus the mutual information $I(\mathbf{S};\mathbf{Y})/\log_2 B$ for the Gaussian channel $\mathcal{N}(\mathbf{y}|\mathbf{s}, \mathbf{I}_B\, \Sigma^2(E)/\log_2 B)$ and input distribution $p_0(\mathbf{s})$, up to a constant factor $-(2\ln 2)^{-1}$.

2.5 Threshold Saturation

We now prove threshold saturation for spatially coupled SS codes using methods from [82]. The main strategy is to assume a "bad" fixed point solution of the spatially coupled SE and to calculate the change in potential due to a small *shift* in two different ways: *i*) by second order Taylor expansion (Lemma

2.4 and Lemma 2.6), *ii*) by direct evaluation (Lemma 2.8). We then show by contradiction that as long as $R < R_{\text{pot}}$, the SE converges to the "good" fixed point (Theorem 2.1).

In Section 2.5.1 we start by showing some essential properties of the spatially coupled SE operator.

2.5.1 Properties of the Coupled System

Monotonicity properties of the SE operators T_{un} and T_{co} are key elements in the analysis.

Lemma 2.3. *The SE operator of the coupled system maintains degradation in space, i.e. if* $\boldsymbol{E} \succeq \boldsymbol{G}$*, then* $T_{\text{co}}(\boldsymbol{E}) \succeq T_{\text{co}}(\boldsymbol{G})$*. This property is verified for* T_{un} *for a scalar error as well.*

Proof. Combining Lemma 2.1 with the first equality in Definition 2.8 implies that if $\mathbf{E} \succeq \mathbf{G}$, then $\Sigma_c(\mathbf{E}) \geq \Sigma_c(\mathbf{G})\ \forall\ c$. Now, the SE operator of Definition 2.9 can be interpreted as an average over the spatial dimension of local MMSE's. The local MMSE's for each position $c = 1, \cdots, \Gamma$ are the ones of B-dimensional equivalent AWGN channels with noise $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_B\, \Sigma_c^2/\log_2 B)$. These are non-decreasing functions of Σ_c^2: this is intuitively clear but we provide a justification based on an explicit formula for the derivative below. Thus $[T_{\text{co}}(\mathbf{E})]_r \geq [T_{\text{co}}(\mathbf{G})]_r\ \forall\ r$, which means $T_{\text{co}}(\mathbf{E}) \succeq T_{\text{co}}(\mathbf{G})$.
The derivative of the MMSE of the Gaussian channel with i.i.d. noise, distributed as $\mathcal{N}(\mathbf{0}, \mathbf{I}_B\, \Sigma^2)$, can be computed as

$$\begin{aligned}\frac{d\,\text{mmse}(\Sigma)}{d(\Sigma^{-2})} &= \frac{d}{d(\Sigma^{-2})}\mathbb{E}_{\mathbf{X},\mathbf{Y}}\big[\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_2^2\big] \\ &= -2\mathbb{E}_{\mathbf{X},\mathbf{Y}}\big[\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_2^2 \text{Var}[\mathbf{X}|\mathbf{Y}]\big]. \end{aligned} \tag{2.26}$$

This formula is valid for vector distributions $p_0(\mathbf{x})$, and in particular, for our B-dimensional sections. It confirms that T_{un} (resp. $[T_{\text{co}}]_r$) is a non-decreasing function of Σ (resp. Σ_c). In particular the local MMSE's for each position $c = 1, \cdots, \Gamma$ in definition 2.9 are non-decreasing. □

Corollary 2.1. *The SE operator of the coupled system maintains degradation in time, i.e.* $T_{\text{co}}(\boldsymbol{E}^{(t)}) \preceq \boldsymbol{E}^{(t)}$ *implies* $T_{\text{co}}(\boldsymbol{E}^{(t+1)}) \preceq \boldsymbol{E}^{(t+1)}$*. Similarly,* $T_{\text{co}}(\boldsymbol{E}^{(t)}) \succeq \boldsymbol{E}^{(t)}$ *implies* $T_{\text{co}}(\boldsymbol{E}^{(t+1)}) \succeq \boldsymbol{E}^{(t+1)}$*. Furthermore, if we take the initial conditions* $\mathbf{E}^{(0)} = \mathbf{1}$ *(the all one-vector) or* $\mathbf{E}^{(0)} = \mathbf{0}$ *(the all zero-vector) the* limiting profile

$$\lim_{t\to\infty} \boldsymbol{E}^{(t)} := T_{\text{co}}^{(\infty)}(\boldsymbol{E}^{(0)}), \tag{2.27}$$

exists. Finally under Assumption 2.1 the limiting profile is a fixed point of T_{co}*, i.e.,*

$$T_{\text{co}}(T_{\text{co}}^{(\infty)}(\boldsymbol{E}^{(0)})) = T_{\text{co}}^{(\infty)}(\boldsymbol{E}^{(0)}). \tag{2.28}$$

These properties are verified by T_{un} *for the underlying system as well.*

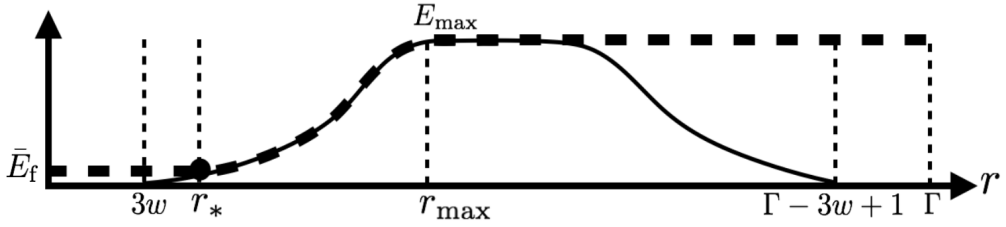

Figure 2.6: A non-symmetric error profile in a typical SE iteration. The solid line corresponds to the original spatially coupled system and the dashed line to the *modified* system. The error profile of the original system has a 0 plateau for all $r \leq 3w$ and it increases until $r_{\max}$ where it reaches its maximum value $E_{\max} \in [0, 1]$. It flattens after $r_{\max}$ then it decreases to reach 0 at $\Gamma - 3w + 1$ and remains null after. The non-symmetric shape of the double-sided wave in Fig. 2.6 emphasises that we are considering the generic case of non-symmetric coupling strength when designing spatially coupled matrices (see Section 2.2). The error profile of the modified system (dashed line) starts with a plateau at $\bar{E}_{\mathrm{f}}$ for all $r \leq r_*$, where $r_* + 1$ is the first position s.t the original profile is at least $\bar{E}_{\mathrm{f}}$, and then matches that of the original system for all $r \in \{r_*, \dots, r_{\max}\}$. It then saturates to $E_{\max}$ for all $r \geq r_{\max}$. Note that if $\mathbf{E} \preceq \bar{\mathbf{E}}_{\mathrm{f}}$ then $r_* = r_{\max}$. By construction, the error profile of the modified system is non-decreasing and degraded with respect to that of the original system.

Proof. First we note $T_{\mathrm{co}}(\mathbf{E}^{(t)}) \preceq \mathbf{E}^{(t)}$ means $\mathbf{E}^{(t+1)} \preceq \mathbf{E}^{(t)}$ and thus by Lemma 2.3 $T_{\mathrm{co}}(\mathbf{E}^{(t+1)}) \preceq T_{\mathrm{co}}(\mathbf{E}^{(t)})$ which means $T_{\mathrm{co}}(\mathbf{E}^{(t+1)}) \preceq \mathbf{E}^{(t+1)}$. The same argument shows that $T_{\mathrm{co}}(\mathbf{E}^{(t)}) \succeq \mathbf{E}^{(t)}$ implies $T_{\mathrm{co}}(\mathbf{E}^{(t+1)}) \succeq \mathbf{E}^{(t+1)}$. Let us show the existence of the limit (2.27) when we start with the initial condition $\mathbf{E}^{(0)} = \mathbf{1}$. This flat profile is maximal at every position thus after one iteration we necessarily have $\mathbf{E}^{(1)} \preceq \mathbf{E}^{(0)}$. Applying t times the operator T_{co} we get $\mathbf{E}^{(t+1)} \preceq \mathbf{E}^{(t)}$ which means $E_r^{(t+1)} \leq E_r^{(t)}$. Thus for every position we have a non-increasing sequence which is non-negative. Thus the sequence converges and $\lim_{t\to\infty} \mathbf{E}^{(t)} = T_{\mathrm{co}}^{(\infty)}(\mathbf{1})$ exists. The same argument applies if we start from the initial condition $\mathbf{E}^{(0)} = \mathbf{0}$ (the limit may be different of course). To show the last statement (2.28) we argue that T_{co} is continuous with respect to $\mathbf{E}$. We already noted after Definition 2.3 that the denoiser $[g_{\mathrm{in}}]_i$ is a continuous function of $\Sigma \geq 0$. Clearly, the denoiser satisfies $0 \leq [g_{\mathrm{in}}]_i \leq 1$ also, and so does the expression $([g_{\mathrm{in}}]_i - s_i)^2$. A look at the Definition 2.9 of $[T_{\mathrm{co}}(\mathbf{E})]_r$ thus shows, by Lebesgue's dominated convergence theorem, that $[T_{\mathrm{co}}(\mathbf{E})]_r$ is jointly continuous in $\Sigma_c(\mathbf{E})$, $c = 1, \cdots, \Gamma$. Thanks to Definition 2.8 and the Assumption 2.1 of continuity of $\Sigma(E)$, we conclude that T_{co} is a continuous function of $\mathbf{E}$. □

Corollary 2.2. *Starting from the error profile $\boldsymbol{E}^{(0)} = \boldsymbol{1}$ and due to the pinning condition, as the SE progresses the error profile must adopt the shape of the solid line shown on Fig. 2.6: it is 0 for $r \leq 3w$, non-decreasing for $3w \leq r \leq$*

$r_{\max}$, *non-increasing for* $r_{\max} \leq r \leq \Gamma - 3w + 1$ *and* 0 *for* $\Gamma - 3w + 1 \leq r \leq \Gamma$.

Proof. For a large enough Γ, the pinning condition (2.25) and the variance symmetry (2.10) ensure that in the first SE iteration $\Sigma_c^2(\mathbf{E}^{(0)} = \mathbf{1})$ satisfies the following ordering along the positions: *i*) it is non-decreasing for all $c \in \{1, \ldots, 4w + 1\}$, *ii*) it is non-increasing for all $c \in \{\Gamma - 4w, \ldots, \Gamma\}$, *iii*) it is constant elsewhere. Using the pinning condition again and the fact that the componentwise SE operator is non-decreasing in Σ_c^2 (see the proof of Lemma 2.3), one can show that after the first SE iteration the error profile $\mathbf{E}^{(1)}$ must adopt the following ordering: *i*) it is non-decreasing for all $r \in \{1, \ldots, 5w+1\}$, *ii*) it is non-increasing for all $r \in \{\Gamma - 5w, \ldots, \Gamma\}$, *iii*) it is constant elsewhere. Repeating the same argument by recursion one deduces that a double-sided wave (solid line shown in Fig. 2.6) propagates inwards as the SE progresses. □

Recall that state evolution is initialized with $\mathbf{E}^{(0)} = \mathbf{1}$. The iterations will eventually converge to a *fixed point profile*

$$\mathbf{E}^{(\infty)} := T_{\text{co}}^{(\infty)}(\mathbf{1}). \tag{2.29}$$

The fixed point reached by SE may be the "good" MSE floor profile $\mathbf{E}_{\text{f}}$ or may be a "bad" profile which is strictly degraded with respect to $\mathbf{E}_{\text{f}}$.

2.5.2 Proof of Threshold Saturation

The goal of this section is to arrive at a proof of the two main results, namely Theorem 2.1 and Corollary 2.3, both formulated at the end of the section. In this section we consider rates in the range $0 < R < R_{\text{pot}}$. Thus the gap given in Definition 2.12 is strictly positive and finite, i.e., $0 < \Delta F_{\text{un}} < +\infty$.

Definition 2.15 (The pseudo error floor)**.** *We fix* $0 < \eta < 1$ *(the reader may as well think of* $\eta = 1/2$ *in all subsequent arguments of this section). It can be shown that continuity of* $\Sigma(E)$ *(Assumption 2.1) implies that the potential function* $F_{\text{un}}(E)$ *is continuous for* $E \in [0, 1]$*. In particular it is continuous at the error floor* E_{f}*. Therefore we can find* $\delta(\eta, B, R) > 0$ *such that* $|F_{\text{un}}(E) - F_{\text{un}}(E_{\text{f}})| \leq \eta \Delta F_{\text{un}}$ *whenever* $|E - E_{\text{f}}| \leq \delta(\eta, B, R)$*. Now we take any* $0 < \epsilon < \delta(\eta, B, R)$ *and set* $\bar{E}_{\text{f}} = E_{\text{f}} + \epsilon$*. We have in particular* $|F_{\text{un}}(\bar{E}_{\text{f}}) - F_{\text{un}}(E_{\text{f}})| \leq \eta \Delta F_{\text{un}}$*. This number* $\bar{E}_{\text{f}}$*, will serve as a "pseudo error floor" in the analysis.*

Definition 2.16 (The modified system)**.** *The modified system is a modification of the SE iterations defined by applying two saturation constraints to the error profile of the original system at every iteration. First recall that the error profile of the original system has a* 0 *plateau for all* $r \leq 3w$ *and increases until* $r_{\max}$ *where it reaches its maximum value* $E_{\max} \in [0, 1]$*. It flattens after* $r_{\max}$ *then it decreases to reach* 0 *at* $\Gamma - 3w + 1$ *and remains null after. Now take any* $0 < \epsilon < \delta(\eta, B, R)$ *and set* $\bar{E}_{\text{f}} = E_{\text{f}} + \epsilon$ *where* E_{f} *is the true error floor. At each iteration the profile of the modified system is defined by applying the following*

two saturation constraints: (i) the profile is set to the pseudo error floor $\bar{E}_{\rm f}$ for all $r \le r_*$, where r_*+1 is the first position s.t the original profile is at least $\bar{E}_{\rm f}$; (ii) the profile is set to $E_{\max}$ for all $r \ge r_{\max}$. For $r \in \{r_*, \dots, r_{\max}\}$ the profiles of the modified and original systems are equal.

Fig. 2.6 gives an illustration of this definition: the full line corresponds to the original system and the dashed one to the modified system. By construction, the error profile of the modified system is non-decreasing and degraded with respect to that of the original system. We note that when the error floor is non-vanishing (e.g. on the AWGN channel) we could take in the analysis $\bar{E}_{\rm f} = E_{\rm f} + \epsilon \to E_{\rm f}$ for fixed code parameters. However for zero error floor we need to have $\epsilon > 0$ in the analysis. For code parameters w and Γ large enough we can make ϵ arbitrarily small.

The fixed point profile of the modified system is degraded with respect to $\mathbf{E}^{(\infty)}$, thus the modified system serves as an upper bound in our proof. Note that the SE iterations of the modified system also satisfy the monotonicity properties of $T_{\rm co}$ (see Section 2.5.1). Moreover, the modified system preserves the shape of the single-sided wave at all times. In the rest of this section we shall work with the modified system.

We now choose a proper shift of the saturated profile in Definition 2.17, and then evaluate the change in potential due to this shift in two different ways in Lemma 2.6 and Lemma 2.8. Theorem 2.1 and Corollary 2.3 will then be easy consequences.

Definition 2.17 (Shift operator)**.** *The* shift operator *is defined pointwise as* $[\mathrm{S}(\boldsymbol{E})]_1 := \bar{E}_{\rm f},\ [\mathrm{S}(\boldsymbol{E})]_r := E_{r-1}$.

Lemma 2.4. *Let $\boldsymbol{E}$ be a fixed point profile of the modified system initialized with $\boldsymbol{E}^{(0)} = \boldsymbol{1}$. Then there exist $\hat{t} \in [0,1]$ such that*

$$F_{\rm co}(\mathrm{S}(\boldsymbol{E})) - F_{\rm co}(\boldsymbol{E}) = \frac{1}{2}\sum_{r,r'=1}^{\Gamma} \Delta E_r \Delta E_{r'} \left[\frac{\partial^2 F_{\rm co}}{\partial E_r \partial E_{r'}}\right]_{\hat{\mathbf{E}}}.$$

where $\Delta E_r := E_r - E_{r-1}$ and $\hat{\mathbf{E}} := (1-\hat{t})\boldsymbol{E} + \hat{t}\mathrm{S}(\boldsymbol{E})$. Note that $\hat{t}$ depends in a non-trivial fashion on $\boldsymbol{E}$.

Proof. Consider $F_{\rm co}(t) := F_{\rm co}(\mathbf{E} + t(\mathrm{S}(\mathbf{E}) - \mathbf{E}))$ and note that $F_{\rm co}(0) = F_{\rm co}(\mathbf{E})$, $F_{\rm co}(1) = F_{\rm co}(\mathrm{S}(\mathbf{E}))$. Since $[\mathrm{S}(\mathbf{E})]_r = E_r + \Delta\mathbf{E}_r$ the mean value theorem yields

$$F_{\rm co}(\mathrm{S}(\mathbf{E})) - F_{\rm co}(\mathbf{E}) = -\sum_{r=1}^{\Gamma} \Delta E_r \left[\frac{\partial F_{\rm co}}{\partial E_r}\right]_{\mathbf{E}} + \frac{1}{2}\sum_{r,r'=1}^{\Gamma} \Delta E_r \Delta E_{r'} \left[\frac{\partial^2 F_{\rm co}}{\partial E_r \partial E_{r'}}\right]_{\hat{\mathbf{E}}}, \tag{2.30}$$

for some suitable $\hat{t} \in [0,1]$. By saturation of $\mathbf{E}$, $\Delta E_r = 0\ \forall\ r \in \mathcal{B} := \{1, \dots, r_*\} \cup \{r_{\max}+1, \dots, \Gamma\}$. Moreover for $r \notin \mathcal{B}$, $E_r = [T_{\rm co}(\mathbf{E})]_r$, and thus by Lemma 2.2 the potential derivative cancels at these positions. Hence the first sum in the right hand side of (2.30) cancels. □

Lemma 2.5. *The fixed point profile of the modified system initialized with* $\boldsymbol{E}^{(0)} = \boldsymbol{1}$ *is* smooth, *meaning that* ΔE_r *satisfies the following*

$$\begin{aligned}|\Delta E_r| &\leq \frac{g_* + \bar{g}}{w\underline{g}} \exp(-c(B)\Sigma^{-2}(E_{r+w})) \\ &\leq \frac{g_* + \bar{g}}{w\underline{g}},\end{aligned}$$

where w *is the coupling window and* $c(B) > 0$ *is a constant depending only on* B*; whereas* g_**,* $\bar{g}$ *and* $\underline{g}$ *correspond to the design function defined in Section 2.2.2.*

Proof. $\Delta E_r = 0$ for all $r \in \mathcal{B}$. By construction of $\{J_{r,c}\}$ we have

$$J_{r,c} \leq \frac{g_w((r-c)/w)}{\underline{g}(2w+1)}.$$

Moreover from Definitions 2.8 and 2.9 of the coupled state evolution operator, the fact that mmse is an increasing function of the noise and Lemma 2.1, we have $\mathrm{mmse}(\Sigma_c(\mathbf{E})) \leq \mathrm{mmse}(\Sigma(E_{r+w}))$ for $c = r-w, \cdots, r+w$. Thus using Lipschitz continuity of g_w, we have for all $r \notin \mathcal{B}$ that

$$\begin{aligned}|\Delta E_r| &= \Big|[T_{\mathrm{co}}(\mathbf{E})]_r - [T_{\mathrm{co}}(\mathbf{E})]_{r-1}\Big| = \Big|\sum_{c=1}^{\Gamma}(J_{r,c} - J_{r-1,c})\mathrm{mmse}(\Sigma_c(\mathbf{E}))\Big| \\ &\leq \frac{\mathrm{mmse}(\Sigma(E_{r+w}))}{(2w+1)\underline{g}} \sum_{c=1}^{\Gamma}\Big|g_w\big(\frac{r-c}{w}\big) - g_w\big(\frac{r-1-c}{w}\big)\Big| \\ &\leq \frac{\mathrm{mmse}(\Sigma(E_{r+w}))}{(2w+1)\underline{g}}\Big(2w\frac{g_*}{w} + |g_w(1)| + |g_w(-1)|\Big) \\ &< \frac{\mathrm{mmse}(\Sigma(E_{r+w}))}{2w\underline{g}}\Big(2g_* + 2\bar{g}\Big) \\ &\leq \frac{g_* + \bar{g}}{w\underline{g}}\exp(-c(B)\Sigma^{-2}(E_{r+w})). \end{aligned} \tag{2.31}$$

The last inequality is obtained by knowing that for an equivalent AWGN channel of variance Σ^2 and under *discrete prior*, $\mathrm{mmse}(\Sigma) \leq \exp(-c\Sigma^{-2})$ where c is some positive number that depends on the prior (see e.g. Appendix D of [104] for an explicit proof). Here the prior is uniform over sections so this number depends only on B. □

Lemma 2.6. *Let* $\boldsymbol{E}$ *be a fixed point profile of the modified system initialized with* $\boldsymbol{E}^{(0)} = \boldsymbol{1}$*. Then the coupled potential verifies*

$$\frac{1}{2}\Big|\sum_{r,r'=1}^{\Gamma}\Delta E_r \Delta E_{r'}\left[\frac{\partial^2 F_{\mathrm{co}}}{\partial E_r \partial E_{r'}}\right]_{\hat{\mathbf{E}}}\Big| < \frac{K(B,\bar{g},\underline{g},g_*)}{(E_{\mathrm{f}}+\epsilon)^{2\beta}Rw}.$$

where $K(B,\bar{g},\underline{g},g_*) > 0$*. The important point here is that the estimate is* $\mathcal{O}(w^{-1})$*.*

Proof. First remark that a fixed point of the modified system satisfies $\mathbf{E} \succeq \mathbf{E}_{\mathrm{f}}$. For $\mathbf{E} = \mathbf{E}_{\mathrm{f}}$ the result is immediate since $\Delta E_r = 0$. It remains to prove this lemma for $\mathbf{E}$ a fixed point of the modified system such that $\mathbf{E} \succ \mathbf{E}_{\mathrm{f}}$. In Appendix 2.8.3 we prove that

$$\Big[\frac{\partial^2 F_{\mathrm{co}}}{\partial E_r \partial E_{r'}}\Big]_{\hat{\mathbf{E}}} \leq \delta_{r,r'} \frac{K_1(B,\bar{g},\underline{g})}{(E_{\mathrm{f}}+\epsilon)R} + 1_{|r-r'|\leq 2w+1} \frac{K_2(B,\bar{g},\underline{g})}{(E_{\mathrm{f}}+\epsilon)^{2\beta}R(2w+1)} \tag{2.32}$$

for some finite positive $K_1(B,\bar{g},\underline{g})$ and $K_2(B,\bar{g},\underline{g})$ independent of w and Γ. Since $\Delta E_r \geq 0$, using the triangle inequality we get

$$\begin{aligned}
&\frac{1}{2}\Big| \sum_{r,r'=1}^{\Gamma} \Delta E_r \Delta E_{r'} \left[\frac{\partial^2 F_{\mathrm{co}}}{\partial E_r \partial E_{r'}}\right]_{\hat{\mathbf{E}}} \Big| \\
&\leq \frac{K_1(B,\bar{g},\underline{g})}{2(E_{\mathrm{f}}+\epsilon)R} \sum_{r=1}^{\Gamma} \Delta E_r^2 + \frac{K_2(B,\bar{g},\underline{g})}{2(E_{\mathrm{f}}+\epsilon)^{2\beta}R(2w+1)} \sum_{r=1}^{\Gamma} \Delta E_r \sum_{r'=r-w}^{r+w} \Delta E_{r'} \\
&\leq \frac{K_1(B,\bar{g},\underline{g})}{2(E_{\mathrm{f}}+\epsilon)R} \max_{r'} \Delta E_{r'} \sum_{r=r_*+1}^{r_{\max}} \Delta E_r + \frac{K_2(B,\bar{g},\underline{g})}{2(E_{\mathrm{f}}+\epsilon)^{2\beta}R} \max_{r'} \Delta E_{r'} \sum_{r=r_*+1}^{r_{\max}} \Delta E_r \\
&\leq \frac{K_1'(B,\bar{g},\underline{g},g_*)}{2(E_{\mathrm{f}}+\epsilon)Rw} + \frac{K_2'(B,\bar{g},\underline{g},g_*)}{2(E_{\mathrm{f}}+\epsilon)^{2\beta}Rw}.
\end{aligned}$$

To get the last inequality we used Lemma 2.5 and $\sum_{r=r_*+1}^{r_{\max}} \Delta E_r = E_{\max} - E_{r_*+1} < 1$. Finally, one can find $K(B,\bar{g},\underline{g},g_*) > 0$ such that the last estimate is smaller than

$$\frac{K(B,\bar{g},\underline{g},g_*)}{(E_{\mathrm{f}}+\epsilon)^{2\beta}Rw}$$

. □

The change in potential due to the shift can be also computed by direct evaluation as shown in the following lemmas.

Lemma 2.7. *Let $\boldsymbol{E}$ be a fixed point profile of the modified system initialized with $\boldsymbol{E}^{(0)} = \boldsymbol{1}$. If $\boldsymbol{E} \succ \bar{\boldsymbol{E}}_{\mathrm{f}}$, then $E_{\max}$ cannot be in the basin of attraction to the MSE floor, i.e., $E_{\max} \notin \mathcal{V}_0$.*

Proof. Knowing that $\mathbf{E} \succ \bar{\mathbf{E}}_{\mathrm{f}}$ and also that $\mathbf{E}$ is non-decreasing implies $\bar{E}_{\mathrm{f}} < E_{\max}$. Moreover, we have that

$$\begin{aligned}
E_{\max} = [T_{\mathrm{co}}(\mathbf{E})]_{r_{\max}} &= \sum_{c=1}^{\Gamma} J_{r_{\max},c} \,\mathrm{mmse}(\Sigma_c(\mathbf{E})) \\
&\leq \sum_{c=1}^{\Gamma} J_{r_{\max},c} \,\mathrm{mmse}(\Sigma(E_{\max})) \leq T_{\mathrm{un}}(E_{\max}),
\end{aligned} \tag{2.33}$$

where the first inequality follows from the fact that $\Sigma_c(\mathbf{E}) \leq \Sigma(E_{\max})$ due to the variance symmetry (2.10) at $r_{\max}$ and the fact that $\mathbf{E}$ is non-decreasing.

The second inequality follows from the variance normalization (2.4). Applying the monotonicity of T_{un} on (2.33) yields

$$E_{\mathrm{f}} < \bar{E}_{\mathrm{f}} < E_{\max} \leq T_{\mathrm{un}}^{(\infty)}(E_{\max}), \tag{2.34}$$

which implies that $E_{\max} \notin \mathcal{V}_0$. □

Lemma 2.8. *Let $0 < \eta < 1$ fixed and $\bar{E}_{\mathrm{f}} = E_{\mathrm{f}} + \epsilon$ with any $0 < \epsilon < \delta(\eta, B, R)$ where $\delta(\eta, B, R)$ has been constructed in Definition 2.15. Let $\boldsymbol{E} \succ \bar{\mathbf{E}}_{\mathrm{f}}$ be a fixed point profile of the modified system initialized with $\boldsymbol{E}^{(0)} = \boldsymbol{1}$. Then $\boldsymbol{E}$ satisfies*

$$F_{\mathrm{co}}(\mathrm{S}(\boldsymbol{E})) - F_{\mathrm{co}}(\boldsymbol{E}) \leq -(1-\eta)\Delta F_{\mathrm{un}},$$

where ΔF_{un} is the free energy gap of the underlying system given in Definition 2.12.

Proof. The contribution of the change in the "energy" term is a perfect telescoping sum:

$$U_{\mathrm{co}}(\mathrm{S}(\mathbf{E})) - U_{\mathrm{co}}(\mathbf{E}) = U_{\mathrm{un}}(\bar{E}_{\mathrm{f}}) - U_{\mathrm{un}}(E_{\max}). \tag{2.35}$$

We now deal with the contribution of the change in the "entropy" term. Using the properties of the construction of $J_{r,c}$ we notice that for all $c \in \{2w+1, \ldots, \Gamma - 2w - 1\}$

$$\begin{aligned}\Sigma_{c+1}^{-2}(\mathrm{S}(\mathbf{E})) = \sum_{r=c+1-w}^{c+1+w} \frac{J_{r,c+1}}{\Sigma^2(E_{r-1})} &= \sum_{r=c-w}^{c+w} \frac{J_{r+1,c+1}}{\Sigma^2(E_r)} \\ &= \sum_{r=c-w}^{c+w} \frac{J_{r,c}}{\Sigma^2(E_r)} = \Sigma_c^{-2}(\mathbf{E})\end{aligned} \tag{2.36}$$

which yields

$$\begin{aligned}S_{\mathrm{co}}(\mathbf{E}) - S_{\mathrm{co}}(\mathrm{S}(\mathbf{E})) = S_{\mathrm{un}}(\Sigma_{\Gamma-2w}(\mathbf{E})) - S_{\mathrm{un}}(\Sigma_{2w+1}(\mathrm{S}(\mathbf{E}))) \\ - \sum_{c \in \mathcal{S}} [S_{\mathrm{un}}(\Sigma_c(\mathrm{S}(\mathbf{E}))) - S_{\mathrm{un}}(\Sigma_c(\mathbf{E}))],\end{aligned} \tag{2.37}$$

where $\mathcal{S} := \{1, \ldots, 2w\} \cup \{\Gamma - 2w + 1, \ldots, \Gamma\}$. By the saturation of the modified system, $\mathbf{E}$ possesses the following property

$$[\mathrm{S}(\mathbf{E})]_r = [\mathbf{E}]_r \quad \text{for all} \quad r \in \{1, \ldots, r_*\} \cup \{r_{\max} + 1, \ldots, \Gamma\}. \tag{2.38}$$

Hence, $\Sigma_c(\mathrm{S}(\mathbf{E})) = \Sigma_c(\mathbf{E})$ for all $c \in \mathcal{S}$ and thus the sum in (2.37) cancels. Furthermore, one can show, using the saturation of $\mathbf{E}$ and the variance symmetry (2.10), that $\Sigma_{2w+1}(\mathrm{S}(\mathbf{E})) = \Sigma(\bar{E}_{\mathrm{f}})$. The same arguments and the fact that $r_{\max} \leq \Gamma - 3w$ for $\mathbf{E} \succ \bar{\mathbf{E}}_{\mathrm{f}}$ lead to $\Sigma_{\Gamma-2w}(\mathbf{E}) = \Sigma(E_{\max})$. Hence, (2.37) yields

$$S_{\mathrm{co}}(\mathbf{E}) - S_{\mathrm{co}}(\mathrm{S}(\mathbf{E})) = S_{\mathrm{un}}(\Sigma(E_{\max})) - S_{\mathrm{un}}(\Sigma(\bar{E}_{\mathrm{f}})). \tag{2.39}$$

Combining (2.35) with (2.39) gives

$$\begin{aligned} F_{\rm co}(\mathrm{S}(\mathbf{E})) - F_{\rm co}(\mathbf{E}) &= -(F_{\rm un}(E_{\max}) - F_{\rm un}(\bar{E}_{\rm f})) \\ &= -(F_{\rm un}(E_{\max}) - F_{\rm un}(E_{\rm f})) + (F_{\rm un}(\bar{E}_{\rm f}) - F_{\rm un}(E_{\rm f})). \end{aligned}$$

Using the definition of the free energy gap (Definition 2.12), the fact that $E_{\max} \notin \mathcal{V}_0$ (Lemma 2.7), and $F_{\rm un}(\bar{E}_{\rm f}) - F_{\rm un}(E_{\rm f}) \leq \eta \Delta F_{\rm un}$ we find

$$F_{\rm co}(\mathrm{S}(\mathbf{E})) - F_{\rm co}(\mathbf{E}) \leq -(1-\eta)\Delta F_{\rm un}.$$

□

Using Lemmas 2.4, 2.6, 2.8 we now prove threshold saturation.

Theorem 2.1. *Let $0 < \eta < 1$ fixed and $\bar{E}_{\rm f} = E_{\rm f} + \epsilon$ with any $0 < \epsilon < \delta(\eta, B, R)$ where $\delta(\eta, B, R)$ has been constructed in Definition 2.15. Fix*

$$R < R_{\rm pot} \quad \text{and} \quad w > \frac{K(B, \bar{g}, \underline{g}, g_*)}{(E_{\rm f} + \epsilon)^{2\beta} R (1-\eta) \Delta F_{\rm un}} \tag{2.40}$$

Then the fixed point profile $\boldsymbol{E}^{(\infty)}$ of the coupled SE must satisfy $\boldsymbol{E}^{(\infty)} \preceq \bar{\boldsymbol{E}}_{\rm f}$.

Proof. Assume that, under these hypotheses, the fixed point profile of the modified system initialized with $\mathbf{E}^{(0)} = \mathbf{1}$ is such that $\mathbf{E} \succ \bar{\mathbf{E}}_{\rm f}$. On one hand by Lemma 2.8 we have for $R < R_{\rm pot}$ a positive $\Delta F_{\rm un}$ and

$$|F_{\rm co}(\mathbf{E}) - F_{\rm co}(\mathrm{S}(\mathbf{E}))| \geq (1-\eta)\Delta F_{\rm un}.$$

On the other hand by Lemmas 2.4 and 2.6

$$|F_{\rm co}(\mathbf{E}) - F_{\rm co}(\mathrm{S}(\mathbf{E}))| \leq \frac{K(B, \bar{g}, \underline{g}, g_*)}{(E_{\rm f} + \epsilon)^{2\beta} R w}.$$

Thus we get

$$w \leq \frac{K(B, \bar{g}, \underline{g}, g_*)}{(E_{\rm f} + \epsilon)^{2\beta} R (1-\eta) \Delta F_{\rm un}}$$

which is a contradiction. Hence, $\mathbf{E} \preceq \bar{\mathbf{E}}_{\rm f}$. Since $\mathbf{E} \succeq \mathbf{E}^{(\infty)}$ we have $\mathbf{E}^{(\infty)} \preceq \bar{\mathbf{E}}_{\rm f}$. □

The most important consequence of this theorem is a statement on the GAMP threshold,

Corollary 2.3. *By first taking $\Gamma \to \infty$ and then $w \to \infty$, the GAMP threshold of the coupled ensemble satisfies $R_{\rm co} \geq R_{\rm pot}$.*

This result follows from Theorem 2.1 and Definition 2.10. Once the limit $w \to +\infty$ is taken we can send $\epsilon \to 0$ and the pseudo error floor tends to the true error floor $\bar{E}_{\rm f} \to E_{\rm f}$.

2.5.3 Discussion

Corollary 2.3 says that the GAMP threshold for the coupled codes saturates the potential threshold in the limit $w \to +\infty$. It is in fact not possible to have the strict inequality $R_{\text{co}} > R_{\text{pot}}$, so in fact equality holds, but the proof would require a separate argument that we omit here because it is not so informative. Besides, this argument is not needed in order to conclude that sparse superposition codes universally achieve capacity under GAMP decoding when $B \to +\infty$. Indeed we have necessarily $R_{\text{co}} < C$ and we know (Section 2.6) that $\lim_{B\to\infty} R_{\text{pot}} = C$. Thus $\lim_{B\to\infty} R_{\text{co}} = C$.

We emphasize that Theorem 2.1 and Corollary 2.3 hold for a large class of estimation problems with random linear mixing [110]. Both the SE and potential formulations of Section 2.4 as well as the proof given in the present section are not restricted to SS codes. Indeed all the definitions and results are obtained for any memoryless channel P_{out} and can be generalized for any factorizable (over B-dimensional sections) prior of the message (or signal) $\mathbf{s}$.

Theorem 2.1 states that for w large enough the state evolution iterations will drive the MSE profile below some pseudo error floor $\bar{E}_{\text{f}} = E_{\text{f}} + \epsilon$. This is then enough information to deduce that the threshold saturation phenomenon happens in the limit where $w \to +\infty$ (and note we do not expect full threshold saturation, i.e., $R_{\text{co}} \to R_{\text{pot}}$ for finite w). However, it is worth pointing out that the condition (2.40) in Theorem 2.1 on the size of the coupling window is most probably *not* optimal. We conjecture that a better bound should hold where $w > C/\Delta F_{\text{un}}$ for some $C > 0$ which does *not* diverge when $E_{\text{f}} + \epsilon \to 0$. The appearance of the error floor in the denominator can be traced back to inequality (2.32) whose derivation is detailed in Appendix 2.8.3. One possible way to cancel this divergence would be to obtain a better bound on ΔE_r than the one given by (2.31). More precisely if E_{r+w} can be replaced by E_r then the proof of Lemma 2.6 and Theorem 2.1 would give a more reasonable lower bound for w. Carrying out this program presents technical difficulties in the analysis of coupled state evolution which we have not overcome in this work. The present difficulties do not appear in the analysis of spatially coupled LDPC codes [27].

2.6 Large Alphabet Size Analysis and Connection with Shannon's Capacity

We now show that as the alphabet size B increases, the potential threshold of SS codes approaches Shannon's capacity $R_{\text{pot}}^{\infty} := \lim_{B\to\infty} R_{\text{pot}} = C$ (Fig. 2.7), and also that $\lim_{B\to\infty} E_{\text{f}} = 0$. These are "static" or "information theoretic" properties of the code independent of the decoding algorithm. Nevertheless this result has an algorithmic consequence. The threshold saturation established in Corollary 2.3 for spatially coupled SS codes implies that

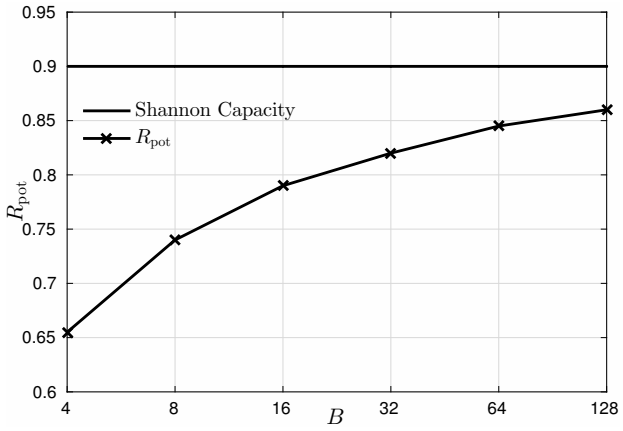

Figure 2.7: The potential threshold v.s the alphabet size B for the BEC with erasure probability $\epsilon = 0.1$.

optimal decoding can actually be performed using the GAMP decoder, i.e. $\lim_{B\to\infty} R_{\text{co}} = C$, because $R_{\text{pot}} \le R_{\text{co}} \le C$.

The potential of the underlying system contains all the information about R_{pot} and R_{un}. Hence, we proceed by computing the potential in the large B regime,

$$\varphi_{\text{un}}(E) := \lim_{B\to\infty} F_{\text{un}}(E). \tag{2.41}$$

The limit (2.41) is computed in [78, 126] for the AWGN channel. Extending this computation to the present setting, one obtains

$$\varphi_{\text{un}}(E) = U_{\text{un}}(E) - \max\Big(0, 1 - \frac{1}{2\ln(2)\Sigma^2(E)}\Big). \tag{2.42}$$

The extension from the AWGN case is straightforward, the $U_{\text{un}}(E)$ term in $F_{\text{un}}(E)$ is independent of B while the $S_{\text{un}}(\Sigma(E))$ term remains the same. The difference is only in the computation of the effective noise $\Sigma(E)$, which is independent of B. We note that (2.42) is not a trivial asymptotic calculation because the "entropy" term $S_{\text{un}}(\Sigma(E))$ involves a B-dimensional integral (see Definition 2.11). Since $B \to \infty$, this amounts to compute a "partition function" (or equivalently solve a non-linear estimation problem where the signal has one non-zero component). We have not attempted to make this asymptotic computation rigorous but we expect that the results of [78, 126] could be made rigorous using the recent work [103, 104, 122, 127].

The analysis of (2.42) for $E \in [0, 1]$ leads to the following

Claim 2.1. *For a fixed rate R and $E \in [0, 1]$, the only possible local minima of $\varphi_{\text{un}}(E)$ are at $E = 0$ and $E = 1$. Furthermore, for $E' \in \big\{E \in [0, 1] \mid 2\ln(2)\Sigma^2(E) < 1\big\}$ the minimum is at $E' = 0$ and for $E' \in \big\{E \in [0, 1] \mid 2\ln(2)\Sigma^2(E) > 1\big\}$ the minimum is at $E' = 1$.*

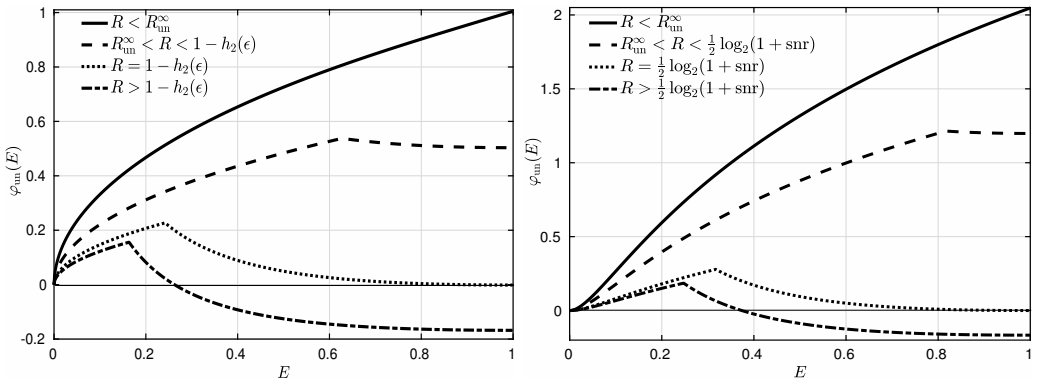

Figure 2.8: The large alphabet potential $\varphi_{\rm un}(E)$ (2.42) as a function of the error parameter E for the BSC (left) and AWGN (right) channels with $\epsilon = 0.1$ and ${\rm snr} = 10$ respectively. $\varphi_{\rm un}(E)$ is scaled such that $\varphi_{\rm un}(0) = 0$. For R below the "asymptotic" GAMP threshold $R_{\rm un}^{\infty}$, there is a unique minimum at $E = 0$ while just above $R_{\rm un}^{\infty}$, this minimum coexists with a local one at $E = 1$. At the optimal threshold of the code, that coincides with the Shannon capacity, the two minima are equal. Then, for $R > C$ the minimum at $E = 1$ becomes the global one, and thus decoding is impossible.

Note that this result was rigorously proven for the AWGN channel in [78] and then verified for several memoryless channels in [120]. A fully rigorous analysis of the function $\varphi_{\rm un}(E)$ would be lengthy; we thus only claim the result here, which is confirmed by numerical analysis.

The existence of a minimum at $E = 0$ means that the error floor $E_{\rm f}$, if it exists, vanishes as B increases (Fig. 2.8). Moreover, if $\Sigma^2(E) < (2\ln(2))^{-1}\ \forall\ E \in [0,1]$, which corresponds to the region $R < (2\ln(2))^{-1}\mathbb{E}_{p|1}[\mathcal{F}(p|1)]$, then $\varphi_{\rm un}(E)$ has a unique minimum at $E = 0$. Similarly if $\Sigma^2(E) > (2\ln 2)^{-1}\ \forall\ E \in [0,1]$, corresponding to $R > (2\ln(2))^{-1}\mathbb{E}_{p|0}[\mathcal{F}(p|0)]$, then $\varphi_{\rm un}(E)$ has a unique minimum at $E = 1$. For *intermediate rates* both minima exist.

Therefore, we identify the algorithmic GAMP threshold of the underlying ensemble, when $B \to +\infty$, as the smallest rate such that a second minimum appears,

$$R_{\rm un}^{\infty} := \lim_{B\to\infty} R_{\rm un} = \frac{\mathbb{E}_{p|1}[\mathcal{F}(p|1)]}{2\ln(2)} = \frac{\mathcal{F}(0|1)}{2\ln(2)}. \tag{2.43}$$

Recall $R_{\rm pot}$ is defined by the point where $\Delta F_{\rm un}$ switches sign (Definition 2.13). Thus $R_{\rm pot}^{\infty}$ can be obtained by equating the two minima of $\varphi_{\rm un}(E)$. The potential (2.42) takes the following values at the two minimizers

$$\varphi_{\rm un}(0) = -\frac{1}{R}\mathbb{E}_z\left[\int dy\,\phi(y|z,0)\log_2\left(\phi(y|z,0)\right)\right],$$
$$\varphi_{\rm un}(1) = -\frac{1}{R}\mathbb{E}_z\left[\int dy\,\phi(y|z,1)\log_2\left(\phi(y|z,1)\right)\right] - 1,$$

where $\phi(y|z,E)$ is given in Definition 2.11. Then, setting $\varphi_{\text{un}}(1) = \varphi_{\text{un}}(0)$ yields

$$R_{\text{pot}}^{\infty} = -\int\int dzdy\mathcal{N}(z|0,1)P_{\text{out}}(y|z)\log_2\Big(\int d\tilde{z}\mathcal{N}(\tilde{z}|0,1)P_{\text{out}}(y|\tilde{z})\Big) + \int\int dzdy\mathcal{N}(z|0,1)P_{\text{out}}(y|z)\log_2\Big(P_{\text{out}}(y|z)\Big). \tag{2.44}$$

We will now recognize that this expression is the Shannon capacity of W for a proper choice of the map π.

Let $\mathcal{A}$ and $\mathcal{B}$ be the input and output alphabet of W respectively, where $\mathcal{A},\mathcal{B} \subseteq \mathbb{R}$ have discrete or continuous supports. Call $\mathcal{P}$ the capacity-achieving input distribution associated with W. Choose $\pi : \mathbb{R} \to \mathcal{A}$ such that $i)$ $P_{\text{out}}(y|z) = W(y|\pi(z))$ and $ii)$ if $Z \sim \mathcal{N}(0,1)$, then $\pi(Z) \sim \mathcal{P}$. This map converts a standard Gaussian random variable Z onto a channel-input random variable $\pi(Z) = A$ with capacity-achieving distribution $\mathcal{P}(a)$. Recall that π can be viewed equivalently as part of the code or of the channel.

Now using the relation

$$\int dz\mathcal{N}(z|0,1)P_{\text{out}}(y|z) = \int dz\mathcal{N}(z|0,1)W(y|\pi(z)) = \int da\mathcal{P}(a)W(y|a),$$

(2.44) can be expressed equivalently as

$$R_{\text{pot}}^{\infty} = -\int\int dyda\mathcal{P}(a)W(y|a)\log_2\Big(\int d\tilde{a}\mathcal{P}(\tilde{a})W(y|\tilde{a})\Big) + \int\int dyda\mathcal{P}(a)W(y|a)\log_2\Big(W(y|a)\Big). \tag{2.45}$$

The first term in (2.45) is nothing but the Shannon entropy $H(Y)$ of the channel output-distribution. The second term eaquals minus the conditional entropy $H(Y|A)$ of the channel-output distribution given the input $A = \pi(Z)$ with capacity-achieving distribution. Thus, R_{pot}^{∞} is the Shannon capacity of W. Combining this result with Corollary 2.3, we can argue that spatially coupled SS codes allow to communicate reliably up to Shannon's capacity over any memoryless channel under low complexity GAMP decoding.

An essential question remains on how to find the proper map π for a given memoryless channel. In the case of discrete input memoryless symmetric channels, Shannon's capacity can be attained by inducing a uniform input distribution $\mathcal{P} = \mathcal{U}_{\mathcal{A}}$. Let us call q the cardinality of $\mathcal{A} = \{a_1,\ldots,a_q\}$. In this case the mapping π is simply $\pi(z) = a_i$ if $z \in]z_{(i-1)/q}, z_{i/q}]$, where $z_{i/q}$ is the i^{th} q-quantile[5] of the Gaussian distribution, with $z_0 = -\infty, z_1 = \infty$. For asymmetric channels, one can use some standard methods such as Gallager's mapping or more advanced ones [128] that introduce bias in the channel-input

[5]With $z_{i/q} = Q^{-1}(1-i/q)$, where $Q^{-1}(\cdot)$ is the inverse of the Gaussian Q-function defined by $Q(x) = \int_x^{+\infty} dt\, \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}$.

distribution in order to match the capacity-achieving one. This task is known as *distribution matching* and it has applications beyond achieving the capacity of asymmetric channels. Interestingly, SS codes can be employed to solve the distribution matching problem as we will see in Chapter 3. Furthermore, SS codes can perform joint distribution matching and channel coding.

We now illustrate our findings for various channels as depicted in Fig. 2.9 and Fig. 2.10.

2.6.1 AWGN Channel

We start showing that our results for the AWGN channel [118] are a special case of the present general framework. No map π is required and the Shannon capacity is directly obtained from (2.44) because the capacity-achieving input distribution for the AWGN channel is Gaussian. Thus, by replacing $P_{\text{out}}(y|z) = \mathcal{N}(y|z, 1/\text{snr})$ in (2.44), one recovers the Shannon capacity $R_{\text{pot}}^{\infty} = \frac{1}{2}\log_2(1+\text{snr})$. Furthermore, from (2.43) one obtains the following algorithmic threshold as $B \to \infty$

$$R_{\text{un}}^{\infty} = \frac{1}{2\ln(2)(1+\text{snr}^{-1})}. \tag{2.46}$$

2.6.2 Binary Symmetric Channel

The BSC with flip probability[6] ϵ has transition probability $W(y|a) = (1-\epsilon)\delta(y-a) + \epsilon\delta(y+a)$, where $\mathcal{A} = \mathcal{B} = \{-1, 1\}$. The proper map is $\pi(z) = \text{sign}(z)$. For $Z \sim \mathcal{N}(0,1)$, this map induces uniform input distribution $\mathcal{U}_{\mathcal{A}} = 1/2$. So by replacing W and $\mathcal{U}_{\mathcal{A}}$ in (2.45), or equivalently $P_{\text{out}}(y|z) = (1-\epsilon)\delta(y-\pi(z)) + \epsilon\delta(y+\pi(z))$ into (2.44), one obtains the Shannon capacity of the BSC channel $R_{\text{pot}}^{\infty} = 1 - h_2(\epsilon)$ where h_2 is the binary entropy function. Using (2.43) this map also gives the algorithmic threshold

$$R_{\text{un}}^{\infty} = \frac{(1-2\epsilon)^2}{\pi\ln(2)}. \tag{2.47}$$

2.6.3 Binary Erasure Channel

Note that the BEC is also symmetric. Therefore, the same mapping $\pi(z) = \text{sign}(z)$ is used and leads to the Shannon capacity $R_{\text{pot}}^{\infty} = 1 - \epsilon$, where ϵ is the erasure probability. Moreover, from (2.43) the algorithmic threshold for the BEC when $B \to \infty$ is

$$R_{\text{un}}^{\infty} = \frac{1-\epsilon}{\pi\ln(2)}. \tag{2.48}$$

[6]With a slight abuse of notation, we use ϵ here as a channel parameter. Not to confuse with ϵ of Section 2.5 (Definition 2.15).

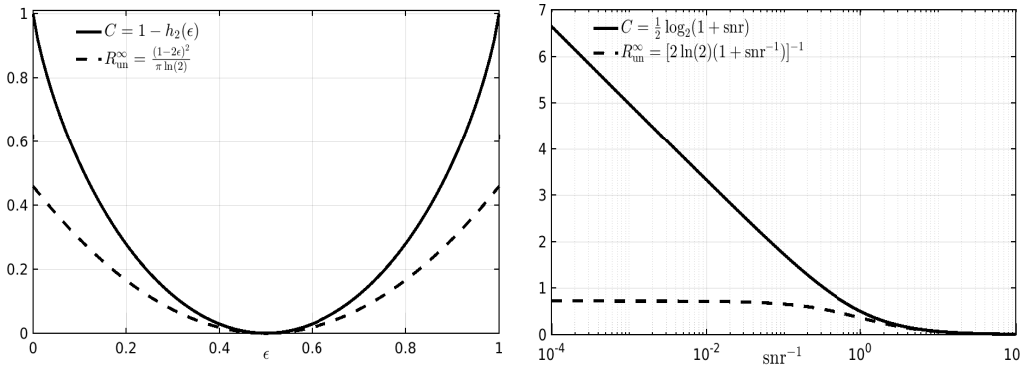

Figure 2.9: The capacities and GAMP thresholds in the infinite alphabet limits for the BSC (left) and AWGN (right) channels.

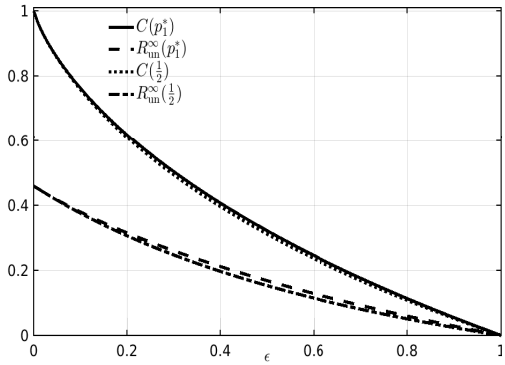

Figure 2.10: Capacity and GAMP threshold of the Z channel in the infinite alphabet limits. $C(p_1^*)$ and $R_{\rm un}^{\infty}(p_1^*)$ are the values under capacity-achieving input distribution, whereas $C(\frac{1}{2})$ and $R_{\rm un}^{\infty}(\frac{1}{2})$ are the values under uniform distribution.

2.6.4 Z Channel

The Z channel is the "most asymmetric" discrete channel. It has binary input and output $\mathcal{A} = \mathcal{B} = \{-1, 1\}$ with transition probability $W(y|a) = \delta(a-1)\delta(y-a) + \delta(a+1)[(1-\epsilon)\delta(y-a) + \epsilon\delta(y+a)]$, where ϵ is the flip probability of the -1 input. The map $\pi(z) = \text{sign}(z)$ leads to the *symmetric capacity* of the Z channel

$$R_{\rm pot}^{\infty}\big(\frac{1}{2}\big) = C\big(\frac{1}{2}\big) = h_2((1-\epsilon)/2) - h_2(\epsilon)/2, \tag{2.49}$$

where $C(\frac{1}{2})$ denotes the symmetric capacity, in other words the input-output mutual information when the input is uniformly distributed with $\mathcal{U}_{\mathcal{A}} = 1/2$. Under the same map $\pi(z)$, one obtains the following algorithmic threshold in the limit $B \to +\infty$

$$R_{\rm un}^{\infty}\big(\frac{1}{2}\big) = \frac{1-\epsilon}{\pi \ln(2)(1+\epsilon)}. \tag{2.50}$$

Note that the expression of $R_{\text{pot}}^{\infty}(\frac{1}{2})$ differs from Shannon's capacity. However, one can introduce bias in the input distribution and hence match the capacity-achieving one. To do so, the proper map defined in terms of the Q-function[7] is $\pi(z) = \text{sign}(z - Q^{-1}(p_1))$, where p_1 is the induced input probability of the bit 1 (see Chapter 3 for more details about the choice of the map).

By optimizing over p_1, one can obtain Shannon's capacity of the Z channel

$$R_{\text{pot}}^{\infty}(p_1^*) = C(p_1^*) = h_2((1-p_1^*)(1-\epsilon)) - (1-p_1^*)h_2(\epsilon), \tag{2.51}$$

with

$$p_1^* = 1 - [(1-\epsilon)(1+2^{h_2(\epsilon)/(1-\epsilon)})]^{-1}. \tag{2.52}$$

Using this optimal map, one obtains the following algorithmic threshold as depicted in Fig. 2.10

$$R_{\text{un}}^{\infty}(p_1^*) = \frac{(1-\epsilon)\left(e^{-[Q^{-1}(p_1^*)]^2/2}\right)^2}{4\pi \ln(2)(1-p_1^*)((1-p_1^*)\epsilon + p_1^*)}. \tag{2.53}$$

2.7 Open Challenges

We end up pointing out some open problems. In order to have a fully rigorous capacity achieving scheme over any memoryless channel, using spatially coupled SS codes and GAMP decoding, it must be shown that SE tracks the asymptotic performance of GAMP for the B-dimensional prior. We conjecture that this is indeed the case. The proof is beyond the scope of this work and would follow by extending the analysis of [69, 115] to the SS codes setting as done in [37] for AMP. It is also desirable to consider practical coding schemes using Hadamard-based operators or, more generally, row-orthogonal matrices. Another important point is to estimate at what rate the error floor vanishes as B increases (when it exists e.g., in the AWGN channel). Finally, finite size effects should be considered in order to assess the practical performance of these codes.

2.8 Appendix

2.8.1 Vectorial GAMP Algorithm

The GAMP algorithm was introduced for general estimation with random linear mixing in [110]. The extension to the present context of SS codes with B-dimensional prior was given in Section 2.3 of this chapter. On a dense graphical model, an important notion of equivalent AWGN channel is used to simplify the BP messages. This notion is due to the linear mixing and it is independent of the physical channel. The physical channel P_{out} is reflected

[7] Here $Q(x) = \int_x^{+\infty} dt \, \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}$.

Table 2.1: The expressions for g_{out} and $\mathcal{F}$.

	$[g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})]_i$	$\mathcal{F}(p\|E)$
General	$(\mathbb{E}[Z_i\|p_i,y_i,\tau_i]-p_i)/\tau_i$ $Y_i\sim P_{\text{out}}(\cdot\|z_i), Z_i\sim \mathcal{N}(p_i,\tau_i)$	See Definition 2.2
AWGNC	$\frac{y_i-p_i}{\tau_i+1/\text{snr}}$	$\frac{1}{1/\text{snr}+E}$
BSC	$\frac{(p_i-k_i)v_i^+ +(p_i+k_i)v_i^-}{\mathcal{Z}_{\text{BSC}}\tau_i}-\frac{p_i}{\tau_i}$	$\frac{Q'^2(1-2\epsilon)^2}{(Q+\epsilon-2\epsilon Q)(1-Q-\epsilon+2\epsilon Q)}$
BEC	$\frac{(p_i-k_i)h_i^+ +(p_i+k_i)h_i^- +2\epsilon\delta(y_i)p_i}{\mathcal{Z}_{\text{BEC}}\tau_i}-\frac{p_i}{\tau_i}$	$\frac{Q'^2(1-\epsilon)}{Q(1-Q)}$
ZC	$\frac{(p_i-k_i)v_i^+ +(p_i+k_i)\delta(y_i-1)}{\mathcal{Z}_{\text{ZC}}\tau_i}-\frac{p_i}{\tau_i}$	$\frac{Q'^2(1-\epsilon)^2}{Q+\epsilon(1-Q)}+\frac{Q'^2(1-\epsilon)}{1-Q}$

$$h_i^+=(1-\epsilon)\delta(y_i+1),\quad h_i^-=(1-\epsilon)\delta(y_i-1),$$
$$v_i^+=(1-\epsilon)\delta(y_i+1)+\epsilon\delta(y_i-1),\quad v_i^-=(1-\epsilon)\delta(y_i-1)+\epsilon\delta(y_i+1),$$
$$k_i=\exp\left(\tfrac{-p_i^2}{2\tau_i}\right)\sqrt{2\tau_i/\pi}+\text{erf}\left(\tfrac{p_i}{\sqrt{2\tau_i}}\right)p_i,\quad k_i'=k_ip_i+\text{erf}\left(\tfrac{p_i}{\sqrt{2\tau_i}}\right)\tau_i,$$
$$Q=\tfrac{1}{2}\text{erfc}\left(\tfrac{-p}{\sqrt{2E}}\right),\qquad Q'=\exp\left(\tfrac{-p^2}{2E}\right)\Big/\sqrt{2\pi E}$$
$$\mathcal{Z}_{\text{BEC}}=\text{erfc}\left(\tfrac{p_i}{\sqrt{2\tau_i}}\right)h_i^+ +\left(1+\text{erf}\left(\tfrac{p_i}{\sqrt{2\tau_i}}\right)\right)h_i^- +2\epsilon\delta(y_i),$$
$$\mathcal{Z}_{\text{ZC}}=\text{erfc}\left(\tfrac{p_i}{\sqrt{2\tau_i}}\right)v_i^+ +\left(1+\text{erf}\left(\tfrac{p_i}{\sqrt{2\tau_i}}\right)\right)\delta(y_i-1),$$
$$\mathcal{Z}_{\text{BSC}}=\text{erfc}\left(\tfrac{p_i}{\sqrt{2\tau_i}}\right)v_i^+ +\left(1+\text{erf}\left(\tfrac{p_i}{\sqrt{2\tau_i}}\right)\right)v_i^-$$

in the computation of the equivalent AWGN channel's parameter through the function $g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})$. This function is acting componentwise and can be interpreted as a *score function* of the parameter p_i associated with the distribution of Y_i. The general expression is

$$\begin{aligned}[g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})]_i &= (\mathbb{E}[z_i|p_i,y_i,\tau_i]-p_i)/\tau_i \\ &= \frac{\int dz_i P_{\text{out}}(y_i|z_i)\mathcal{N}(z_i|p_i,\tau_i)(z_i-p_i)/\tau_i}{\int dz_i P_{\text{out}}(y_i|z_i)\mathcal{N}(z_i|p_i,\tau_i)}\end{aligned} \tag{2.54}$$

This expression is also equal to $\partial_{p_i}\ln f(y_i|p_i,\tau_i)$ where f is the function occurring in Definition 2.2 of the Fisher information.

In Table 2.1 and Table 2.2[8] we give the explicit expressions for various channels as well as their derivatives used in the GAMP algorithm of Section 2.3 (where snr is the signal-to-noise ratio of the AWGN channel, ϵ the erasure or flip probability of the BSC, BEC and ZC). The expressions of the Fisher information used in SE of Section 2.4 are given as well. These involve the Gaussian error function $\text{erf}(x)=\frac{\sqrt{2}}{\pi}\int_0^x dt\, e^{-t^2}$ and its complement $\text{erfc}(x)=$

[8]Based on a joint work with E. Bıyık and J. Barbier [121].

Table 2.2: The expressions for $-\frac{\partial}{\partial \mathbf{p}} g_{\text{out}}$.

	$[-\frac{\partial}{\partial \mathbf{p}} g_{\text{out}}(\mathbf{p}, \mathbf{y}, \boldsymbol{\tau})]_i$
General	$(\tau_i - \text{Var}[Z_i \vert p_i, y_i, \tau_i])/\tau_i^2$ $Y_i \sim P_{\text{out}}(\cdot \vert z_i), Z_i \sim \mathcal{N}(p_i, \tau_i)$
AWGNC	$\frac{1}{\tau_i + 1/\text{snr}}$
BSC	$\frac{1}{\tau_i} - \frac{(p_i^2+\tau_i-k_i')v_i^+ + (p_i^2+\tau_i+k_i')v_i^-}{\mathcal{Z}_{\text{BSC}}{\tau_i}^2} + \left([g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})]_i + \frac{p_i}{\tau_i}\right)^2$
BEC	$\frac{1}{\tau_i} - \frac{(p_i^2+\tau_i-k_i')h_i^+ + (p_i^2+\tau_i+k_i')h_i^- + 2\epsilon\delta(y_i)(\tau_i+p_i^2)}{\mathcal{Z}_{\text{BEC}}{\tau_i}^2} +$ $\left([g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})]_i + \frac{p_i}{\tau_i}\right)^2$
ZC	$\frac{1}{\tau_i} - \frac{(p_i^2+\tau_i-k_i')v_i^+ + (p_i^2+\tau_i+k_i')\delta(y_i-1)}{\mathcal{Z}_{\text{ZC}}{\tau_i}^2} +$ $\left([g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})]_i + \frac{p_i}{\tau_i}\right)^2$

$1 - \text{erf}(x)$. Note that, for the sake of simplicity, all the expressions for the binary input channels of Table 2.1 (BSC, BEC and ZC) are given using the map $\pi(z) = \text{sign}(z)$. This map leads to a sub-optimal performance for the asymetric Z channel. The optimal map would require a bias in the input distribution as explained in Section 2.6.4.

2.8.2 State Evolution and Potential Function

In this appendix we prove Lemma 2.2. Namely, we show that the stationarity condition $\partial F_{\text{un}}/\partial E = 0$ for the potential function in Definition 2.11 implies the state evolution equation in Definition 2.3. We present a detailed derivation for the underlying uncoupled system. The proof of Lemma 2.2 for the coupled system follows exactly the same steps.

The calculation is best done by looking at F_{un} as a function of E and $\Sigma(E)^{-2}$, so that

$$\frac{dF_{\text{un}}}{dE} = -\frac{1}{2\ln(2)\Sigma(E)^2} - \frac{1}{R}\frac{\partial}{\partial E}\mathbb{E}_Z\Big[\int dy\, \phi(y|Z,E)\log_2 \phi(y|Z,E)\Big]$$
$$- \Big\{\frac{E}{2\ln(2)} + \frac{\partial}{\partial \Sigma(E)^{-2}}\mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\log_B \int d^B\mathbf{x}\, p_0(\mathbf{x})\theta(\mathbf{x},\mathbf{S},\mathbf{Z},\Sigma(E))\Big]\Big\}\frac{d}{dE}\Sigma(E)^{-2}. \tag{2.55}$$

We first look at the derivative of the bracket $\{\cdots\}$ with respect to Σ^{-2}. In the next few lines the following notation is used for the "Gibbs" average

$$\langle A(\mathbf{x})\rangle_{\text{den}} = \frac{\int d^B\mathbf{x}\, A(\mathbf{x})p_0(\mathbf{x})\theta(\mathbf{x},\mathbf{S},\mathbf{Z},\Sigma(E))}{\int d^B\mathbf{x}\, p_0(\mathbf{x})\theta(\mathbf{x},\mathbf{S},\mathbf{Z},\Sigma(E))}.$$

Using the explicit expression of $\theta(\mathbf{x}, \mathbf{S}, \mathbf{Z}, \Sigma(E))$ we have

$$\begin{aligned}
&\frac{\partial}{\partial \Sigma(E)^{-2}} \mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\log_B \int d^B\mathbf{x}\, p_0(\mathbf{x}) \theta(\mathbf{x}, \mathbf{S}, \mathbf{Z}, \Sigma(E))\Big] \\
&= \frac{\partial}{\partial \Sigma(E)^{-2}} \mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\log_B \int d^B\mathbf{x}\, p_0(\mathbf{x}) e^{-\frac{1}{2}\left(\|\mathbf{x}-\mathbf{S}\|^2 \Sigma(E)^{-2} \frac{\ln B}{\ln 2} - \frac{2\mathbf{Z}\cdot(\mathbf{x}-\mathbf{S})}{\Sigma(E)}\sqrt{\frac{\ln B}{\ln 2}} + \|Z\|^2\right)}\Big] \\
&= -\frac{1}{2\ln 2}\mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\langle \|\mathbf{x}-\mathbf{S}\|^2\rangle_{\rm den}\Big] + \frac{1}{2}\mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\mathbf{z}\cdot\langle \mathbf{x}-\mathbf{S}\rangle_{\rm den}\Big]\frac{\Sigma(E)}{\sqrt{(\ln B)(\ln 2)}} \\
&= -\frac{1}{2\ln 2}\mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\langle \|\mathbf{x}-\mathbf{S}\|^2\rangle_{\rm den}\Big] + \frac{1}{2}\mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\nabla_{\mathbf{Z}}\cdot\langle \mathbf{x}-\mathbf{S}\rangle_{\rm den}\Big]\frac{\Sigma(E)}{\sqrt{(\ln B)(\ln 2)}} \\
&= -\frac{1}{2\ln 2}\Big(\mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\langle \|\mathbf{x}-\mathbf{S}\|^2\rangle_{\rm den}\Big] - \mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\langle \|\mathbf{x}-\mathbf{S}\|^2\rangle_{\rm den}\Big] + \mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\|\langle \mathbf{x}-\mathbf{S}\rangle_{\rm den}\|^2\Big]\Big) \\
&= -\frac{1}{2\ln 2}\mathbb{E}_{\mathbf{S},\mathbf{Z}}\Big[\|\langle \mathbf{x}\rangle_{\rm den} - \mathbf{S}\|^2\Big] \\
&= -\frac{1}{2\ln 2}\mathrm{mmse}(\Sigma(E)).
\end{aligned}$$

We show below that

$$\frac{1}{\Sigma(E)^2} = -\frac{2}{R}\frac{\partial}{\partial E}\mathbb{E}_Z\Big[\int dy\, \phi(y|Z,E)\ln\phi(y|Z,E)\Big] \tag{2.56}$$

so that (2.55) becomes

$$\frac{dF_{\rm un}}{dE} = \Big\{\mathrm{mmse}(\Sigma(E)) - E\Big\}\frac{1}{2\ln 2}\frac{d}{dE}\Sigma(E)^{-2} \tag{2.57}$$

which obviously shows that $dF_{\rm un}/dE = 0$ implies the SE equation $E = T_{\rm un}(E)$. We point out as a side remark that this is the correct "integrating factor" which allows to recover the potential function from the SE equation.

It remains to derive (2.56). We will start from the derivative with respect to E in (2.56) and show that this relation can be transformed into Definition 2.2, namely

$$\frac{1}{\Sigma(E)^2} = \frac{1}{R}\int dp\, \frac{e^{-\frac{p^2}{2(1-E)}}}{\sqrt{2\pi(1-E)}}\int dy f(y|p,E)(\partial_p \ln f(y|p,E))^2 \tag{2.58}$$

where

$$f(y|p,E) = \int dx P_{\rm out}(y|x)\frac{e^{-\frac{(x-p)^2}{2E}}}{\sqrt{2\pi E}}. \tag{2.59}$$

We first note that $\phi(y|z,E) = f(y|z\sqrt{1-E},E)$ so the derivative w.r.t E

on the right hand side of (2.56) becomes

$$\begin{aligned}
&\frac{\partial}{\partial E}\mathbb{E}_Z\Big[\int dy\,\phi(y|Z,E)\ln\phi(y|Z,E)\Big]\\
&=\frac{\partial}{\partial E}\int dz\,\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\int dy\,f(y|z\sqrt{1-E},E)\ln f(y|z\sqrt{1-E},E)\\
&=\int dz\,\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\int dy\,(1+\ln f(y|z\sqrt{1-E},E))\partial_E f(y|z\sqrt{1-E},E). \qquad (2.60)
\end{aligned}$$

An exercise in differentiation of Gaussians shows[9]

$$\partial_E\Big\{\frac{e^{-\frac{(x-z\sqrt{1-E})^2}{2E}}}{\sqrt{2\pi E}}\Big\}=\frac{e^{\frac{z^2}{2}}}{2(1-E)}\partial_z\Big\{e^{-\frac{z^2}{2}}\partial_z\Big\{\frac{e^{-\frac{(x-z\sqrt{1-E})^2}{2E}}}{\sqrt{2\pi E}}\Big\}\Big\}.$$

Thus from (2.59)

$$\partial_E f(y|z\sqrt{1-E},E)=\frac{e^{\frac{z^2}{2}}}{2(1-E)}\partial_z\Big\{e^{-\frac{z^2}{2}}\partial_z f(y|z\sqrt{1-E},E)\Big\}$$

and (2.60) becomes

$$\begin{aligned}
&\frac{\partial}{\partial E}\mathbb{E}_Z\Big[\int dy\,\phi(y|Z,E)\ln\phi(y|Z,E)\Big]\\
&=\frac{1}{2(1-E)}\int dz\int dy(1+\ln f(y|z\sqrt{1-E},E))\partial_z\Big\{\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\partial_z f(y|z\sqrt{1-E},E)\Big\}\\
&=-\frac{1}{2(1-E)}\int dz\,\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\int dy\,\frac{\big(\partial_z f(y|z\sqrt{1-E},E)\big)^2}{f(y|z\sqrt{1-E},E)}\\
&=-\frac{1}{2}\int dz\,\frac{e^{-\frac{z^2}{2(1-E)}}}{\sqrt{2\pi(1-E)}}\int dy\,\frac{\big(\partial_z f(y|z,E)\big)^2}{f(y|z,E)}\\
&=-\frac{1}{2}\int dz\,\frac{e^{-\frac{z^2}{2(1-E)}}}{\sqrt{2\pi(1-E)}}\int dy\,f(y|z,E)\big(\partial_z\ln f(y|z,E)\big)^2.
\end{aligned}$$

This result explicitly shows that (2.56) and (2.58) are equivalent as announced.

[9] We thank Christophe Schülke for pointing out this trick.

2.8.3 Bounds on the Second Derivative of the Potential Function

In this appendix, we provide an upper bound on the second derivative

$$\Big[\frac{\partial^2 F_{\rm co}}{\partial E_r \partial E_{r'}}\Big]_{\hat{\mathbf{E}}} = \Big[\frac{\partial^2}{\partial E_r \partial E_{r'}} \sum_{r=1}^{\Gamma} U_{\rm un}(E_r)\Big]_{\hat{\mathbf{E}}} - \frac{\partial^2}{\partial E_r \partial E_{r'}}\Big[\sum_{c=1}^{\Gamma} S_{\rm un}(\Sigma_c(\mathbf{E}))\Big]_{\hat{\mathbf{E}}}$$
$$= \delta_{r,r'}\Big[\frac{\partial^2 U_{\rm un}(E_r)}{\partial E_r^2}\Big]_{\hat{\mathbf{E}}} - \sum_{c=1}^{\Gamma}\Big[\frac{\partial^2 S_{\rm un}(\Sigma_c(\mathbf{E}))}{\partial E_r \partial E_{r'}}\Big]_{\hat{\mathbf{E}}} \tag{2.61}$$

of the potential function needed in the proof of Lemma 2.6. We first perform the analysis for general memoryless channels satisfying our two Assumptions 2.1, 2.2. We then briefly show how to improve the estimate in the special case of the AWGN because of the non vanishing error floor.

General Channel

Energy term:

Using relation (2.56) of Appendix 2.8.2 one obtains for the first derivative of the energy term

$$\frac{\partial U_{\rm un}(E_r)}{\partial E_r} = -\frac{E_r}{2\ln 2}\frac{\partial \Sigma^{-2}}{\partial E_r}.$$

Differentiating once more

$$\frac{\partial^2 U_{\rm un}(E_r)}{\partial E_r^2} = -\frac{1}{2\ln 2}\frac{\partial \Sigma^{-2}}{\partial E_r} - \frac{E_r}{2\ln 2}\frac{\partial^2 \Sigma^{-2}}{\partial E_r^2}.$$

Using Assumption 2.2 we immediately get

$$\frac{\partial^2 U_{\rm un}(E_r)}{\partial E_r^2} \le \frac{C}{2(\ln 2) R E_r^{\beta}} + \frac{C E_r}{2(\ln 2) R E_r^{\beta}}.$$

Now recall that in the proof of Lemma 2.6 we have $\hat{E}_r > \bar{E}_{\rm f} = E_{\rm f} + \epsilon$ where $E_{\rm f}$ is the (true) error floor and $\epsilon > 0$. Therefore

$$\Big[\frac{\partial^2 U_{\rm un}}{\partial E_r^2}\Big]_{\hat{\mathbf{E}}} \le \frac{C}{2(\ln 2) R (E_{\rm f}+\epsilon)^{\beta}} + \frac{C(E_{\rm f}+\epsilon)}{2(\ln 2) R (E_{\rm f}+\epsilon)^{\beta}}$$
$$\le \frac{C(2+\epsilon)}{2(\ln 2) R (E_{\rm f}+\epsilon)^{\beta}}. \tag{2.62}$$

Of course this is the worse possible bound and is valid all the way up to the left boundary of the *modified* system. As one moves towards the right of the spatially coupled system one could use bigger values for E_r and tigthen the bound. This however is not needed to prove Lemma 2.6 as long as $\epsilon > 0$.

Entropy term:

For the second derivative of the "entropy" term we first apply the chain rule

$$\begin{aligned}\frac{\partial^2 S_{\mathrm{un}}(\Sigma_c(\mathbf{E}))}{\partial E_r \partial E_{r'}} &= \frac{\partial}{\partial E_{r'}}\Big(\frac{\partial S_{\mathrm{un}}}{\partial \Sigma_c^{-2}}\frac{\partial \Sigma_c^{-2}}{\partial E_r}\Big)\\ &= \frac{\partial^2 S_{\mathrm{un}}}{\partial (\Sigma_c^{-2})^2}\frac{\partial \Sigma_c^{-2}}{\partial E_r}\frac{\partial \Sigma_c^{-2}}{\partial E_{r'}} + \frac{\partial S_{\mathrm{un}}}{\partial \Sigma_c^{-2}}\frac{\partial^2 \Sigma_c^{-2}}{\partial E_r \partial E_{r'}}\\ &= J_{r,c}J_{r',c}\frac{\partial^2 S_{\mathrm{un}}}{\partial (\Sigma_c^{-2})^2}\frac{\partial \Sigma^{-2}}{\partial E_r}\frac{\partial \Sigma^{-2}}{\partial E_{r'}} + \delta_{rr'}J_{r,c}\frac{\partial S_{\mathrm{un}}}{\partial \Sigma_c^{-2}}\frac{\partial^2 \Sigma^{-2}}{\partial E_r^2},\end{aligned}$$

where to get the last line we used

$$\frac{\partial \Sigma_c^{-2}}{\partial E_r} = J_{r,c}\frac{\partial \Sigma^{-2}}{\partial E_r}1_{c-w\le r\le c+w}, \qquad \frac{\partial^2 \Sigma_c^{-2}}{\partial E_r \partial E_{r'}} = \delta_{rr'}J_{r,c}\frac{\partial^2 \Sigma^{-2}}{\partial E_r^2}1_{c-w\le r\le c+w}$$

which follow directly from the definition of $\Sigma_c^{-2}(\mathbf{E})$. Recall that by construction $J_{r,c}/\Gamma \le (\bar{g}/\underline{g})(2w+1)^{-1}$. Recall also Assumption 2.2. We thus have

$$\begin{aligned}&\Big|\frac{\partial^2 S_{\mathrm{un}}(\Sigma_c(\mathbf{E}))}{\partial E_r \partial E_{r'}}\Big|\\ &\le \frac{\bar{g}^2}{\underline{g}^2(2w+1)^2}\Big|\frac{\partial^2 S_{\mathrm{un}}}{\partial (\Sigma_c^{-2})^2}\Big|\Big|\frac{\partial \Sigma^{-2}}{\partial E_r}\Big|\Big|\frac{\partial \Sigma^{-2}}{\partial E_{r'}}\Big|1_{c-w\le r\le c+w}1_{c-w\le r'\le c+w}\\ &+ \frac{\delta_{rr'}\bar{g}}{\underline{g}(2w+1)}\Big|\frac{\partial S_{\mathrm{un}}}{\partial \Sigma_c^{-2}}\Big|\Big|\frac{\partial^2 \Sigma^{-2}}{\partial E_r^2}\Big|1_{c-w\le r\le c+w}\\ &\le \frac{\bar{g}^2 C^2}{\underline{g}^2(2w+1)^2 R^2 E_r^\beta E_{r'}^\beta}\Big|\frac{\partial^2 S_{\mathrm{un}}}{\partial (\Sigma_c^{-2})^2}\Big|1_{c-w\le r\le c+w}1_{c-w\le r'\le c+w}\\ &+ \frac{\delta_{rr'}\bar{g}C}{\underline{g}(2w+1)RE^\beta}\Big|\frac{\partial S_{\mathrm{un}}}{\partial \Sigma_c^{-2}}\Big|1_{c-w\le r\le c+w}\end{aligned} \tag{2.63}$$

The next step is to compute and estimate the partial derivatives of S_{un} in this expression. Using Definition 2.11 we find that (this involves differentiating under integral signs which can be justified by the ensuing bounds)

$$\frac{\partial S_{\mathrm{un}}}{\partial \Sigma_c^{-2}} = \sum_{i=2}^{B}\mathbb{E}_{\mathbf{Z}}\Big[\Big(\frac{(Z_i - Z_1)\Sigma_c}{2\sqrt{\ln 2 \ln B}} - \frac{1}{\ln 2}\Big)\frac{e_i}{1+\sum_{j=2}^{B} e_j}\Big], \tag{2.64}$$

and

$$\begin{aligned}&\frac{\partial^2 S_{\mathrm{un}}}{\partial (\Sigma_c^{-2})^2}\\ &= (\ln B)\sum_{i=2}^{B}\mathbb{E}_{\mathbf{Z}}\Big[\Big(\Big(\frac{(Z_i - Z_1)\Sigma_c}{2\sqrt{\ln 2 \ln B}} - \frac{1}{\ln 2}\Big)^2 - \frac{(Z_i - Z_1)\Sigma_c^3}{2\sqrt{\ln 2 \ln B}}\Big)\frac{e_i}{1+\sum_{j=2}^{B} e_j}\\ &- (\ln B)\sum_{i,j=2}^{B}\Big(\frac{(Z_i - Z_1)\Sigma_c}{2\sqrt{\ln 2 \ln B}} - \frac{1}{\ln 2}\Big)\Big(\frac{(Z_j - Z_1)\Sigma_c}{2\sqrt{\ln 2 \ln B}} - \frac{1}{\ln 2}\Big)\frac{e_i e_j}{(1+\sum_{j=2}^{B} e_j)^2}\Big].\end{aligned} \tag{2.65}$$

Since $e_i \geq 0$ we have for $2 \leq i \leq n$

$$\frac{e_i}{1+\sum_{j=2}^{B} e_j} \leq 1,$$

which easily implies the following bounds for (2.64) and (2.65)

$$\begin{aligned}\Big|\frac{\partial S_{\text{un}}}{\partial \Sigma_c^{-2}}\Big| &\leq (B-1)\Big(\frac{\Sigma_c}{\sqrt{\pi \ln 2 \ln B}} + \frac{1}{\ln 2}\Big) \\ &\leq C_1(B) + C_2(B)\Sigma_c, \end{aligned} \tag{2.66}$$

$$\begin{aligned}\Big|\frac{\partial S_{\text{un}}}{\partial \Sigma_c^{-2}}\Big| &\leq (\ln B)(B-1)\Big(\frac{\Sigma_c^2}{2\ln 2 \ln B} + \frac{1}{(\ln 2)^2} + \frac{2\Sigma_c}{\ln 2\sqrt{\pi \ln 2 \ln B}}\Big) \\ &\quad + (\ln B)(B-1)^2\Big(\frac{(\frac{2\sqrt{3}}{\pi}+\frac{1}{3})\Sigma_c^2}{4\ln 2 \ln B} + \frac{1}{(\ln 2)^2} + \frac{2\Sigma_c}{\ln 2\sqrt{\pi \ln 2 \ln B}}\Big) \\ &\quad + (B-1)\frac{\Sigma_c^3}{\sqrt{\pi \ln 2 \ln B}} \\ &\leq C_3(B) + C_4(B)\Sigma_c + C_5(B)\Sigma_c^2 + C_6(B)\Sigma_c^3. \end{aligned} \tag{2.67}$$

where $C_i(B)$, $i = 1, \cdots, 6$ are constants that depend only on B. Furthermore, from the definition of $\Sigma_c(\mathbf{E})$ and Assumption 2.1 we remark that $\Sigma_c(\mathbf{E}) \leq \sup_{E\in[0,1]} \Sigma(E) = \Sigma(1)$. Hence, we can replace Σ_c by $\Sigma(1)$ in the bounds (2.66), (2.67). Then using these two bounds the estimate (2.63) becomes

$$\begin{aligned}&\Big|\frac{\partial^2 S_{\text{un}}(\Sigma_c(\mathbf{E}))}{\partial E_r \partial E_{r'}}\Big| \\ &\leq \frac{\bar{g}^2 C^2}{\underline{g}^2(2w+1)^2 R^2 E_r^\beta E_{r'}^\beta}\Big(C_3(B) + C_4(B)\Sigma(1) + C_5(B)\Sigma^2(1) + C_6(B)\Sigma^3(1)\Big) \\ &\times 1_{c-w\leq r\leq c+w} 1_{c-w\leq r'\leq c+w} \\ &+ \frac{\delta_{rr'}\bar{g}C}{\underline{g}(2w+1)RE_r^\beta}\Big(C_1(B) + C_2(B)\Sigma(1)\Big)1_{c-w\leq r\leq c+w}\end{aligned}$$

Since

$$\begin{cases} 1_{c-w\leq r\leq c+w} 1_{c-w\leq r'\leq c+w} \leq 1_{r-w\leq c\leq r+w} 1_{|r-r'|\leq 2w+1} \\ 1_{c-w\leq r\leq c+w} = 1_{r-w\leq c\leq r+w}, \end{cases}$$

when we sum over c we get

$$\begin{aligned}&\sum_{c=1}^{\Gamma} \frac{\partial^2 S_{\text{un}}(\Sigma_c(\mathbf{E}))}{\partial E_r \partial E_{r'}} \\ &\leq \frac{\bar{g}^2 C^2}{\underline{g}^2 R^2 (2w+1) E_r^\beta E_{r'}^\beta}\Big(C_3(B) + C_4(B)\Sigma(1) + C_5(B)\Sigma^2(1) + C_6(B)\Sigma^3(1)\Big) \\ &\times 1_{|r-r'|\leq 2w+1} + \frac{\delta_{rr'}\bar{g}C}{\underline{g}RE_r^\beta}\Big(C_1(B) + C_2(B)\Sigma(1)\Big)\end{aligned}$$

Finally, using again $\hat{E}_r \geq \bar{E}_\mathrm{f} = E_\mathrm{f} + \epsilon$ we obtain

$$\begin{aligned}
&\Big[\sum_{c=1}^{\Gamma} \frac{\partial^2 S_{\mathrm{un}}(\Sigma_c(\mathbf{E}))}{\partial E_r \partial E_{r'}}\Big]_{\hat{\mathbf{E}}} \\
&\leq \frac{\bar{g}^2 C^2}{\underline{g}^2 R^2 (2w+1)(E_\mathrm{f}+\epsilon)^{2\beta}} \\
&\times \Big(C_3(B) + C_4(B)\Sigma(1) + C_5(B)\Sigma^2(1) + C_6(B)\Sigma^3(1)\Big) 1_{|r-r'|\leq 2w+1} \\
&+ \frac{\delta_{rr'}\bar{g}C}{\underline{g}R(E_\mathrm{f}+\epsilon)^{\beta}}\Big(C_1(B) + C_2(B)\Sigma(1)\Big)
\end{aligned} \tag{2.68}$$

Final bound:

Putting (2.61), (2.62) and (2.68) together the triangle inequality implies the important result

$$\Big[\frac{\partial^2 F_{\mathrm{co}}}{\partial E_r \partial E_{r'}}\Big]_{\hat{\mathbf{E}}} \leq \delta_{r,r'} \frac{K_1(B,\bar{g},\underline{g})}{(E_\mathrm{f}+\epsilon)R} + 1_{|r-r'|\leq 2w+1} \frac{K_2(B,\bar{g},\underline{g})}{(E_\mathrm{f}+\epsilon)^{2\beta} R(2w+1)} \tag{2.69}$$

for some finite positive $K_1(B,\bar{g},\underline{g})$ and $K_2(B,\bar{g},\underline{g})$ independent of w and Γ.

AWGN Channel

For the AWGN channel we have an explicit expression for the effective noise, $\Sigma(E)^2 = (\mathrm{snr}^{-1} + E)R$ which implies

$$\begin{cases}
\Sigma^2(E_r) & \leq & R(\mathrm{snr}^{-1}+1) \\
\frac{\partial \Sigma^{-2}}{\partial E_r} & \leq & \frac{\mathrm{snr}^2}{R} \\
\frac{\partial^2 \Sigma^{-2}}{\partial E_r^2} & \leq & \frac{\mathrm{snr}^3}{R}.
\end{cases} \tag{2.70}$$

Then using these bounds at the appropriate places in the previous analysis we get

$$\Big[\frac{\partial^2 F_{\mathrm{co}}}{\partial E_r \partial E_{r'}}\Big]_{\hat{\mathbf{E}}} \leq \delta_{r,r'} K_1'(B,\bar{g},\underline{g})\mathrm{snr}^2 + 1_{|r-r'|\leq 2w+1} \frac{K_2'(B,\bar{g},\underline{g})\mathrm{snr}^4}{2w+1} \tag{2.71}$$

for new constants $K_1'(B,\bar{g},\underline{g})$, $K_2'(B,\bar{g},\underline{g})$ (independent of w, Γ). We can see that the qualitative behaviour of the bound when $\mathrm{snr} \to +\infty$ is the same as in the case of vanishing error floor $E_\mathrm{f} = 0$ and $\epsilon \to 0$.

2.8.4 Potential Function and Replica Calculation

The potential functions of the uncoupled and coupled systems, used in this chapter, can be viewed as a mathematical tool and we are not really concerned how they are found. However in practice it is important to have a more or less systematic method which allows to write down "good" potential functions. There are essentially two ways. One is to "integrate" the SE equations as

done in [82] by using an appropriate "integrating factor". With this method there is some amount of guess involved. For example in the present problem it is not entirely obvious that the correct integrating factor is directly related to the Fisher information (as equation (2.57) in Appendix 2.8.2 shows). The other way is to perform a formal and brute force replica or cavity calculation of the free energy which is then given as a variational expression involving the potential function. The disadvantage of such a calculation is that it is painful and maybe also that it is formal, but the advantage is that it is quite systematic. For completeness we give the replica calculation. We stress that the results of the chapter do *not* rest on this formal calculation and the reader can entirely skip it.

We treat the prototypical case of a spatially coupled compressed-sensing like system where the signal has *scalar* components x_i, $i = 1, \cdots, N$ i.i.d. distributed according to a general prior $p_0(x)$. The calculation is exactly the same for signals whose components are B-dimensional with arbitray priors and sparse superposition codes fall in this class. The integration symbol $\mathcal{D}v$ is used for $dv\, e^{-\frac{v^2}{2}}$.

The spatially coupled matrix is made of $\Gamma \times \Gamma$ blocks, each with N/Γ columns and $\alpha N/\Gamma$ rows for the blocks part of the r^{th} block-row. The entries inside the block (r, c) are i.i.d. with distribution $\mathcal{N}(0, J_{r,c}\Gamma/N)$. Furthermore, we enforce the per block-row variance normalization $\sum_{c=1}^{\Gamma} J_{r,c} = 1 \ \forall \ r$. We use the notation $\mathbf{x}^0$ for the signal and define $z_\mu^a := \sum_{c=1}^{\Gamma} \sum_{i \in c}^{N/\Gamma} F_{\mu i} x_i^a$ where the matrix structure is made explicit.

The posterior distribution is given by the Bayes rule

$$P(\mathbf{x}|\mathbf{y}) = Z(\mathbf{y})^{-1} \prod_{i=1}^{N} p_0(x_i) \prod_{\mu=1}^{M} P_{\text{out}}(y_\mu|z_\mu)$$

where $Z(\mathbf{y}) = P(\mathbf{y})$ is the observation dependent normalization, or partition function. The (coupled) free energy F_{co} will be calculated using the replica trick in one of its many incarnations

$$F_{\text{co}} := -\lim_{N\to\infty} \lim_{n\to 0} \frac{\partial}{\partial n} \frac{\ln(\mathbb{E}[Z(\mathbf{y})^n])}{N}, \tag{2.72}$$

where $\mathbb{E}$ denotes expectation with respect to the observation $\mathbf{y}(\mathbf{F})$ which depend on the measurement matrix realization (that will be always implicit). We thus need to compute the n^{th} moment of the partition function. For the moment, we consider $n \in \mathbb{N}$ despite that we will let $n \to 0$ at the end.

$Z(\mathbf{y})^n$ can be interpreted as the partition function of n i.i.d. systems, the

replicas $a = 1, \cdots, n$, each generated independently from the posterior $P(\mathbf{x}|\mathbf{y})$

$$Z(\mathbf{y})^n = \int \prod_{a=1}^{n} \left[d\mathbf{x}^a \prod_{i=1}^{N} p_0(x_i^a) \prod_{\mu=1}^{M} P_{\text{out}}(y_\mu|z_\mu^a) \right], \tag{2.73}$$

$$\begin{aligned} \mathbb{E}[Z(\mathbf{y})^n] &= \mathbb{E}_{\mathbf{F}} \int d\mathbf{y} Z(\mathbf{y})^n P(\mathbf{y}) = \mathbb{E}_{\mathbf{F}} \int d\mathbf{y} Z(\mathbf{y})^{n+1} \\ &= \mathbb{E}_{\mathbf{F}} \int d\mathbf{y} \prod_{a=0}^{n} \left[d\mathbf{x}^a \prod_{i=1}^{N} p_0(x_i^a) \prod_{\mu=1}^{M} P_{\text{out}}(y_\mu|z_\mu^a) \right], \end{aligned} \tag{2.74}$$

where the last equality is implied by $P(\mathbf{y}) = Z(\mathbf{y})$. This last point is valid only in the Bayes optimal setting and is known to induce a remarkable set of consequences, among which the correctness of the replica symmetric predictions.

The $\mathbf{F}$ and $\mathbf{x}^a$ random variables being i.i.d., we can treat z_μ^a as a Gaussian random variable by the central limit theorem. Let us compute their distribution. As $\mathbf{F}$ has zero mean, z_μ^a has zero mean also. Its covariance matrix $\tilde{\boldsymbol{q}}_{r_\mu}$ depends on the block-row index $r_\mu \in \{1, \cdots, \Gamma\}$ to which the μ^{th} measurement index belongs. Similarly, $c_i \in \{1, \ldots, \Gamma\}$ is the block-column index to which the i^{th} column belongs. We have

$$\tilde{q}_{r_\mu}^{ab} = \mathbb{E}_{\mathbf{F}}[z_\mu^a z_\mu^b] = \sum_{c,c'=1}^{\Gamma,\Gamma} \sum_{i\in c, j\in c'}^{N/\Gamma, N/\Gamma} \mathbb{E}_{\mathbf{F}}[F_{\mu i} F_{\mu j}] x_i^a x_j^b = \sum_{c}^{\Gamma} \frac{J_{r_\mu, c}}{N} \sum_{i\in c}^{N/\Gamma} x_i^a x_i^b, \tag{2.75}$$

because $\mathbb{E}_{\mathbf{F}}[F_{\mu i} F_{\mu j}] = \delta_{ij} J_{r_\mu, c_i}/N$ in the present spatial coupling construction. We introduce the macroscopic replica overlap matrix, that takes into account the block structure in the signal induced by the matrix structure. Let

$$q_c^{ab} := \frac{\Gamma}{N} \sum_{i\in c}^{N/\Gamma} x_i^a x_i^b \;\forall\; a, b \in \{0, \cdots, n\}. \tag{2.76}$$

Then (2.75) becomes $\tilde{q}_r^{ab} = \sum_{c=1}^{\Gamma} J_{r,c} q_c^{ab}$.

We now introduce the replica symmetric ansatz. According to this ansatz, the overlap should not depend on the replica index $q_c^{ab} = q_c \;\forall\; a \neq b$, $q_c^{aa} = Q_c \;\forall\; a$. This implies

$$\tilde{q}_r^{ab} = \tilde{q}_r = \sum_{c=1}^{\Gamma} J_{r,c} q_c \;\forall\; a \neq b,\; \tilde{q}_r^{aa} = \tilde{Q}_r = \sum_{c=1}^{\Gamma} J_{r,c} Q_c \;\forall\; a. \tag{2.77}$$

Using the variance normalization $Q_c = \tilde{Q}_r$. Then, one can show that in Bayes optimal inference we have furthermore $Q_c = \tilde{Q}_r = \mathbb{E}[S^2] \;\forall\; c, r \in \{1, \cdots, \Gamma\}$, where $\mathbb{E}[s^2] = \int ds p_0(s) s^2$. In the physics litterature this is often called a "Nishimori identity".

Thus the self overlap Q_c is fixed and the condition (2.76) for $a = b$ does not need to be enforced. On the other hand, the cross overlap for $a \neq b$

is unknown and so we must keep $\{q_c\}$ as variables. Define a distribution of replicated variables at fixed overlap matrices $\{\boldsymbol{q}_c, c \in \{1, \cdots, \Gamma\}\}$

$$P(\{\mathbf{x}^a\}|\{\boldsymbol{q}_c\}) := \frac{1}{\Xi(\{\boldsymbol{q}_c\})} \prod_{a=0}^{n} \left[\prod_{i=1}^{N} p_0(x_i^a) \prod_{c=1}^{\Gamma} \prod_{b<a}^{n} \delta\left(\frac{1}{2i\pi} \left[\frac{N}{\Gamma} q_c^{ab} - \sum_{i \in c}^{N/\Gamma} x_i^a x_i^b \right] \right) \right], \tag{2.78}$$

where $\Xi(\{\boldsymbol{q}_c\})$ is the associated normalization. The role of the $2i\pi$ appearing in the delta function is purely formal and will become clear later on. Plugging this expression inside (2.74) we get

$$\mathbb{E}[Z(\mathbf{y})^n] = \mathbb{E}_{\mathbf{F}} \int d\mathbf{y} \prod_{c=1}^{\Gamma} d\boldsymbol{q}_c P(\{\mathbf{x}^a\}|\{\boldsymbol{q}_c\}) \Xi(\{\boldsymbol{q}_c\}) \prod_{a=0}^{n} \left[d\mathbf{x}^a \prod_{r=1}^{\Gamma} \prod_{\mu \in r}^{\alpha N/\Gamma} P_{\text{out}}(y_\mu|z_\mu^a) \right] \tag{2.79}$$

$$= \int \prod_{c=1}^{\Gamma} d\boldsymbol{q}_c \Xi(\{\boldsymbol{q}_c\}) \int d\mathbf{y} P(\{\mathbf{z}^a\}|\{\boldsymbol{q}_c\}) \prod_{a=0}^{n} \left[d\mathbf{z}^a \prod_{r=1}^{\Gamma} \prod_{\mu \in r}^{\alpha N/\Gamma} P_{\text{out}}(y_\mu|z_\mu^a) \right]. \tag{2.80}$$

The second equality is obtained after noticing that the integrand in (2.79) depends on $\{x_i^a\}$ only through $\{z_\mu^a\}$, this allows to replace the integration on $\{x_i^a\}$ by an integration on $\{z_\mu^a\}$. As already explained, by the central limit theorem

$$\begin{aligned} P(\{\mathbf{z}^a\}|\{\boldsymbol{q}_c\}) &= \prod_{\mu=1}^{M} \mathcal{N}(\mathbf{z}_\mu|0, \tilde{\boldsymbol{q}}_{r_\mu}) = \prod_{r=1}^{\Gamma} \prod_{\mu \in r}^{\alpha N/\Gamma} \mathcal{N}(\mathbf{z}_\mu|0, \tilde{\boldsymbol{q}}_r) \\ &= \prod_{r=1}^{\Gamma} \left[(2\pi)^{n+1} \det(\tilde{\boldsymbol{q}}_r) \right]^{-\frac{\alpha N}{2\Gamma}} \prod_{\mu \in r}^{\alpha N/\Gamma} e^{-\frac{1}{2} \sum_{a,b=0}^{n,n} z_\mu^a [\tilde{\boldsymbol{q}}_r^{-1}]_{ab} z_\mu^b}. \end{aligned} \tag{2.81}$$

This is a product of multivariate centered Gaussian distributions, where $\mathbf{z}_\mu := [z_\mu^a, a \in \{0, \ldots, n\}]$, $\mathbf{z}^a := [z_\mu^a, \mu \in \{1, \ldots, M\}]$. Recall $\tilde{\boldsymbol{q}}_r$ is a function of $\{\boldsymbol{q}_c\}$. Let

$$\mathbb{E}[Z(\mathbf{y})^n] = \int \prod_{c=1}^{\Gamma} d\boldsymbol{q}_c \exp\Big[N \Big(f(\{\boldsymbol{q}_c\}) + g(\{\boldsymbol{q}_c\}) \Big) \Big], \tag{2.82}$$

$$f(\{\boldsymbol{q}_c\}) := \frac{1}{N} \ln \Big[\Xi(\{\boldsymbol{q}_c\}) \Big], \tag{2.83}$$

$$g(\{\boldsymbol{q}_c\}) := \frac{1}{N} \ln \left[\int d\mathbf{y} P(\{\mathbf{z}^a\}|\{\boldsymbol{q}_c\}) \prod_{a=0}^{n} \left[d\mathbf{z}^a \prod_{r=1}^{\Gamma} \prod_{\mu \in r}^{\alpha N/\Gamma} P_{\text{out}}(y_\mu|z_\mu^a) \right] \right]. \tag{2.84}$$

Now we perform a saddle point estimation. This requires to take the limit $N \to \infty$ limit before letting $n \to 0$, and we assume without justification that

the final result does not depend on the order of limits n and N. This gives for the free energy, using (2.72)

$$\begin{aligned} F_{\text{co}} &= -\lim_{n\to 0}\frac{\partial}{\partial n}\underset{\{q_c\}}{\text{extr}}\Big(f(\{\boldsymbol{q}_c\})+g(\{\boldsymbol{q}_c\})\Big) \\ &= -\underset{\{q_c\}}{\text{extr}}\left(\lim_{n\to 0}\frac{\partial f(\{\boldsymbol{q}_c\})}{\partial n}+\lim_{n\to 0}\frac{\partial g(\{\boldsymbol{q}_c\})}{\partial n}\right). \end{aligned} \tag{2.85}$$

Now the replica symmetric ansatz allows to simplify g since $P(\{\mathbf{z}^a\}|\{\boldsymbol{q}_c\})$ becomes

$$P(\{\mathbf{z}^a\}|\{\boldsymbol{q}_c\}) = \prod_{r=1}^{\Gamma}\left[(2\pi)^{n+1}\det(\tilde{\boldsymbol{q}}_r)\right]^{-\frac{\alpha N}{2\Gamma}}\prod_{\mu\in r}^{\alpha N/\Gamma} e^{-\frac{C_{1,r}}{2}\sum_{a=0}^{n}(z_\mu^a)^2-\frac{C_{2,r}}{2}\sum_{a=0,b\neq a}^{n,n} z_\mu^a z_\mu^b}, \tag{2.86}$$

where $C_{1,r}$ and $C_{2,r}$ depend on $\tilde{q}_r$ and $\mathbb{E}[s^2]$ as they are obtained from the matrix inversion $\tilde{\boldsymbol{q}}_r^{-1}$. Thanks to the simple structure of $\tilde{\boldsymbol{q}}_r$ under the replica symmetric ansatz, one can easily show that

$$C_{1,r} = \frac{\mathbb{E}[s^2]+(n-2)\tilde{q}_r}{\mathbb{E}[s^2](\mathbb{E}[s^2]+(n-2)\tilde{q}_r)+(1-n)\tilde{q}_r^2} \underset{n\to 0}{\to} \frac{\mathbb{E}[s^2]-2\tilde{q}_r}{(\mathbb{E}[s^2]-\tilde{q}_r)^2}, \tag{2.87}$$

$$C_{2,r} = -\frac{\tilde{q}_r}{\mathbb{E}[s^2](\mathbb{E}[s^2]+(n-2)\tilde{q}_r)+(1-n)\tilde{q}_r^2} \underset{n\to 0}{\to} -\frac{\tilde{q}_r}{(\mathbb{E}[s^2]-\tilde{q}_r)^2}. \tag{2.88}$$

The replicated variables $\{\mathbf{z}^a\}$ are correlated through $P(\{\mathbf{z}^a\}|\{\boldsymbol{q}_c\})$. In order to simplify g, we decorrelate them by linearizing the exponent of $P(\{\mathbf{z}^a\}|\{\boldsymbol{q}_c\})$ using the Gaussian transformation formula for a given $K>0$

$$e^{\frac{K}{2}\sum_{a=0,b\neq a}^{n,n} z_\mu^a z_\mu^b} = \int \mathcal{D}\xi_\mu\, e^{\xi_\mu\sqrt{K}\sum_{a=0}^{n} z_\mu^a - \frac{K}{2}\sum_{a=0}^{n}(z_\mu^a)^2}, \tag{2.89}$$

i.e. the previously correlated $\{z_\mu^a, a\in\{0,\dots,n\}\}$ are now i.i.d. Gaussian variables, but that all interact with a common random Gaussian effective field ξ_μ. Using this with $K=-C_{2,r}$ as we know that $C_{2,r}\leq 0$, the integration in g can now be performed starting from (2.84)

$$\begin{aligned} &g(\{\boldsymbol{q}_c\}) \\ &= \frac{1}{N}\ln\left[\prod_{r=1}^{\Gamma}\prod_{\mu\in r}^{\alpha N/\Gamma}\int \mathcal{D}\xi_\mu dy_\mu\left(\int dz_\mu \mathcal{N}\Big(z_\mu\Big|m(\xi_\mu,\tilde{q}_r),V(\mathbb{E}[s^2],\tilde{q}_r)\Big)P_{\text{out}}(y_\mu|z_\mu)\right)^{n+1}\right] \\ &= \frac{1}{\Gamma}\sum_{r=1}^{\Gamma}\alpha\ln\left[\int \mathcal{D}\xi dy\left(\int dz\mathcal{N}\Big(z\Big|m(\xi,\tilde{q}_r),V(\mathbb{E}[s^2],\tilde{q}_r)\Big)P_{\text{out}}(y|z)\right)^{n+1}\right] \\ &= \frac{1}{\Gamma}\sum_{r=1}^{\Gamma}\alpha\ln\left[\int \mathcal{D}\xi dy\left(\int \mathcal{D}z P_{\text{out}}\Big(y\Big|m(\xi,\tilde{q}_r)+z\sqrt{V(\mathbb{E}[s^2],\tilde{q}_r)}\Big)\right)^{n+1}\right]. \end{aligned} \tag{2.90}$$

As assumed, g does no depend on N. Let us compute $m(\xi_\mu, \tilde{q}_r), V(\mathbb{E}[s^2], \tilde{q}_r)$. Combining (2.86) with (2.89), we get that for a $\mu \in r$ and up to a normalization, $z_\mu \sim \exp(-z_\mu^2(C_{1,r}-C_{2,r})/2+z_\mu\xi_\mu\sqrt{-C_{2,r}})$ which becomes using the $n \to 0$ limit of (2.87), (2.88) $z_\mu \sim \mathcal{N}(z_\mu|\xi_\mu\sqrt{\tilde{q}_r}, \mathbb{E}[s^2]-\tilde{q}_r)\exp(\xi_\mu^2\tilde{q}_r/[2(\mathbb{E}[s^2]-\tilde{q}_r)])$. Normalizing $P(z_\mu)$, the term $\exp(\xi_\mu^2\tilde{q}_r/[2(\mathbb{E}[s^2]-\tilde{q}_r)])$ disappears being independent of z_μ and thus $P(z_\mu) = \mathcal{N}(z_\mu|\xi_\mu\sqrt{\tilde{q}_r}, \mathbb{E}[s^2]-\tilde{q}_r)$. Thus $m(\xi_\mu, \tilde{q}_r) = \xi_\mu\sqrt{\tilde{q}_r}$, $V(\mathbb{E}[s^2], \tilde{q}_r) = \mathbb{E}[s^2]-\tilde{q}_r$. Now performing the $\lim_{n\to 0}\partial_n$ operation and using the identity

$$\lim_{n\to 0}\frac{\partial}{\partial n}\ln\left[\int duX(u)^{n+1}\right] = \frac{\int duX(u)\ln(X(u))}{\int dvX(v)}, \tag{2.91}$$

we directly obtain

$$\begin{aligned}\lim_{n\to 0}\frac{\partial g(\{\boldsymbol{q}_c\})}{\partial n} =& \sum_{r=1}^{\Gamma}\frac{\alpha}{\Gamma}\int\left\{\mathcal{D}\xi dy\mathcal{D}\hat{z}\, P_{\text{out}}\Big(y\Big|\xi\sqrt{\tilde{q}_r}+\hat{z}\sqrt{\mathbb{E}[s^2]-\tilde{q}_r}\Big)\right.\\ &\left.\times\ln\left[\int\mathcal{D}z\, P_{\text{out}}\Big(y\Big|\xi\sqrt{\tilde{q}_r}+z\sqrt{\mathbb{E}[s^2]-\tilde{q}_r}\Big)\right]\right\},\end{aligned} \tag{2.92}$$

where we used the normalization $\int dydu\, P_{\text{out}}(y|u)\mathcal{N}(u|a,b) = 1$, such that the denominator in (2.91) sums to one. Let us now deal with $f(\{\boldsymbol{q}_c\})$. We will use the following representation of the delta function $\delta(x) = \int_{\mathbb{R}} d\hat{q}\exp(2i\pi\hat{q}x) \Leftrightarrow \delta(x/(2i\pi)) = \int_{\mathbb{R}} d\hat{q}\exp(\hat{q}x)$ where $\hat{q}$ can be interpreted as an auxiliary external field.[10]

We assume the replica symmetric ansatz for the auxillary fields similarly as for the overlap matrix $\hat{q}_c^{ab} = -\hat{q}_c\ \forall\ a, b \neq a$. The minus sign is just introduced for convenience. Using again the Gaussian transformation formula, we get

$$\begin{aligned}&f(\{\boldsymbol{q}_c\})\\ &=\frac{1}{N}\ln\left[\prod_{c=1}^{\Gamma}\int d\hat{q}_c\prod_{a=0}^{n}\prod_{k\in c}^{N/\Gamma}\left[p_0(x_k^a)dx_k^a\right]e^{-\hat{q}_c\sum_{a=0,b<a}^{n,n}\left(\frac{N}{\Gamma}q_c-\sum_{k\in c}^{N/\Gamma}x_k^ax_k^b\right)}\right]\\ &=\frac{1}{N}\sum_{c=1}^{\Gamma}\ln\left[\int d\hat{q}_c e^{-\frac{N(n+1)n}{2\Gamma}\hat{q}_cq_c}\left(\int\prod_{a=0}^{n}\left[dx^ap_0(x^a)\right]e^{\frac{\hat{q}_c}{2}\sum_{a=0,b\neq a}^{n,n}x^ax^b}\right)^{\frac{N}{\Gamma}}\right]\\ &=\frac{1}{N}\sum_{c=1}^{\Gamma}\ln\left[\int d\hat{q}_c e^{-\frac{N(n+1)n}{2\Gamma}\hat{q}_cq_c}\left(\int\mathcal{D}\xi\prod_{a=0}^{n}\left[dx^ap_0(x^a)\right]e^{-\frac{\hat{q}_c}{2}\sum_a(x^a)^2+\xi\sqrt{\hat{q}_c}\sum_a x^a}\right)^{\frac{N}{\Gamma}}\right]\\ &=\frac{1}{\Gamma}\sum_{c=1}^{\Gamma}\underset{\hat{q}_c}{\text{extr}}\left(-\frac{(n+1)n}{2}\hat{q}_cq_c+\ln\left[\int\mathcal{D}\xi\left(\int dxp_0(x)e^{-\frac{\hat{q}_c}{2}x^2+\xi\sqrt{\hat{q}_c}x}\right)^{n+1}\right]\right),\end{aligned} \tag{2.93}$$

[10] We now understand that the presence of the $2i\pi$ in (2.78) is thus just a trick to make the integral real.

where we have assumed that we can treat $\hat{q}_c$ as a positive variable for the Gaussian transformation transform. This will be verified a posteriori at the end of the computation. The saddle point method employed for the estimation of the integral over the auxiliary fields is justified similarly as before, as the $N \to \infty$ as already been assumed. Finally, using again (2.91), we obtain

$$\lim_{n\to 0} \frac{\partial f(\{\boldsymbol{q}_c\})}{\partial n} = \frac{1}{\Gamma} \sum_{c=1}^{\Gamma} \underset{\hat{q}_c}{\text{extr}} \Bigg(-\frac{\hat{q}_c q_c}{2} + \int \Big\{ \mathcal{D}\xi ds p_0(s) e^{-\frac{\hat{q}_c}{2}s^2 + \xi\sqrt{\hat{q}_c}s} \times \ln\Big[\int dx p_0(x) e^{-\frac{\hat{q}_c}{2}x^2 + \xi\sqrt{\hat{q}_c}x} \Big] \Big\} \Bigg). \tag{2.94}$$

Using (2.85), (2.92) and this last expression, we get a first version of the replica formula for the free energy

$$\begin{aligned} \Gamma F_{\text{co}} = \underset{\{q_c,\hat{q}_c\}}{\text{extr}} \Bigg(& -\sum_{r=1}^{\Gamma} \alpha \int \Big\{ \mathcal{D}\xi dy \mathcal{D}\hat{z}\, P_{\text{out}}\Big(y\Big|\xi\sqrt{\tilde{q}_r} + \hat{z}\sqrt{\mathbb{E}[s^2] - \tilde{q}_r}\Big) \\ & \times \ln\Big[\int \mathcal{D}z\, P_{\text{out}}\Big(y\Big|\xi\sqrt{\tilde{q}_r} + z\sqrt{\mathbb{E}[s^2] - \tilde{q}_r}\Big)\Big] \Big\} \\ & + \sum_{c=1}^{\Gamma} \Bigg(\frac{\hat{q}_c q_c}{2} - \int \mathcal{D}\xi ds p_0(s) e^{-\frac{\hat{q}_c}{2}s^2 + \xi\sqrt{\hat{q}_c}s} \ln\Big[\int dx p_0(x) e^{-\frac{\hat{q}_c}{2}x^2 + \xi\sqrt{\hat{q}_c}x} \Big] \Bigg) \Bigg). \end{aligned} \tag{2.95}$$

Recall that in this expression $\tilde{q}_r = \sum_{c=1}^{\Gamma} J_{r,c} q_c$.

To make contact with the potential function introduced in this chapter we still have to partially solve the extremization problem and reduce (2.95) to a variational problem over one variable. Differentiating the function of $\{q_c, \hat{q}_c\}$ in (2.95) with respect to q_c and setting the derivative to zero we find

$$\hat{q}_c = 2 \sum_{r=1}^{\Gamma} J_{r,c} \alpha\, \partial_{\tilde{q}_r} \Bigg(\int dy \mathcal{D}t f(y|t\sqrt{\tilde{q}_r}, \mathbb{E}[s^2] - \tilde{q}_r) \ln\Big[f(y|t\sqrt{\tilde{q}_r}, \mathbb{E}[s^2] - \tilde{q}_r) \Big] \Bigg), \tag{2.96}$$

where

$$f(y|\mu, \sigma^2) := \int dx \mathcal{N}(x|\mu, \sigma^2) P_{\text{out}}(y|x) = \int \mathcal{D}x P_{\text{out}}(y|x\sigma + \mu). \tag{2.97}$$

One can show the following identity (this has already been shown and used in Appendix 2.8.2)

$$\partial_{\tilde{q}_r} f(y|t\sqrt{\tilde{q}_r}, \mathbb{E}[s^2] - \tilde{q}_r) = \frac{e^{\frac{t^2}{2}}}{2\tilde{q}_r}\, \partial_t \Bigg(e^{-\frac{t^2}{2}}\, \partial_t \Big(f(y|t\sqrt{\tilde{q}_r}, \mathbb{E}[s^2] - \tilde{q}_r) \Big) \Bigg). \tag{2.98}$$

Hence, (2.96) can be rewritten as

$$\begin{aligned}
\hat{q}_c &= 2\sum_{r=1}^{\Gamma} J_{r,c}\alpha \int \Big\{ dy\mathcal{D}t\Big(1+\ln\Big[f_{\text{out}}(y|t\sqrt{\tilde{q}_r},\mathbb{E}[s^2]-\tilde{q}_r)\Big]\Big) \\
&\times \partial_{\tilde{q}_r} f_{\text{out}}(y|t\sqrt{\tilde{q}_r},\mathbb{E}[s^2]-\tilde{q}_r)\Big\} \\
&= -\sum_{r=1}^{\Gamma} J_{r,c}\frac{\alpha}{\tilde{q}_r}\int \Big\{ dydt\frac{1}{\sqrt{2\pi}}\Big(1+\ln\Big[f_{\text{out}}(y|t\sqrt{\tilde{q}_r},\mathbb{E}[s^2]-\tilde{q}_r)\Big]\Big) \\
&\times \partial_t\Big(e^{\frac{-t^2}{2}}\partial_t\big(f_{\text{out}}(y|t\sqrt{\tilde{q}_r},\mathbb{E}[s^2]-\tilde{q}_r)\big)\Big)\Big\} \\
&= \sum_{r=1}^{\Gamma} J_{r,c}\frac{\alpha}{\tilde{q}_r}\int dy\mathcal{D}t\frac{\Big(\partial_t f_{\text{out}}(y|t\sqrt{\tilde{q}_r},\mathbb{E}[s^2]-\tilde{q}_r)\Big)^2}{f_{\text{out}}(y|t\sqrt{\tilde{q}_r},\mathbb{E}[s^2]-\tilde{q}_r)} \\
&= \sum_{r=1}^{\Gamma} J_{r,c}\alpha \int dydpdz\frac{\exp\big(-\frac{p^2}{2\tilde{q}_r}\big)}{\sqrt{2\pi\tilde{q}_r}} f(y|p,\mathbb{E}[s^2]-\tilde{q}_r)\big(\partial_p \ln f(y|p,\mathbb{E}[s^2]-\tilde{q}_r)\big)^2 \\
&= \sum_{r=1}^{\Gamma} J_{r,c}\alpha\mathbb{E}_{p|\tilde{q}_r}[\mathcal{F}\big(p|\mathbb{E}(s^2)-\tilde{q}_r\big)]. \qquad (2.99)
\end{aligned}$$

The final step consists in replacing the stationarity condition (2.99) in (2.95). First we remark

$$\begin{aligned}
\sum_{c=1}^{\Gamma}\frac{\hat{q}_c q_c}{2} &= \frac{1}{2}\sum_{r=1}^{\Gamma}\sum_{c=1}^{\Gamma} J_{r,c}q_c\alpha\mathbb{E}_{p|\tilde{q}_r}[\mathcal{F}\big(p|\mathbb{E}(s^2)-\tilde{q}_r\big)] \\
&= \frac{1}{2}\sum_{r=1}^{\Gamma}\tilde{q}_r\alpha\mathbb{E}_{p|\tilde{q}_r}[\mathcal{F}\big(p|\mathbb{E}(s^2)-\tilde{q}_r\big)] \\
&= \frac{1}{2}\sum_{r=1}^{\Gamma}\tilde{q}_r\Sigma^{-2}(\mathbb{E}(s^2)-\tilde{q}_r) \qquad (2.100)
\end{aligned}$$

where in the last line we have set

$$\Sigma^{-2}(\mathbb{E}(s^2)-\tilde{q}_r) = \alpha\mathbb{E}_{p|\tilde{q}_r}[\mathcal{F}\big(p|\mathbb{E}(s^2)-\tilde{q}_r\big)]. \qquad (2.101)$$

We also set

$$\Sigma_c^{-2}(\{\tilde{q}_r\}) = \sum_{r=1}^{\Gamma} J_{r,c}\alpha\mathbb{E}_{p|\tilde{q}_r}[\mathcal{F}\big(p|\mathbb{E}(s^2)-\tilde{q}_r\big)] \qquad (2.102)$$

so that $\hat{q}_c = \Sigma_c^{-2}(\{\tilde{q}_r\})$. Then (2.95) becomes

$$\begin{aligned}
&\operatorname*{extr}_{\{\tilde{q}_r\}}\Bigg(-\sum_{r=1}^{\Gamma}\Big(\alpha\int\Big\{\mathcal{D}\xi dy\mathcal{D}\hat{z}\,P_{\rm out}\Big(y\Big|\xi\sqrt{\tilde{q}_r}+\hat{z}\sqrt{\mathbb{E}[s^2]-\tilde{q}_r}\Big)\\
&\times\ln\Big[\int\mathcal{D}z\,P_{\rm out}\Big(y\Big|\xi\sqrt{\tilde{q}_r}+z\sqrt{\mathbb{E}[s^2]-\tilde{q}_r}\Big)\Big]\Big\}+\frac{1}{2}\tilde{q}_r\Sigma^{-2}(\tilde{q}_r)\Big)\\
&-\sum_{c=1}^{\Gamma}\Big(\int\Big\{\mathcal{D}\xi ds p_0(s)e^{-\frac{\Sigma_c^{-2}(\{\tilde{q}_r\})}{2}s^2+\xi\sqrt{\Sigma_c^{-2}(\{\tilde{q}_r\})}s}\\
&\times\ln\Big[\int dx p_0(x)e^{-\frac{\hat{q}_c}{2}x^2+\xi\sqrt{\Sigma_c^{-2}(\{\tilde{q}_r\})}x}\Big]\Big)\Big\}\Bigg).
\end{aligned}\tag{2.103}$$

The (courageous) reader can now compare with Definitions (2.14) and (2.11). The sum over r yields an "internal energy" contribution $\sum_r U_{\rm un}(E_r)$ and the sum over c an "entropic" contribution $\sum_c S_{\rm un}(\Sigma_c(\mathbf{E}))$. To adapt the formula to sparse superposition codes one must replace all scalars by B-dimensional vectors, replace $E[s^2]\to 1$, $\alpha\to 1/R$ and set $\tilde{q}_r\to E[s^2]-E_r = 1-E_r$.

Distribution Matching via Sparse Superposition Codes

3

In this chapter,[1] we formulate the fixed-length distribution matching as a Bayesian inference problem. Our proposed solution is inspired from the compressed sensing paradigm and the sparse superposition (SS) codes of Chapter 2. The distribution matching task requires a *matcher* at the transmitting end and a *dematcher* at the receiving end. In this chapter, we present a simple and exact matcher based on position modulation (PM), that introduces sparsity in the source, and Gaussian signal quantization. At the receiver, the dematcher exploits the sparsity in the source and performs low-complexity dematching based on generalized approximate message-passing (GAMP). We show that GAMP dematcher and spatial coupling lead to an asymptotically optimal performance, in the sense that the rate tends to the entropy of the target distribution with vanishing reconstruction error in a proper limit. Furthermore, we assess the performance of the dematcher on practical Hadamard-based operators. A remarkable inherent feature of the proposed solution is the possibility to: *i*) perform matching at the symbol level (nonbinary); *ii*) perform joint channel coding and matching. Note that all the theoretical guarantees of the proposed solution can be derived from Chapter 2. However, the formulation of distribution matching in a way that leverages the GAMP algorithm for a source coding problem is very novel and interesting by its own.

3.1 Introduction

Distribution matching has recently attracted lots of attention in long-haul fiber optical communications. As an inverse of data compression, a distribution

[1]The content of this chapter is based on a joint work with V. Aref and L. Schmalen [129].

matcher maps a discrete memoryless source, namely i.i.d. Bernoulli(1/2) bits, into a sequence of symbols distributed according to a target distribution. A dematcher is required to perform the inverse operation and recover the original source with a certain reliability.

As a primary application, distribution matching is used for probabilistic shaping [130] in order to imitate the capacity achieving input distribution of the channel and increase the spectral efficiency. The distribution matching task in probabilistic shaping can be done at the bit level by introducing bias in the binary source followed by a high-order modulation scheme that yields nonuiform symbols. However, one can perform distribution matching at the symbol level by directly mapping the binary sequence into the desired symbols, e.g. nonunifrom pulse-amplitude modulated (PAM) or quadrature-amplitude modulated (QAM) symbols. Distribution matching is also used to achieve the capacity of asymmetric channels [128] and for rate adaptation [15].

Optimal variable-length distribution matching schemes with offline algorithms were proposed in [131, 132, 133, 134]. A low-complexity online algorithm based on arithmetic coding was introduced in [135, 136]. Variable-length schemes require large buffer sizes and suffer from error propagation and synchronization problems [131]. Recently, an asymptotically optimal fixed-length and low-complexity distribution matcher was introduced in [137].

All the aforementioned schemes are lossless. However, their practical implementations require a separate forward error correction code to be added on top of the distribution matcher [138], which might incur a rate loss and error propagation for finite blocklengths. In this chapter, we propose a scheme which inherently performs joint channel coding and distribution matching. In particular, we formulate the fixed-length distribution matching as a Bayesian inference problem. The formulation is inspired from the compressed sensing paradiagm [38, 39] and sparse superposition (SS) codes [29, 78, 37, 118]. Moreover, we provide a low-complexity algorithm based on generalized approximate message-passing (GAMP) [60, 139] and spatial coupling. The proposed scheme is asymptotically optimal and it is motivated by the recent success of GAMP in quantized compressed sensing [140] and SS codes [116, 120, 121].

For the proposed scheme, the algorithmic performance under GAMP dematcher and the Bayes-optimal performance, under optimal dematcher, can be tracked by the state evolution (SE) and potential function. We show via SE analysis and numerical simulations that GAMP operates up to an "algorithmic rate" with a nonnegligible gap to the information theoretical rate. However, we illustrate that the GAMP dematcher on a spatially coupled version of the problem is asymptotically optimal in the blocklength, in the sense that the algorithmic rate saturates the Bayes-optimal rate which, in turn, tends to the entropy of the target distribution in a proper limit. Furthermore, unlike the existing approaches, the target distribution is attained for all blocklengths due to the simplicity of the matcher which is based on quantizing a Gaussian signal.

Bearing in mind practical implementations, we assess the performance of the dematcher on Hadmard-based operators that allow for substantial decrease

in the complexity and memory needs. It is noteworthy to mention that our approach provides a single-shot solution by performing distribution matching at the symbol level in addition to the possibility of implementing joint channel coding on memoryless channels, a promising solution for probabilistically-shaped coded modulation schemes [112, 113, 114].

3.2 Distribution Matching

In binary distribution matching, one is ideally interested in mapping a binary sequence $\mathbf{u} \in \{0,1\}^m$ with i.i.d. Bernoulli(1/2) bits into another discrete sequence $\mathbf{y} \in \mathcal{A}^M$ having a target marginal distribution P_Y. The mapping is done such that $\mathbf{u}$ is perfectly reconstructed from $\mathbf{y}$. We call $\mathbf{u}$ the *source* and $\mathbf{y}$ the *target*. Let Y be the target random variable with alphabet $\mathcal{A}$ and distribution P_Y. The maximal achievable rate (or the information theoretical rate) of lossless distribution matching is given by

$$R = \frac{m}{M} \leq H(Y), \tag{3.1}$$

where $H(Y)$ is the entropy of Y. In the binary-to-binary case, $\mathbf{u}$ is mapped to another binary sequence $\mathbf{y} \in \{0,1\}^M$ with M Bernoulli($p^\star$) bits, where $p^\star$ represents the target distribution. (see Fig. 3.1). The maximal achievable rate in this case is given by

$$R = \frac{m}{M} \leq h_2(p^\star), \tag{3.2}$$

where $h_2(\cdot)$ is the binary entropy function.

Note that one can frame this problem as the inverse of the lossless source coding problem. In source coding, one is normally interested in mapping a discrete memoryless source, possibly binary, with a presumably nonuniform distribution into the shortest possible binary sequence. The lower bound on the lossless compression rate is given by the entropy of the original source. In distribution matching, we need to expand the original binary uniform source into another discrete source, possibly binary as well, in order to reach a target distribution. The lower bound on the lossless "expansion rate" (here M/m) is given by the inverse of the target distribution's entropy.

Consequently, a natural approach to solve the distribution matching problem is to use variable-length prefix-free source coding schemes such as Huffman codes [131, 132, 133, 134] or low-complexity arithmetic codes [135, 136]. In this case, perfect reconstruction is guaranteed for all blocklengths, while the distortion measure is defined to be the normalized Kullback-Leibler (KL) divergence between the matcher distribution and the target distribution [137]. As the blocklength increases, the rate of the aforementioned schemes tends to the maximal achievable rate (3.1) with vanishing normalized KL divergence between the matcher and target distributions. However, the main limitation of

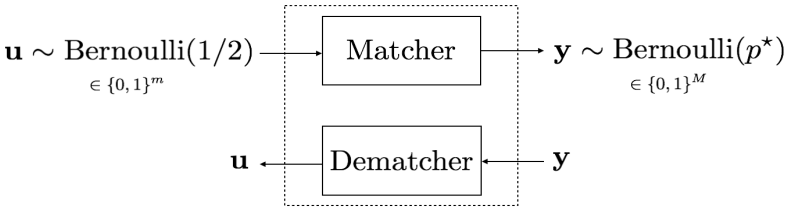

Figure 3.1: Binary-to-binary distribution matching block diagram.

these schemes is the varying transmission rate. Recently, a fixed-rate approach based on constant composition arithmetic coding was introduced in [137].

Another approach to solve the distribution matching is to employ a forward error correcting code [128] as done in lossless source coding [141, 142, 143]. In this approach, the target distribution can be matched while the distortion measure is the reconstruction error, between the original source and the reconstructed one, that vanishes in the blocklength. Although this approach might be erroneous at finite blocklengths, it remains very useful for many application scenarios because of its amenability to perform joint channel coding and matching. Following this second approach, we propose a solution that employs the SS codes for distribution matching and relies on the recent success of GAMP algorithm for such codes [120, 121, 116].

3.3 Application: Probabilistic Shaping for Optical Channels

One of the main applications of distribution matching is probabilistic shaping as mentioned earlier. Driven by the recent advances in fiber optical communications, probabilistic signal shaping has attracted lots of attention in the optical community. As the optical channel is bandwidth-efficient, the adoption of high-order modulation schemes is the pathway toward increasing the spectral efficiency. This can be done without increasing the power which is limited by the nonlinearities of the channel.

The nonlinearities in the optical channels are often parameterized by some variants of the Gaussian noise (GN) model [144, 145, 146, 147]. Therefore, the adoption of constellations that follow Maxwell-Boltzmann (MB) distributions is very favorable as it allows to operate close to capacity [131]. The use of classical uniform signalling with equidistant constellations incurs a performance loss in terms of data rate. The mismatch between the signal distribution and the capacity-achieving input distribution induces a gap to capacity, which might be significant in the high signal-to-noise ratio (SNR) regime. Therefore, optimizing the signalling is essential in order to take full advantage of the optical channel and increase the spectral efficiency, specially with high-order modulation schemes. The two possible techniques to achieve this and

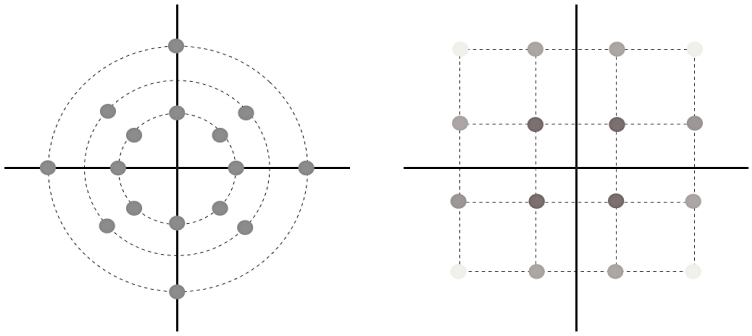

Figure 3.2: Left: 16-QAM constellation with geometric shaping. A nonequidistant constellation is used where all the symbols occur with a uniform probability $1/16$. Right: 16-QAM constellation with probabilistic shaping. An equidistant constellation is used instead but the symbols occur with nonuniform probability. The nonuniform probability is highlighted via the gray scale for the 16 symbols (e.g. the symbols at the outer corners occur with the lowest probability).

circumvent the performance loss are geometric shaping and probabilistic shaping. While the former uses uniformly distributed (or equiprobable) symbols on nonuniform (or nonequidistant) constellations, the latter uses nonuniformly distributed symbols on uniform constellations (see Fig. 3.2).

Using nonuniform constellations, as in geometric shaping, necessitates drastic modifications in the signal processing algorithms. Therefore, probabilistic shaping represents a viable alternative as it relies on the uniform equidistant constellations, and hence uses the preexisting signal processing algorithms and digital-to-analog converters. Moreover, probabilistic shaping seems to outperform the geometric shaping in terms of achievable rates [138]. The probabilistic shaping task requires a distribution matcher at the transmitter in order to map the original source, typically binary with i.i.d. Bernoulli(1/2) bits, into the nonuniform symbols. An inverse operation is required at the receiver to recover the source. In the sequel, we will present a new distribution matching scheme that can be used for probabilistic shaping. A key feature of the proposed scheme is the ability to perform distribution matching directly at the symbol level. Moreover, the distribution matching can be done jointly with channel coding, and thus spare any additional forward error correction code.

3.4 Compressed Sensing Approach for Distribution Matching

Our proposed solution, depicted in Fig. 3.3, employs the SS codes used for general channel in Chapter 2 to perform distribution matching. One can formulate the distribution matching as a SS code on a deterministic nonlinear

channel, and hence leverage the GAMP algorithm to perform the dematching. The GAMP dematcher identifies an *effective channel*, which can be a concatenation of the deterministic matcher (quantizer) with a noisy channel, over which the estimation is performed. Thus, distribution matching and channel coding can be jointly performed. In this chapter, we focus on the distribution matching part while the channel coding part was already investigated in Chapter 2. Hence, our channel is a quantization of a Gaussian signal that yields the target distribution. Note that in Chapter 2, SS codes were used for forward error correction over general channel while other techniques [128] were proposed to perform the distribution matching task. The main contribution of this chapter is to show that SS codes can be used to perform distribution matching concurrently for any discrete alphabet.

3.4.1 Matcher

In order to use the compressed sensing and AMP paradigms, we need to introduce sparsity in $\mathbf{u}$. This can be done via simple position modulation (PM) scheme (see Fig. 3.3). Following the same construction of Chapter 2, we take $m = L\log_2(B)$ with B chosen to be a power of 2. The original source can be seen as a vector made of L *sections*, $\mathbf{u} = [\mathbf{u}_1, \dots, \mathbf{u}_L]$, where each section $\mathbf{u}_l$, $l \in \{1, \dots, L\}$ is a $\log_2(B)$-dimensional vector. We call B the *section size*. The original source is then mapped to a sparse *signal* $\mathbf{s}$ made of L sections. Each section $\mathbf{s}_l$ is a B-dimensional vector with a single nonzero component that is equal to 1. The position of the non-zero component in $\mathbf{s}_l$ is specified by the binary representation of $\mathbf{u}_l$. For example if $B = 4$ and $L = 5$, a valid source is $\mathbf{u} = [00, 01, 11, 10, 01]$ which corresponds to $\mathbf{s} = [0001, 0010, 1000, 0100, 0010]$. We set $N = LB$.

A fixed *coding matrix* $\mathbf{F} \in \mathbb{R}^{M\times N}$ is taken with i.i.d real Gaussian entries distributed as $\mathcal{N}(0, 1/L)$. We use this matrix to obtain a *codeword* $\mathbf{z} = \mathbf{Fs} \in \mathbb{R}^M$ with i.i.d. standard Gaussian entries. The matching task consists of quantizing the Gaussian codeword entries through a quantizer $\Phi(\cdot)$ acting componentwise with

$$y_i = \Phi(z_i) = \Phi([\mathbf{Fs}]_i), \quad i = 1, \dots, M, \tag{3.3}$$

such that the output is distributed according to a given target distribution. Note that one can look at $\Phi(\cdot)$ as a deterministic nonlinear channel leading to the target distribution.

For the binary case and a target distribution of Bernoulli($p^\star$), the quantizer takes the following form as depicted in Fig. 3.4

$$y_i = \Phi(z_i) = \ \text{sign}(z_i - Q^{-1}(p^\star)), \quad i = 1, \dots, M, \tag{3.4}$$

where $Q^{-1}(\cdot)$ is the inverse of the Gaussian Q-function defined by $Q(x) = \int_x^{+\infty} dt\, \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}$. The output $\mathbf{y}$ of the quantizer is in $\{-1, +1\}^M$ which can be mapped to $\{0, 1\}^M$.

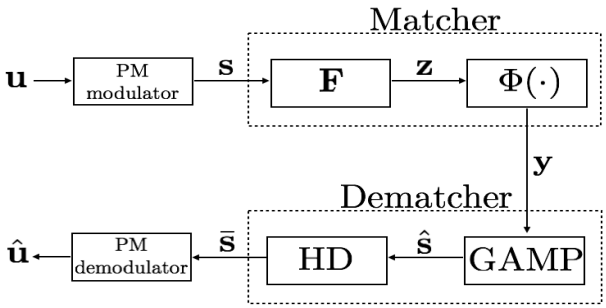

Figure 3.3: The original source **u** is mapped to a sparse signal **s** through a PM modulator. A quantizer $\Phi(\cdot)$ is then applied to obtain the target sequence **y**. The GAMP algorithm provides soft valued estimate $\hat{\mathbf{s}}$ of **s** in the MMSE sense. A simple hard decision (HD) scheme is used to provide the binary decoded message $\bar{\mathbf{s}}$ by setting the most biased component in each section of $\hat{\mathbf{s}}$ to 1 and the others to 0. The reconstructed version $\hat{\mathbf{u}}$ of the original source **u** can be easily recovered from $\bar{\mathbf{s}}$ using PM demodulator.

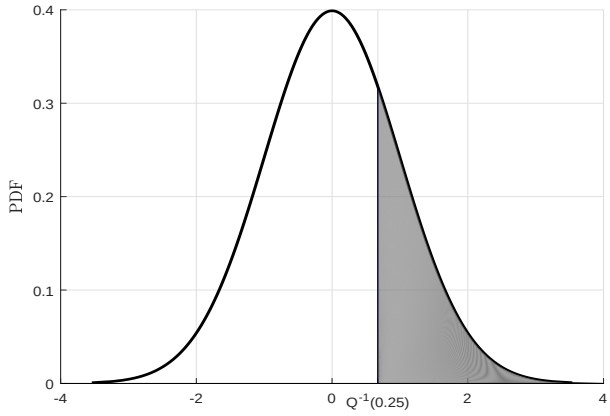

Figure 3.4: Quantization of a standard Gaussian to Bernoulli$(1/4)$ bits.

This approach is readily generalizable beyond the binary case to q-ary discrete target distributions (e.g. PAM or QAM symbols). In this case, the quantizer Φ uses biased q-quantiles of the Gaussian distribution for quantization. Specifically, let Y be a discrete random variable with q-ary alphabet $\mathcal{A} = \{a_1, \ldots, a_q\}$ and distribution $P_Y(a_k) = P_k$ $(k \in \{1, \ldots, q\})$. The quantizer is defined by

$$\Phi(z) = a_k \quad \text{if } z \in]c_{k-1}, c_k], \tag{3.5}$$

with

$$c_k = \begin{cases} -\infty & k = 0, \\ Q^{-1}(1 - \sum_{j=1}^{k} P_j) & k = 1, \ldots, q, \end{cases} \tag{3.6}$$

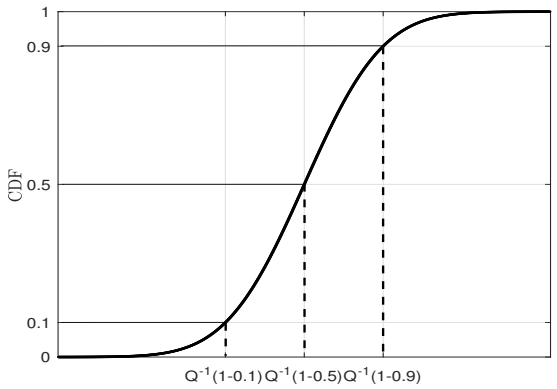

Figure 3.5: Quantization intervals of a standard Gaussian signal to nonuniform 4-PAM constellation $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$ with $P_Y = [0.1, 0.4, 0.4, 0.1]$.

Note that this simple matching operation ensures that the target distribution is attained for all blocklengths. Note also that the quantizer $\Phi(\cdot)$ was denoted $\pi(\cdot)$ in the binary case of Chapter 2. We make this distinction to stress that this map can be applied to any q-ary alphabet. The quantization intervals of a nonuniform 4-PAM constellation based on the cumulative distribution function of a standard Gaussian are shown in Fig. 3.5.

3.4.2 Dematcher

The dematching task is to recover a sparse signal $\mathbf{s}$, and hence $\mathbf{u}$, from quantized random projections $\mathbf{y}$ as depicted in Fig. 3.3. The sparsity introduced in the signal by PM can be harnessed at the dematcher in a compressed sensing fashion. Namely, the dematching can be interpreted as a compressed sensing problem with structured sparsity. Consequently, the same algorithms and analysis tools used in compressed sensing theory, such as GAMP algorithm and SE, can be used for the dematching task as done for SS codes of Chapter 2. For a Bernoulli(1/2) source and PM scheme, the sections of $\mathbf{s}$ are uniformly distributed over all the possible B-dimensional vectors with a single nonzero component that is equal to 1. The prior of each section is denoted by $p_0(\mathbf{s}_l)$.

In a Bayesian setting, the estimation of the signal $\mathbf{s}$, based on the observed target $\mathbf{y}$ and a fixed matrix $\mathbf{F}$, can be done in a minimum mean-square error (MMSE) sense or maximum a-posteriori (MAP) sense. This necessitates the computation of the posterior distribution of $\mathbf{s}$ given $\mathbf{y}$ and $\mathbf{F}$ on a dense graphical model. Therefore, one can use an iterative message-passing algorithm such as GAMP, which was first introduced in compressed sensing [60, 139] and then adapted to account for any structured B-dimensional prior distribution [78, 37, 116]. The GAMP algorithm uses Gaussian and quadratic approximations that yield a sequence of disjoint estimation problems under an *equivalent Gaussian noise*. The real physical channel appears in the computation of the

moments of the equivalent Gaussian noise. The physical channel in our distribution matching problem is a deterministic highly nonlinear channel defined by the quantizer (3.3). Thus, the GAMP algorithm of Chapter 2 (Algorithm 2.1) can be adapted to act on such channel. Note that for joint channel coding and matching, the channel would be a concatenation of the quantizer and the noisy channel.

Algorithm 2.1 of Chapter 2 shows the steps of GAMP. The same algorithm can be used here and the only difference resides in the computation of Steps 8 and 9, which depend on the the quantizer (3.3). For the q-ary quantizer given in (3.5) and (3.6), the i^{th} entry of g_{out} and $\frac{\partial}{\partial \mathbf{p}} g_{\text{out}}$ take the following forms respectively

$$\begin{aligned}
[g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})]_i &= \frac{\sum_{k=1}^{q} \delta(y_i - a_k)\big(Q'_{k-1}(p_i,\tau_i) - Q'_k(p_i,\tau_i)\big)}{\sum_{k=1}^{q} \delta(y_i - a_k)\big(Q_{k-1}(p_i,\tau_i) - Q_k(p_i,\tau_i)\big)}, \\
[\frac{\partial}{\partial \mathbf{p}} g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})]_i &= \big([g_{\text{out}}(\mathbf{p},\mathbf{y},\boldsymbol{\tau})]_i\big)^2 \\
&\quad - \frac{\sum_{k=1}^{q} \delta(y_i - a_k)\big(Q''_{k-1}(p_i,\tau_i) - Q''_k(p_i,\tau_i)\big)}{\sum_{k=1}^{q} \delta(y_i - a_k)\big(Q_{k-1}(p_i,\tau_i) - Q_k(p_i,\tau_i)\big)},
\end{aligned}$$

for $i = 1, \ldots, M$, with

$$\begin{cases}
Q_k(p,\tau) = Q(\frac{c_k - p}{\sqrt{\tau}}) \\
Q'_k(p,\tau) = \frac{e^{-(c_k-p)^2/(2\tau)}}{\sqrt{2\pi\tau}} \\
Q''_k(p,\tau) = Q'_k(p,\tau)\frac{c_k - p}{\tau},
\end{cases} \tag{3.7}$$

where $Q(\cdot)$ in the first equation of (3.7) denotes the standard Gaussian Q-function while the c_k's are given in (3.6). Steps 12 and 13 of Algorithm 2.1 depend on the prior p_0, which is the same for SS codes.

The GAMP algorithm requires an exchange of $\mathcal{O}(N)$ messages. The complexity of computing each message is dominated by a matrix-vector multiplication. In fact, both the matcher and the GAMP dematcher involve matrix-vector multiplication. Hence, the worst case complexity, per message, is $\mathcal{O}(MN)$. This can be simplified using structured operators such as Fourier, wavelet or Hadamard. While random Gaussian matrices are mathematically more tractable and easier for analysis, the structured matrices provide practical advantages and more robust finite-length performance [77]. Hadamard-based matrices constructed as in [77], with random sub-sampled modes of the full Hadamard operator, allow to achieve a complexity of $\mathcal{O}\big(m \ln(N)\big)$ and drastically reduce the storage need. Note that using such matrices might necessitate fine tuning the quantizer (3.3) as the codeword's distribution deviates from Gaussian. Moreover, one would need to use other variants of AMP that are better suited to general matrices [148, 149, 150].

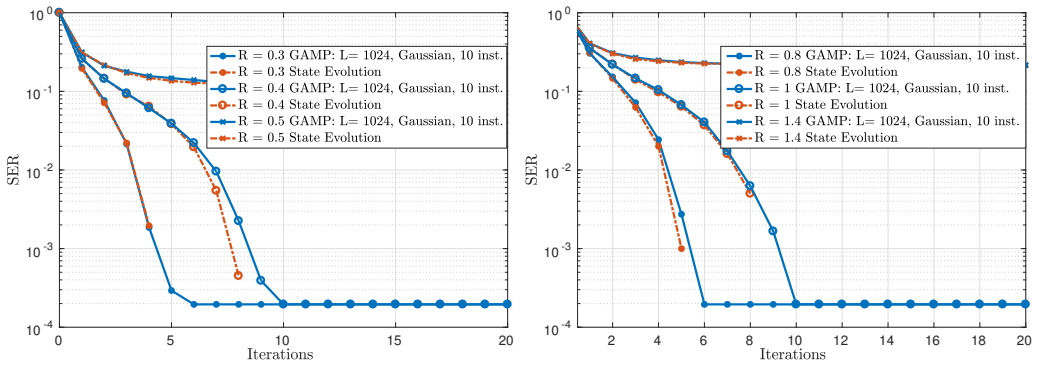

Figure 3.6: The performance of GAMP at each iteration in terms of SER under Gaussian coding matrices. Left: binary-to-binary distribution matching with target $p^\star = 1/4$; the information theoretical rate of such scheme is $R = h_2(p^\star) = 0.8113$. Right: binary to 4-ary distribution matching with target distribution $P_Y = [0.1, 0.4, 0.3, 0.2]$; the information theoretical rate of such scheme is $R = H(Y) = 1.8464$. We fix $B = 4$ and we simulate GAMP under various rates. As long as the rate is small enough (i.e. $R < R_{\text{GAMP}}$), GAMP (solid blue line) performs the dematching task up to a finite-length error floor that vanishes with L. As L increases, the GAMP performance coincides with the SE prediction (dotted red line) and the error floor vanishes.

3.5 Performance Evaluation

The same analysis tools used in Chapter 2, namely the state evolution (SE) and potential function, are adapted and used for analysis here. An important aspect of GAMP algorithm is that its asymptotic performance can be analytically tracked at each iteration by the SE equation [151]. SE is a simple recursion analogous to the density evolution used to track the performance of low-density parity-check (LDPC) codes on sparse graphical models. Moreover, the ultimate Bayes-optimal performance of our proposed scheme, i.e. the performance under optimal algorithm, can be obtained from the potential function [78, 116] inspired from statistical physics techniques and elaborated in Chapter 2. Note that the GAMP performance is typically assessed using the mean-square error (MSE) between $\mathbf{s}$ and $\hat{\mathbf{s}}$ or the section error rate (SER) between $\mathbf{s}$ and $\bar{\mathbf{s}}$ (i.e. the fraction of sections that are wrongly reconstructed, see Fig. 3.3 for $\hat{\mathbf{s}}$ and $\bar{\mathbf{s}}$).

Numerical simulations as well as SE analysis show the following: for any fixed section size B, the GAMP algorithm exhibits asymptotically in L a "phase transition" at an algorithmic rate (or threshold) denoted by R_{GAMP}. Formally, R_{GAMP} is the maximum rate at which the GAMP algorithm performs the dematching task with vanishing reconstruction error. As soon as the rate exceeds this threshold, GAMP algorithm fails. These observations are depicted in Fig. 3.6 for both binary-to-binary and binary to q-ary distribution matching with $B = 4$. Our empirical results shown in Fig. 3.6 confirm

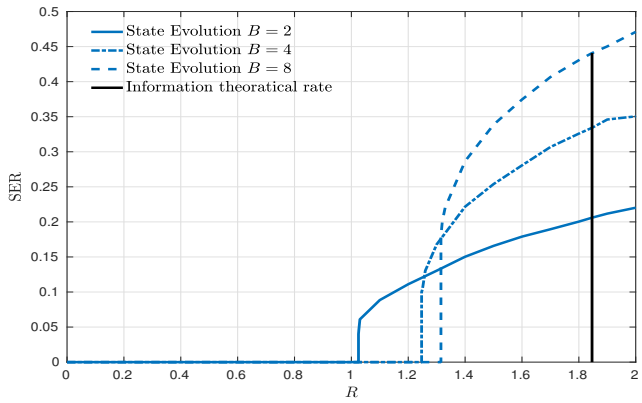

Figure 3.7: The performance of GAMP as predicted by the SE for binary to 4-ary distribution matching with target distribution $P_Y = [0.1, 0.4, 0.3, 0.2]$. We perform the SE analysis for three different section sizes. A sharp phase transition occurs for each section size: $R_{\text{GAMP}} = 1.025$ for $B = 2$, $R_{\text{GAMP}} = 1.247$ for $B = 4$, $R_{\text{GAMP}} = 1.315$ for $B = 8$. Note that although the gap to the information theoretical rate varies with B, a nonnegligible gap persists with the current construction as B increases.

that the SE tracks the asymptotic performance of GAMP dematcher. An alternative way to see the phase transition behavior is presented in Fig. 3.7 where the final SER, after SE convergence, is plotted as a function of the rate for three different section sizes.

The empirical algorithmic rate R_{GAMP}, obtained from running GAMP on real instances, as well as the one obtained from the numerical SE analysis for the current "uncoupled" construction are shown on the upper curve of Fig. 3.8 for different values of B. Under Gaussian coding matrices, the performance is accurately predicted by the SE for all values of B. Using Hadamard-based matrices incurs a small performance loss, in terms of R_{GAMP}, that vanishes with B. However, a gap to the information theoretical rate persists as B increases.

The gap to the information theoretical rate is due to the sub-optimality of GAMP, which is a low-complexity iterative algorithm. In order to predict the Bayes-optimal performance of our proposed scheme under optimal algorithm, which is computationally intractable, we use the potential function. Numerical simulations show that the Bayes-optimal rate (or potential threshold) denoted by R_{opt} approaches the information theoretical rate as the section size increases (see Fig. 3.8). Moreover, using the same analysis of the potential function as in Chapter 2, one can argue that R_{opt} indeed tends to the information theoretical rate as $B \to \infty$.

An effective approach to boost the algorithmic performance of GAMP is to apply spatial coupling as done for capacity achieving SS codes of Chapter 2. There are different ways to construct spatially coupled coding matrix and to impose the *seed* at the boundaries. One way to impose the seed was already

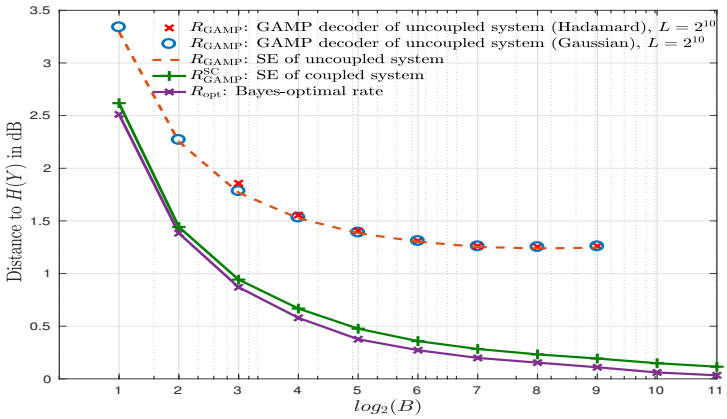

Figure 3.8: The distance between the algorithmic rate and $H(Y)$ in dB as a function of the section size. Binary-to-binary distribution matching is performed with target $p^\star = 1/4$. For the uncoupled system: the gap to the information theoretical rate persists even with large section sizes (dotted line). For the spatially coupled system: the algorithmic rate (green curve) follows the optimal rate (purple curve) that tends to $H(Y)$ as B increases. Spatial coupling is performed with the following coupling parameters: $L_c = 32$, $L_r = 33$, $w_b = 2$, $w_f = 2$, $\beta = 1.2$ and $J = 0.1$. The small mismatch between the purple and green curves is due to the finite length of coupling parameters. As the coupling parameters increase, the two curves coincide (threshold saturation phenomenon).

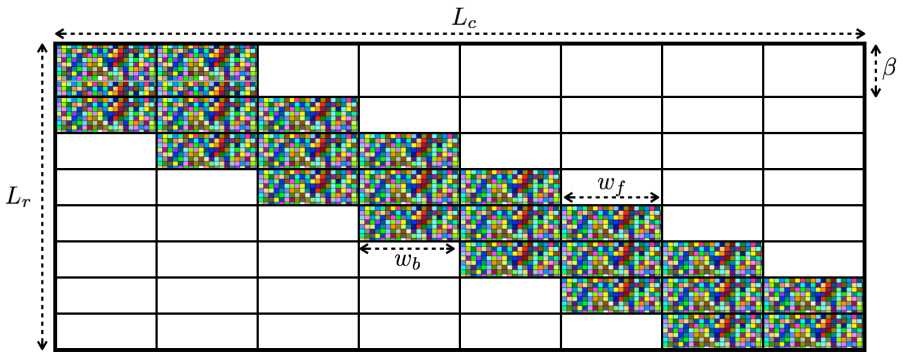

Figure 3.9: A spatially coupled coding matrix **F** with $L_r \times L_c$ blocks. Besides the diagonal blocks, there are w_b and w_f off-diagonal blocks with nonzero elements representing the backward and forward coupling windows respectively. Each block is of dimension $M \times N$ except the blocks in the first block-row that have a dimension of $\beta M \times N$, where β represents the seed rate. Each nonzero block is composed of i.i.d. real Gaussian entries with zero mean and a certain variance such that the variances of each block-row add up to 1. The variances can be tuned in a uniform or nonuniform fashion using the coupling strength parameter J as done in [78, 77].

introduced in 2.2. The seed was imposed using the *pinning condition* which is very suited for the analysis. In this chapter, we provide an alternative practical way to impose the seed using higher measurement rate at the boundary as depicted in Fig. 3.9.

It turns out that spatial coupling significantly improves the performance and decreases the gap to the information theoretical rate $H(Y)$. The gap can be made arbitrarily small by increasing B (see Fig. 3.8). Fig. 3.8 shows that the performance of GAMP under Gaussian matrices (circles) is accurately predicted by the SE analysis. Applying GAMP, originally developed for Gaussian matrices, on Hadamard-based matrices (crosses) yields a small mismatch w.r.t. the SE prediction. This lack of accuracy in SE for non-Gaussian matrices can be handled by using other variants of AMP [148, 149, 150].

Our SE simulations show that the algorithmic rate of the spatially coupled system denoted by $R^{\mathrm{SC}}_{\mathrm{GAMP}}$ follows the Bayes-optimal rate, which tends to the information theoretical rate. Actually, one can show that the algorithmic rate of the spatially coupled system equals the Bayes-optimal rate in a proper limit as done for SS codes of Chapter 2. This phenomenon turns to be quite general and it is coined *threshold saturation*. It was first introduced in the context of spatially coupled LDPC codes [152] and then generalized in [82] to any problem tracked by a scalar recursion.

Symmetric Rank-One Matrix Factorization

4

Factorizing low-rank matrices is a problem with many applications in machine learning and statistics, ranging from sparse principal component analysis to community detection and sub-matrix localization. For probabilistic models in the Bayes optimal setting, general expressions for the mutual information have been proposed using powerful heuristic statistical physics computations via the replica and cavity methods, and proven in few specific cases by a variety of methods. In this Chapter,[1] we use the spatial coupling methodology developed in the framework of error correcting codes, to rigorously derive the mutual information for the symmetric rank-one case. We characterize the detectability phase transitions in a large set of estimation problems, where we show that there exists a gap between what currently known polynomial algorithms (in particular spectral methods and approximate message-passing) can do and what is expected information theoretically. Moreover, we show that the computational gap vanishes for the proposed spatially coupled model, a promising feature with many possible applications. Our proof technique has an interest on its own and exploits three essential ingredients: the interpolation method first introduced in statistical physics, the analysis of approximate message-passing algorithms first introduced in compressed sensing, and the theory of threshold saturation for spatially coupled systems first developed in coding theory. Our approach is very generic and can be applied to many other open problems in statistical estimation where heuristic statistical physics predictions are available.

[1]The content of this chapter is based on a joint work with J. Barbier, N. Macris, F. Krzakala and L. Zdeborová [153]

4.1 Introduction

We consider the following probabilistic rank-one matrix factorization (or rank-one matrix estimation) problem: one has access to noisy observations $\mathbf{w} = (w_{ij})_{i,j=1}^{n} \in \mathbb{R}^{n,n}$ of the pair-wise product of the components of a vector $\mathbf{s} = (s_i)_{i=1}^{n} \in \mathbb{R}^n$ where the components are i.i.d random variables distributed according to $S_i \sim P_0$, $i = 1, \dots, n$.[2] The matrix elements of $\mathbf{w}$ are observed through a noisy element-wise (possibly non-linear) output probabilistic channel $P_{\text{out}}(w_{ij}|s_i s_j)$, with $i, j = 1, \dots, n$. The goal is to estimate the vector $\mathbf{s}$ from $\mathbf{w}$, up to a global flip of sign in general, assuming that both distributions P_0 and P_{out} are known. We assume the noise to be symmetric so that $w_{ij} = w_{ji}$. There are many important problems in statistics and machine learning that can be expressed in this way, among which:

- *Sparse PCA*: Sparse principal component analysis (PCA) is a dimensionality reduction technique where one looks for a low-rank representation of a data matrix with sparsity constraints [43]. The following is the simplest probabilistic symmetric version where one estimates a rank-one matrix. Consider a sparse random vector $\mathbf{S}$, for instance drawn from a Gauss-Bernoulli distribution, and take an additive white Gaussian noise (AWGN) channel where the observations are $W_{ij} = S_i S_j/\sqrt{n} + \Delta Z_{ij}$ whith $Z_{ij} \sim \mathcal{N}(0, 1)$. Here[3] $P_{\text{out}}(w_{ij}|s_i s_j) = \mathcal{N}(w_{ij}|s_i s_j/\sqrt{n}, \Delta)$.

- *Spiked Wigner model*: In this model the noise is still Gaussian, but the vector $\mathbf{S}$ is assumed to be a Bernoulli random vector with i.i.d components $S_i \sim \text{Ber}(\rho)$. This formulation is a particular case of the spiked covariance model in statistics introduced by [154, 155]. It has also attracted a lot of attention in the framework of random matrix theory (see for instance [44] and references therein).

- *Community detection*: In its simplest setting, one uses a Rademacher vector $\mathbf{S}$ where each variable take values $S_i \in \{-1, 1\}$ depending on the "community" it belongs to. The observation model then introduces missing information and errors such that, for instance, $P_{\text{out}}(w_{ij}|s_i s_j) = p_1\delta(w_{ij} - s_i s_j) + p_2\delta(w_{ij} + s_i s_j) + (1 - p_1 - p_2)\delta(w_{ij})$, where $\delta(\cdot)$ is the Delta dirac function. These models have recently attracted a lot of attention both in statistics and machine learning contexts (see e.g. [40, 41, 156, 157, 158, 159]).

- *Sub-matrix localization*: This is the problem of finding a submatrix with an elevated mean in a large noisy matrix, as in [160, 161].

[2] Note that in Chapter 1, the dimension of the problem was denoted by the capital letter N for coherence with the sparse superposition codes notations. Here, we will use the small letter n, which is more consistent with the matrix factorization literature. We hope that this will not confuse the reader.

[3] In this chapter $\mathcal{N}(x|m, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-(x - m)^2/2\sigma^2))$

- *Matrix completion*: A last example is the matrix completion problem where a part of the information (the matrix elements) is hidden, while the rest is given with noise. For instance, a classical model is $P_{\text{out}}(w_{ij}|s_i s_j) = p\delta(w_{ij}) + (1-p)\mathcal{N}(w_{ij}|s_i s_j, \Delta)$. Such problems have been extensively discussed over the last decades, in particular because of their connection to collaborative filtering (see for instance [45, 46, 162, 163]).

Here we shall consider the probabilistic formulation of these problems and focus on estimation in the mean square error (MSE) sense. We rigorously derive an explicit formula for the mutual information in the asymptotic limit, and for the information theoretic minimal mean square error (MMSE). Our results imply that in a large region of parameters, the posterior expectation of the underlying signal, a quantity often assumed intractable to compute, can be obtained using a polynomial-time scheme via the approximate message-passing (AMP) framework [164, 165, 166, 42, 167]. We also demonstrate the existence of a region where no *known* tractable algorithm is able to find a solution correlated with the ground truth. Nevertheless, we prove explicitly that it is information theoretically possible to do so (even in this region), and discuss the implications in terms of computational complexity.

The crux of our analysis rests on an "auxiliary" spatially coupled (SC) system. The hallmark of SC models is that one can tune them so that the gap between the algorithmic and information theoretic limits is eliminated, while at the same time the mutual information is maintained unchanged for the coupled and original models. Roughly speaking, this means that it is possible to algorithmically compute the information theoretic limit of the original model because a suitable algorithm is optimal on the coupled system.

Our proof technique has an interest by its own as it combines recent rigorous results in coding theory along the study of capacity-achieving SC codes [79, 25, 82, 85, 116] with other progress coming from developments in mathematical physics of spin glass theory [168]. Moreover, our proof exploits the "threshold saturation" phenomenon of the AMP algorithm and uses spatial coupling as a proof technique. From this point of view, we believe that the theorem proven in this chapter is relevant in a broader context going beyond low-rank matrix estimation and can be applied for a wide range of inference problems where message-passing algorithm and spatial coupling can be applied. Furthermore, our work provides important results on the exact formula for the MMSE and on the optimality of the AMP algorithm.

Hundreds of papers have been published in statistics, machine learning or information theory using the non-rigorous statistical physics approach. We believe that our result helps setting a rigorous foundation of a broad line of work. While we focus on rank-one symmetric matrix estimation, our proof technique is readily extendable to more generic low-rank symmetric matrix or low-rank symmetric tensor estimation. We also believe that it can be extended to other problems of interest in machine learning and signal processing. It has already been extended to linear estimation and compressed sensing [122, 104].

We conclude this introduction by giving a few pointers to the recent literature on rigorous results. For rank-one symmetric matrix estimation problems, AMP has been introduced by [164], who also computed the state evolution formula to analyze its performance, generalizing techniques developed by [69], [115] and [169]. State evolution was further studied by [166] and [42]. In [170, 167], the generalization to larger rank was also considered. The mutual information was already computed in the special case when $S_i = \pm 1 \sim \mathrm{Ber}(1/2)$ in [171] where an equivalent spin glass model was analyzed. The results of [171] were first generalized in [47] who, notably, obtained a generic matching upper bound. The same formula was also rigorously computed following the study of AMP in [166] for spike models (provided, however, that the signal was not *too* sparse) and in [42] for strictly symmetric community detection. The general formula proposed by [170] for the conditional entropy and the MMSE on the basis of the heuristic cavity method from statistical physics was first demonstrated in full generality in [102]. This chapter represents an extended version of [102] that includes all the proofs and derivations along with more detailed discussions. All preexisting proofs could not reach the more interesting regime where a gap between the algorithmic and information theoretic performances appears (i.e. in the presence of "first order" phase transition), leaving a gap with the statistical physics conjectured formula. Following the work of [102], the replica formula for rank-one symmetric matrix estimation has been proven again several times using totally different techniques that involve the concentration's proof of the overlaps [172, 173]. Our proof strategy does not require any concentration and it uses AMP and spatial coupling as proof techniques. Hence, our result has more practical implications in terms of proving the range of optimality of the AMP algorithm for both the underlying (uncoupled) and spatially coupled models.

This chapter is organized as follows: the problem statement and the main results are given in Section 4.2 along with a sketch of the proof, two applications for symmetric rank-one matrix estimation are presented in Section 4.3, the threshold saturation phenomenon and the relation between the underlying and spatially coupled models are proven in Section 4.4 and Section 4.5 respectively, the proof of the main results follows in Section 4.6 and Section 4.7.

A word about notations: in this chapter, we use capital letters for random variables, and small letters for fixed realizations. Matrices and vectors are bold while scalars are not. Components of vectors or matrices are identified by the presence of lower indices.

4.2 Setting and Main Results

4.2.1 Basic Underlying Model

A standard and natural setting is to consider an additive white gausian noise (AWGN) channel with variance Δ assumed to be known. The model reads

$$w_{ij} = \frac{s_i s_j}{\sqrt{n}} + \sqrt{\Delta} z_{ij}, \tag{4.1}$$

where $\mathbf{z} = (z_{ij})_{i,j=1}^n$ is a symmetric matrix with $Z_{ij} \sim \mathcal{N}(0,1)$, $1 \leq i \leq j \leq n$, and $\mathbf{s} = (s_i)_{i=1}^n$ has i.i.d components $S_i \sim P_0$. We set $\mathbb{E}[S^2] = v$. Precise hypothesis on P_0 are given later.

Perhaps surprisingly, it turns out that the study of this Gaussian setting is sufficient to completely characterize all the problems discussed in the introduction, even if we are dealing with more complicated (noisy) observation models. This is made possible by a theorem of channel universality. Essentially, the theorem states that for any output channel $P_{\text{out}}(w|y)$ such that at $y = 0$ the function $y \mapsto \log P_{\text{out}}(w|y)$ is three times differentiable with bounded second and third derivatives, then the mutual information satisfies

$$I(\mathbf{S}; \mathbf{W}) = I(\mathbf{S}\mathbf{S}^\intercal; \mathbf{S}\mathbf{S}^\intercal + \sqrt{\Delta}\, \mathbf{Z}) + O(\sqrt{n}), \tag{4.2}$$

where Δ is the inverse Fisher information (evaluated at $y = 0$) of the output channel

$$\Delta^{-1} := \int dw P_{\text{out}}(w|0) \left(\frac{\partial \log P_{\text{out}}(w|y)}{\partial y}\Big|_{y=0} \right)^2. \tag{4.3}$$

This means that the mutual information per variable $I(\mathbf{S}; \mathbf{W})/n$ is asymptotically equal the mutual information per variable of an AWGN channel. Informally, it implies that we only have to compute the mutual information for an "effective" Gaussian channel to take care of a wide range of problems. The statement was conjectured in [170] and can be proven by an application of the Lindeberg principle [42], [47].

4.2.2 AMP Algorithm and State Evolution

AMP has been applied for the rank-one symmetric matrix estimation problems by [164], who also computed the state evolution formula to analyze its performance, generalizing techniques developed by [69] and [115]. In [174], AMP was used in conjunction with a spectral initialization. State evolution was further studied by [166] and [42]. AMP is an iterative algorithm that provides an estimate $\hat{\mathbf{s}}^{(t)}(\mathbf{w})$, at each iteration $t \in \mathbb{N}$, of the vector $\mathbf{s}$. It turns out that tracking the asymptotic vector and matrix MSE of the AMP algorithm is equivalent to running a simple recursion called *state evolution* (SE).

The AMP algorithm reads

$$\begin{cases} \hat{s}_j^{(t)} = \eta_t((\mathbf{w}\hat{\mathbf{s}}^{(t-1)})_j - b^{(t-1)}\hat{\mathbf{s}}_j^{(t-2)}), \\ b^{(t)} = \frac{1}{n}\sum_{i=1}^n \eta_t'((\mathbf{w}\hat{\mathbf{s}}^{(t-1)})_i - b^{(t-1)}\hat{s}_i^{(t-2)}) \end{cases} \tag{4.4}$$

for $j = 1, \cdots, n$, where $\eta_t(y)$ is called the *denoiser* function and $\eta_t'(y)$ is the derivative w.r.t y. The denoiser is the MMSE estimator associated to an "equivalent scalar denoising problem"

$$y = s + \Sigma(E)z, \qquad \Sigma(E)^{-2} := \frac{v - E}{\Delta}\,. \tag{4.5}$$

with $Z \sim \mathcal{N}(0,1)$ and

$$\eta(y) = \mathbb{E}[X|Y = y] = \frac{\int dx x P_0(x) e^{-\frac{(x-Y)^2}{2\Sigma(E)^2}}}{\int dx P_0(x) e^{-\frac{(x-Y)^2}{2\Sigma(E)^2}}}\,, \tag{4.6}$$

where E is updated at each time instance t according to the recursion (4.10).

Natural performance measures are the "vector" and "matrix" MSE's of the AMP estimator defined below.

Definition 4.1 (Vector and matrix MSE of AMP)**.** *The vector and matrix MSE of the AMP estimator $\hat{\boldsymbol{S}}^{(t)}(\boldsymbol{W})$ at iteration t are defined respectively as follows*

$$\mathrm{Vmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) := \frac{1}{n}\mathbb{E}_{\boldsymbol{S},\boldsymbol{W}}\big[\|\hat{\boldsymbol{S}}^{(t)} - \boldsymbol{S}\|_2^2\big], \tag{4.7}$$

$$\mathrm{Mmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) := \frac{1}{n^2}\mathbb{E}_{\boldsymbol{S},\boldsymbol{W}}\big[\|\hat{\boldsymbol{S}}^{(t)}\hat{\boldsymbol{S}}^{(t)^\intercal} - \boldsymbol{S}\boldsymbol{S}^\intercal\|_{\mathrm{F}}^2\big], \tag{4.8}$$

where $\|A\|_{\mathrm{F}}^2 = \sum_{i,j} A_{ij}^2$ stands for the Frobenius norm of a matrix A.

A remarkable fact that follows from a general theorem of [69] (see [42] for its use in the matrix case) is that the state evolution sequence tracks these two MSE's and thus allows to assess the performance of AMP. Consider the scalar denoising problem (4.5). Hence, the (scalar) mmse function associated to this problem reads

$$\mathrm{mmse}(\Sigma(E)^{-2}) := \mathbb{E}_{S,Y}[(S - \mathbb{E}[X|Y])^2]\,. \tag{4.9}$$

The *state evolution sequence* $E^{(t)}$, $t \in \mathbb{N}$ is defined as

$$E^{(t+1)} = \mathrm{mmse}(\Sigma(E^{(t)})^{-2}), \qquad E^{(0)} = v. \tag{4.10}$$

Since the mmse function is monotone decreasing (its argument has the dimension of a signal to noise ratio) it is easy to see that that $E^{(t)}$ is a decreasing non-negative sequence. Thus $\lim_{t\to+\infty} E^{(t)} := E^{(\infty)}$ exists. One of the basic results of [69], [42] is

$$\lim_{n\to+\infty} \mathrm{Vmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) = E^{(t)}, \ \lim_{n\to+\infty} \mathrm{Mmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) = v^2 - (v - E^{(t)})^2. \tag{4.11}$$

We note that the results in [69], [42] are stronger in the sense that the non-averaged algorithmic mean square errors are tracked by state evolution with probability one.

Note that when $\mathbb{E}[S] = 0$ then v is an unstable fixed point, and as such, state evolution "does not start", in other words we have $E^{(t)} = v$. While this is not really a problem when one runs AMP in practice, for analysis purposes one can circumvent this problem by slightly biasing P_0 and remove the bias at the end of the analysis. For simplicity, we always assume that P_0 is biased so that $\mathbb{E}[S]$ is not zero.

Assumption 4.1. *In this chapter we assume that P_0 is discrete with bounded support. Moreover, we assume that P_0 is biased such that $\mathbb{E}[S]$ is non-zero.*

A fundamental quantity computed by state evolution is the algorithmic threshold.

Definition 4.2 (AMP threshold)**.** *For $\Delta > 0$ small enough, the fixed point equation corresponding to* (4.10) *has a unique solution for all noise values in $]0, \Delta[$. We define Δ_{AMP} as the supremum of all such Δ.*

4.2.3 Spatially Coupled Model

The present spatially coupled construction is similar to the one used for the coupled Curie-Weiss model [79] and is also similar to mean-field spin glass systems introduced in [175, 176]. We consider a chain (or a ring) of underlying systems positioned at $\mu \in \{0, \dots, L\}$ and coupled to neighboring blocks $\{\mu - w, \dots, \mu + w\}$. Positions μ are taken modulo $L+1$ and the integer $w \in \{0, \dots, L/2\}$ equals the size of the *coupling window*. The coupled model is

$$w_{i_\mu j_\nu} = s_{i_\mu} s_{j_\nu} \sqrt{\frac{\Lambda_{\mu\nu}}{n}} + z_{i_\mu j_\nu} \sqrt{\Delta}, \tag{4.12}$$

where the index $i_\mu \in \{1, \dots, n\}$ (resp. j_ν) belongs to the block μ (resp. ν) along the ring, $\mathbf{\Lambda}$ is an $(L+1)\times(L+1)$ matrix which describes the strength of the coupling between blocks, and $Z_{i_\mu j_\nu} \sim \mathcal{N}(0,1)$ are i.i.d. For the analysis to work, the matrix elements have to be chosen appropriately. We assume that:

i) $\mathbf{\Lambda}$ is a doubly stochastic matrix;

ii) $\Lambda_{\mu\nu}$ depends on $|\mu - \nu|$;

iii) $\Lambda_{\mu\nu}$ is not vanishing for $|\mu - \nu| \leq w$ and vanishes for $|\mu - \nu| > w$;

iv) $\mathbf{\Lambda}$ is *smooth* in the sense $|\Lambda_{\mu\nu} - \Lambda_{\mu+1\nu}| = \mathcal{O}(w^{-2})$ and $\Lambda^* := \sup_{\mu,\nu} \Lambda_{\mu\nu} = \mathcal{O}(w^{-1})$;

v) $\mathbf{\Lambda}$ has a non-negative Fourier transform.

All these conditions can easily be met, the simplest example being a triangle of base $2w+1$ and height $1/(w+1)$, more precisely:

$$\Lambda_{\mu\nu} = \begin{cases} \frac{1}{w+1}\left(1 - \frac{|\mu-\nu|}{w+1}\right), & |\mu-\nu| \le w \\ 0, & |\mu-\nu| > w \end{cases} \tag{4.13}$$

We will always denote by $\mathcal{S}_\mu := \{\nu \,|\, \Lambda_{\mu\nu} \neq 0\}$ the set of $2w+1$ blocks coupled to block μ.

The construction of the coupled system is completed by introducing a *seed* in the ring: we assume perfect knowledge of the signal components $\{s_{i_\mu}\}$ for $\mu \in \mathcal{B} := \{-w, \dots, w-1\} \mod L$. This seed is what allows to close the gap between the algorithmic and information theoretic limits and therefore plays a crucial role. We sometimes refer to the seed as the *pinning construction*. Note that the seed can also be viewed as an "opening" of the chain with fixed boundary conditions.

AMP has been applied for the rank-one symmetric matrix estimation problems by [164], who also computed the state evolution formula to analyze its performance, generalizing techniques developed by [69] and [115]. State evolution was further studied by [166] and [42].

The AMP algorithm and the state evolution recursion [166, 42] can be easily adapted to the spatially coupled model as done in Section 4.4. The proof that the state evolution for the symmetric rank-one matrix estimation problem tracks the AMP on a spatially coupled model is an extension of the analysis done in [166, 42] for the uncoupled model. The full re-derivation of such result would be lengthy and beyond the scope of our analysis. We thus assume that state evolution tracks the AMP performance for our coupled problem. However, we believe that the proof will be similar to the one done for the spatially coupled compressed sensing problem [115]. This assumption is vindicated numerically.

Assumption 4.2. *We consider the spatially coupled model* (4.12) *with* P_0 *satisfying Assumption 4.1. We assume that state evolution tracks the AMP algorithm for this model.*

4.2.4 Main Results: Basic Underlying Model

One of our central results is a proof of the expression for the asymptotic mutual information per variable via the so-called *replica symmetric (RS) potential*. This is the function $E \in [0, v] \mapsto i_{\rm RS}(E; \Delta) \in \mathbb{R}$ defined as

$$i_{\rm RS}(E;\Delta) := \frac{(v-E)^2 + v^2}{4\Delta} - \mathbb{E}_{S,Z}\left[\ln\left(\int dx\, P_0(x) e^{-\frac{x^2}{2\Sigma(E)^2} + x\left(\frac{S}{\Sigma(E)^2} + \frac{Z}{\Sigma(E)}\right)}\right)\right], \tag{4.14}$$

with $Z \sim \mathcal{N}(0,1)$, $S \sim P_0$. Most of our results will assume that P_0 is a discrete distribution over a finite bounded real alphabet $P_0(s) = \sum_{\alpha=1}^{\nu} p_\alpha \delta(s - a_\alpha)$ (see

Assumption 4.1). Thus the only continuous integral in (4.14) is the Gaussian over Z. The extension to mixtures of continuous and discrete signals can be obtained by approximation methods not discussed in this chapter (see e.g. the methods in [172]).

It turns out that both the information theoretic and algorithmic AMP thresholds are determined by the set of stationary points of (4.14) (w.r.t E). It is possible to show that for all $\Delta > 0$ there always exist at least one stationary minimum.[4] In this contribution we suppose that at most three stationary points exist, corresponding to situations with at most one phase transition as depicted in Fig. 4.1 (see Assumption 4.3 below). Situations with multiple transitions could also be covered by our techniques.

Assumption 4.3. *We assume that P_0 is such that there exist at most three stationary points for the potential* (4.14).

Remark 4.1. *An important property of the replica symmetric potential is that the stationary points satisfy the state evolution fixed point equation. In other words $\partial i_{\mathrm{RS}}/\partial E = 0$ implies $E = \mathrm{mmse}(\Sigma(E)^{-2})$ and conversely. Moreover it is not difficult to see that the Δ_{AMP} is given by the smallest solution of $\partial i_{\mathrm{RS}}/\partial E = \partial^2 i_{\mathrm{RS}}/\partial E^2 = 0$. In other words the AMP threshold is the "first" horizontal inflexion point appearing in $i_{\mathrm{RS}}(E;\Delta)$ when Δ increases from 0 to $+\infty$.*

One of the main results of this chapter is formulated in the following theorem which provides a proof of the conjectured single-letter formula for the asymptotic mutual information per variable.

Theorem 4.1 (RS formula for the mutual information)**.** *Fix $\Delta > 0$ and let P_0 satisfy Assumptions 4.1-4.3. Then*

$$\lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{S};\boldsymbol{W}) = \min_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta). \tag{4.15}$$

Proof. See Section 4.6. □

The proof of the *existence of the limit* does not require the above hypothesis on P_0. Also, it was first shown in [47] that

$$\limsup_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W}) \leq \min_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta), \tag{4.16}$$

an inequality that *we will use* in the proof section. Note that, interestingly, and perhaps surprisingly, the analysis of [47] leads to a sharp upper bound on the "free energy" for all finite n. We will make extensive use of this inequality and for sake of completeness, we summarize its proof in Appendix 4.8.1.

Theorem 4.1 allows to compute the information theoretic phase transition threshold which we define in the following way.

[4] Note $E=0$ is never a stationary point (except for the trivial case of P_0 a single Dirac mass which we exclude from the discussion) and $E = v$ is stationary only if $\mathbb{E}[S] = 0$.

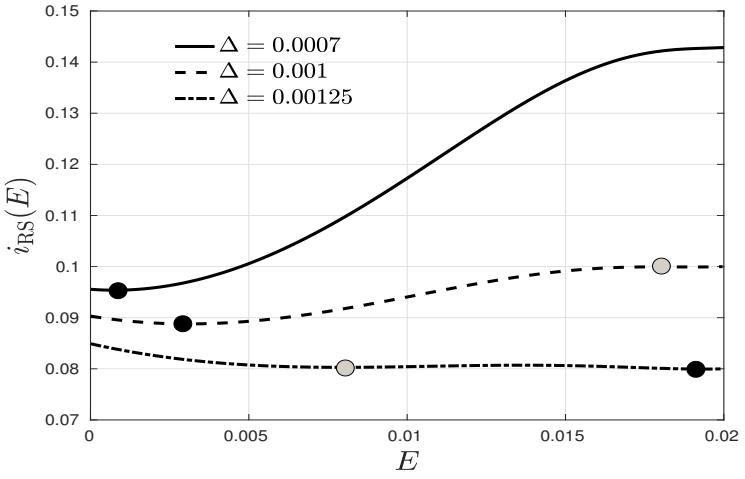

Figure 4.1: The replica symmetric potential $i_{\mathrm{RS}}(E)$ for three values of Δ in the spiked Wigner model with $S_i \sim \mathrm{Ber}(\rho)$. The normalized mutual information is $\min i_{\mathrm{RS}}(E)$ (the black dots, while the gray dots correspond to the local minima). The asymptotic MMSE is $\mathrm{argmin} i_{\mathrm{RS}}(E)$, where $v = \rho$ in this case with $\rho = 0.02$. The curves from top to bottom: i) For low noise values, here $\Delta = 0.0007 < \Delta_{\mathrm{AMP}}$, there exists a unique "good" minimum corresponding to the MMSE and AMP is Bayes-optimal. ii) As the noise increases, a second local "bad" minimum appears: this is the situation at $\Delta_{\mathrm{AMP}} < \Delta = 0.0011 < \Delta_{\mathrm{RS}}$. iii) For $\Delta = 0.00125 > \Delta_{\mathrm{RS}}$, the "bad" minimum becomes the global one and the MMSE suddenly deteriorates. AMP can be seen as a naive minimizer of this curve starting from $E = v = 0.02$.

Definition 4.3 (Information theoretic or optimal threshold)**.** *Define* Δ_{opt} *as the first non-analyticity point of the mutual information as* Δ *increases. Formally*

$$\Delta_{\mathrm{opt}} := \sup\{\Delta | \lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{S}; \boldsymbol{W}) \text{ is analytic in }]0, \Delta[\}. \tag{4.17}$$

The information theoretic threshold is also called "optimal threshold" because we expect $\Delta_{\mathrm{AMP}} \leq \Delta_{\mathrm{opt}}$. This is indeed proven in Lemma 4.12.

When P_0 is s.t the RS potential (4.14) has at most three stationary points, then $\min_{E\in[0,v]} i_{\mathrm{RS}}(E; \Delta)$ has at most one *non-analyticity* point denoted Δ_{RS} (see Fig. 4.1). In case of analyticity over all $\mathbb{R}_+$, we set $\Delta_{\mathrm{RS}} = \infty$. We call Δ_{RS} the RS or potential threshold. Theorem 4.1 gives us a mean to concretely *compute* the information theoretic threshold: $\Delta_{\mathrm{opt}} = \Delta_{\mathrm{RS}}$.

From Theorem 4.1 we will also deduce the expressions for the *vector* MMSE and the *matrix* MMSE defined below.

Definition 4.4 (Vector and matrix MMSE)**.** *The vector and matrix MMSE are defined respectively as follows*

$$\mathrm{Vmmse}_n(\Delta^{-1}) := \frac{1}{n} \mathbb{E}_{\boldsymbol{S},\boldsymbol{W}} \Big[\big\| \boldsymbol{S} - \mathbb{E}[\boldsymbol{X}|\, \boldsymbol{W}] \big\|_2^2 \Big]. \tag{4.18}$$

$$\mathrm{Mmmse}_n(\Delta^{-1}) := \frac{1}{n^2}\mathbb{E}_{\boldsymbol{S},\boldsymbol{W}}\Big[\big\|\boldsymbol{S}\boldsymbol{S}^\intercal - \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\intercal|\,\boldsymbol{W}]\big\|_{\mathrm{F}}^2\Big]. \tag{4.19}$$

The conditional expectation $\mathbb{E}[\,\cdot\,|\mathbf{W}]$ in Definition 4.4 is w.r.t the posterior distribution

$$P(\mathbf{x}|\mathbf{w}) = \frac{1}{\tilde{\mathcal{Z}}(\mathbf{w})} e^{-\frac{1}{2\Delta}\sum_{i\le j}\left(\frac{x_i x_j}{\sqrt{n}} - w_{ij}\right)^2} \prod_{i=1}^n P_0(x_i), \tag{4.20}$$

with the normalizing factor depending on the observation given by

$$\tilde{\mathcal{Z}}(\mathbf{w}) = \int \Big\{\prod_{i=1}^n dx_i P_0(x_i)\Big\} e^{-\frac{1}{2\Delta}\sum_{i\le j}\left(\frac{x_i x_j}{\sqrt{n}} - w_{ij}\right)^2} \tag{4.21}$$

The expectation $\mathbb{E}_{\mathbf{S},\mathbf{W}}[\,\cdot\,]$ is the one w.r.t $P(\mathbf{w})P(\mathbf{s}) = \tilde{\mathcal{Z}}(\mathbf{w})\prod_{i=1}^n P_0(s_i)$. The expressions for the MMSE's in terms of (4.14) are given in the following corollary.

Corollary 4.1 (Exact formula for the MMSE)**.** *For all $\Delta \neq \Delta_{\mathrm{RS}}$, the matrix MMSE is asymptotically*

$$\lim_{n\to\infty} \mathrm{Mmmse}_n(\Delta^{-1}) = v^2 - (v - \mathrm{argmin}_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta))^2. \tag{4.22}$$

Moreover, if $\Delta < \Delta_{\mathrm{AMP}}$ or $\Delta > \Delta_{\mathrm{RS}}$, then the usual vector MMSE satisfies

$$\lim_{n\to\infty} \mathrm{Vmmse}_n(\Delta^{-1}) = \mathrm{argmin}_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta). \tag{4.23}$$

Proof. See Section 4.7. □

It is natural to conjecture that, in the *whole* range $\Delta \neq \Delta_{\mathrm{RS}}$, the vector MMSE is given by $\mathrm{argmin}_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta)$, but our proof does not quite yield the full statement.

Another fundamental consequence of Theorem 4.1 concerns the optimality of the performance of AMP.

Corollary 4.2 (Optimality of AMP)**.** *For $\Delta < \Delta_{\mathrm{AMP}}$ or $\Delta > \Delta_{\mathrm{RS}}$, the AMP is asymptotically optimal in the sense that it yields upon convergence the asymptotic vector-MMSE and matrix-MMSE of Corollary 4.1. Namely,*

$$\lim_{t\to+\infty}\lim_{n\to+\infty} \mathrm{Mmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) = \lim_{n\to\infty} \mathrm{Mmmse}_n(\Delta^{-1}). \tag{4.24}$$

$$\lim_{t\to+\infty}\lim_{n\to+\infty} \mathrm{Vmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) = \lim_{n\to\infty} \mathrm{Vmmse}_n(\Delta^{-1}). \tag{4.25}$$

On the other hand, for $\Delta_{\mathrm{AMP}} < \Delta < \Delta_{\mathrm{RS}}$ the AMP algorithm is strictly suboptimal, namely

$$\lim_{t\to+\infty}\lim_{n\to+\infty} \mathrm{Mmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) > \lim_{n\to\infty} \mathrm{Mmmse}_n(\Delta^{-1}). \tag{4.26}$$

$$\lim_{t\to+\infty}\lim_{n\to+\infty} \mathrm{Vmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) > \lim_{n\to\infty} \mathrm{Vmmse}_n(\Delta^{-1}). \tag{4.27}$$

Proof. See Section 4.7. □

This leaves the region $\Delta_{\mathrm{AMP}} < \Delta < \Delta_{\mathrm{RS}}$ algorithmically open for efficient polynomial time algorithms. A natural conjecture, backed up by many results in spin glass theory, coding theory, planted models and the planted clique problems, is:

Conjecture 4.1. *For $\Delta_{\mathrm{AMP}} < \Delta < \Delta_{\mathrm{RS}}$, no polynomial time efficient algorithm that outperforms AMP exists.*

4.2.5 Main Results: Coupled Model

In this chapter, the spatially coupled construction is used for the purposes of the proof. However, one can also imagine interesting applications of the spatially coupled estimation problem, specially in view of the fact that AMP turns out to be optimal for the spatially coupled system. In coding theory for example, the use of spatially coupled systems as a proof device historically followed their initial construction which was for engineering purposes and led to the construction of capacity achieving codes.

Our first crucial result states that the mutual information of the coupled and original systems are the same in a suitable limit. The mutual information of the coupled system of length L and with coupling window w is denoted $I_{w,L}(\mathbf{S}; \mathbf{W})$.

Theorem 4.2 (Equality of mutual informations)**.** *For any fixed w s.t. P_0 satisfies Assumption 4.1, the following limits exist and are equal*

$$\lim_{L\to\infty} \lim_{n\to\infty} \frac{1}{n(L+1)} I_{w,L}(\boldsymbol{S}; \boldsymbol{W}) = \lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{S}; \boldsymbol{W}). \tag{4.28}$$

Proof. See Section 4.5. □

An immediate corollary is that the non-analyticity points (w.r.t Δ) of the mutual informations are the same in the coupled and underlying models. In particular, define the optimal threshold of the spatially coupled model defined by $\Delta^{\mathrm{c}}_{\mathrm{opt}} := \sup\{\Delta \mid \lim_{L\to\infty} \lim_{n\to\infty} I_{w,L}(\mathbf{S}; \mathbf{W})/(n(L+1)) \text{ is analytic in }]0, \Delta[\}$, we have $\Delta^{\mathrm{c}}_{\mathrm{opt}} = \Delta_{\mathrm{opt}}$.

The second crucial result states that the AMP threshold of the spatially coupled system is at least as good as Δ_{RS} (threshold saturation result of Theorem 4.3). The analysis of AMP applies to the coupled system as well [69, 115] and it can be shown that the performance of AMP is assessed by SE. Let

$$E^{(t)}_{\mu} := \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\|\mathbf{S}_{\mu} - \hat{\mathbf{S}}^{(t)}_{\mu}\|_2^2] \tag{4.29}$$

be the asymptotic average vector-MSE of the AMP estimate $\hat{\mathbf{S}}^{(t)}_{\mu}$ at time t for the μ-th "block" of $\mathbf{S}$. We associate to each position $\mu \in \{0, \dots, L\}$ an

independent scalar system with AWGN of the form $y = s + \Sigma_\mu(\mathbf{E}; \Delta) z$, with

$$\Sigma_\mu(\mathbf{E})^{-2} := \frac{v - \sum_{\nu=0}^{L} \Lambda_{\mu\nu} E_\nu}{\Delta} \tag{4.30}$$

and $S \sim P_0$, $Z \sim \mathcal{N}(0,1)$. Taking into account knowledge of the signal components in the seed $\mathcal{B}$, SE reads

$$\begin{cases} E_\mu^{(t+1)} = \mathrm{mmse}(\Sigma_\mu(\mathbf{E}^{(t)}; \Delta)^{-2}), \ E_\mu^{(0)} = v \text{ for } \mu \in \{0, \dots, L\} \setminus \mathcal{B}, \\ E_\mu^{(t)} = 0 \text{ for } \mu \in \mathcal{B}, t \geq 0 \end{cases} \tag{4.31}$$

where the mmse function is defined as in (4.9).

From the monotonicity of the mmse function we have $E_\mu^{(t+1)} \leq E_\mu^{(t)}$ for all $\mu \in \{0, \dots, L\}$, a partial order which implies that $\lim_{t\to\infty} \mathbf{E}^{(t)} = \mathbf{E}^{(\infty)}$ exists. This allows to define an algorithmic threshold for the coupled system on a finite chain:

$$\Delta_{\mathrm{AMP},w,L} := \sup\{\Delta | E_\mu^{(\infty)} \leq E_{\mathrm{good}}(\Delta) \ \forall \ \mu\}$$

where $E_{\mathrm{good}}(\Delta)$ is the trivial fixed point solution of the SE starting with the initial condition $E^{(0)} = 0$. A more formal but equivalent definition of $\Delta_{\mathrm{AMP},w,L}$ is given in Section 4.4.

Theorem 4.3 (Threshold saturation)**.** *Let $\Delta_{\mathrm{AMP}}^{\mathrm{c}}$ be the algorithmic threshold on an infinite chain, $\Delta_{\mathrm{AMP}}^{\mathrm{c}} := \liminf_{w\to\infty} \liminf_{L\to\infty} \Delta_{\mathrm{AMP},w,L}$, s.t. P_0 satisfies Assumptions 4.1 and 4.2. We have $\Delta_{\mathrm{AMP}}^{\mathrm{c}} \geq \Delta_{\mathrm{RS}}$.*

Proof. See Section 4.4. □

Our techniques also allow to prove the equality $\Delta_{\mathrm{AMP}}^{\mathrm{c}} = \Delta_{\mathrm{RS}}$, but this is not directly needed.

4.2.6 Roadmap of the Proof of the Replica Symmetric Formula

Here we give a roadmap of the proof of Theorem 4.1 that will occupy Sections 4.4–4.6. A fruitful idea is to concentrate on the question whether $\Delta_{\mathrm{opt}} = \Delta_{\mathrm{RS}}$. The proof of this equality automatically generates the proof of Theorem 4.1.

We first prove in Section 4.6.1 that $\Delta_{\mathrm{opt}} \leq \Delta_{\mathrm{RS}}$. This proof is based on a joint use of the I-MMSE relation (Lemma 4.9), the replica bound (4.16) and the suboptimality of the AMP algorithm. In the process of proving $\Delta_{\mathrm{opt}} \leq \Delta_{\mathrm{RS}}$, we in fact get as a direct bonus the proof of Theorem 4.1 for $\Delta < \Delta_{\mathrm{opt}}$.

The proof of $\Delta_{\mathrm{opt}} \geq \Delta_{\mathrm{RS}}$ requires the use of spatial coupling. The main strategy is to show

$$\Delta_{\mathrm{RS}} \leq \Delta_{\mathrm{AMP}}^{\mathrm{c}} \leq \Delta_{\mathrm{opt}}^{\mathrm{c}} = \Delta_{\mathrm{opt}}. \tag{4.32}$$

The first inequality in (4.32) is proven in Section 4.4 using methods first invented in coding theory: The algorithmic AMP threshold of the spatially

coupled system $\Delta^{\mathrm{c}}_{\mathrm{AMP}}$ saturates (tends in a suitable limit) towards Δ_{RS}, i.e. $\Delta_{\mathrm{RS}} \leq \Delta^{\mathrm{c}}_{\mathrm{AMP}}$ (Theorem 4.3). To prove the (last) equality we show in Section 4.5 that the free energies, and hence the mutual informations, of the underlying and spatially coupled systems are equal in a suitable asymptotic limit (Theorem 4.2). This implies that their non-analyticities occur at the same point and hence $\Delta^{\mathrm{c}}_{\mathrm{opt}} = \Delta_{\mathrm{opt}}$. This is done through an interpolation which, although similar in spirit, is different than the one used to prove replica bounds (e.g. (4.16)). In the process of showing $\Delta^{\mathrm{c}}_{\mathrm{opt}} = \Delta_{\mathrm{opt}}$, we will also derive the existence of the limit for $I(\mathbf{S};\mathbf{W})/n$. Finally, the second inequality is due the suboptimality of the AMP algorithm. This follows by a direct extension of the SE analysis of [166, 42] to the spatially coupled case as done in [115].

Once $\Delta_{\mathrm{opt}} = \Delta_{\mathrm{RS}}$ is established it is easy to put everything together and conclude the proof of Theorem 4.1. In fact all that remains is to prove Theorem 4.1 for $\Delta > \Delta_{\mathrm{opt}}$. This follows by an easy argument in section 4.6.2 which combines $\Delta_{\mathrm{opt}} = \Delta_{\mathrm{RS}}$, the replica bound (4.16) and the suboptimality of the AMP algorithm. Note that in the proof sections that follow, we assume that Assumptions 4.1-4.3 hold.

4.2.7 Connection with the Planted SK Model

Let us briefly discuss the connection of the matrix factorization problem (4.1) with a statistical mechanical spin glass model which is a variant of the classic Sherrington-Kirkpatrick (SK) model. This is also the occasion to express the mutual information as a "free energy" through a simple relation that will be used in various guises later on.

Replacing $w_{ij} = n^{-1/2} s_i s_j + \sqrt{\Delta} z_{ij}$ in (4.20) and simplifying the fraction after expanding the squares, the posterior distribution can be expressed in terms of $\mathbf{s}, \mathbf{z}$ as follows

$$P(\mathbf{x}|\mathbf{s},\mathbf{z}) = \frac{1}{\mathcal{Z}} e^{-\mathcal{H}(\mathbf{x}|\mathbf{s},\mathbf{z})} \prod_{i=1}^{n} P_0(x_i), \tag{4.33}$$

where

$$\mathcal{H}(\mathbf{x}|\mathbf{s},\mathbf{z}) = \sum_{i \leq j = 1}^{n} \Big(\frac{x_i^2 x_j^2}{2n\Delta} - \frac{s_i s_j x_i x_j}{n\Delta} - \frac{z_{ij} x_i x_j}{\sqrt{n\Delta}} \Big) \tag{4.34}$$

and

$$\mathcal{Z} = \int \Big\{ \prod_{i=1}^{n} dx_i P_0(x_i) \Big\} e^{-\mathcal{H}(\mathbf{x}|\mathbf{s},\mathbf{z})}. \tag{4.35}$$

In the language of statistical mechanics, (4.34) is the "Hamiltonian", (4.35) is the "partition function", and (4.33) is the Gibbs distribution. This distribution is random since it depends on the realizations of $\mathbf{S}$, $\mathbf{Z}$. Conditional expectations with respect to (4.33) are denoted by the Gibbs "bracket" $\langle - \rangle$. More precisely

$$\mathbb{E}_{\mathbf{X}|\mathbf{S},\mathbf{Z}} \big[A(\mathbf{X}) | \mathbf{S} = \mathbf{s}, \mathbf{Z} = \mathbf{z} \big] = \langle A(\mathbf{X}) \rangle. \tag{4.36}$$

The free energy is defined as

$$f_n = -\frac{1}{n}\mathbb{E}_{\mathbf{S},\mathbf{Z}}[\ln \mathcal{Z}]. \tag{4.37}$$

Notice the difference between $\tilde{\mathcal{Z}}$ in (4.20) and $\mathcal{Z}$ in (4.33). The former is the partition function with a complete square, whereas the latter is the partition function that we obtain after expanding the square and simplifying the posterior distribution.

In Appendix 4.8.2, we show that mutual information and free energy are essentially the same object up to a trivial term. For the present model

$$\frac{1}{n}I(\mathbf{S};\mathbf{W}) = -\frac{1}{n}\mathbb{E}_{\mathbf{S},\mathbf{Z}}[\ln \mathcal{Z}] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n}(2\mathbb{E}[S^4] - v^2), \tag{4.38}$$

where recall $v = \mathbb{E}[S^2]$. This relationship turns out to be very practical and will be used several times.

For binary signals we have s_i and $x_i \in \{-1, +1\}$, so the model is a binary spin glass model. The first term in the Hamiltonian is a trivial constant, the last term corresponds exactly to the SK model with random Gaussian interactions, and the second term can be interpreted as an external random field that biases the spins. This is sometimes called a "planted" SK model.

The rest of the chapter is organized as follows. In Section 4.3 we provide two examples of the symmetric rank-one matrix estimation problem. Threshold saturation and the invariance of the mutual information due the spatial coupling are shown in Section 4.4 and 4.5 respectively. The proof of Theorem 4.1 follows in Section 4.6. Section 4.7 is dedicated to the proof of Corollary 4.1 and Corollary 4.2.

4.3 Two Examples: Spiked Wigner and Community Detection

In order to illustrate our results, we shall present them here in the context of two examples: the spiked Wigner model, where we close a conjecture left open by [166], and the case of asymmetric community detection.

4.3.1 Spiked Wigner Model

The first model is defined as follows: we are given data distributed according to the spiked Wigner model where the vector $\mathbf{s}$ is assumed to be a Bernoulli 0/1 random variable with probability ρ. Data then consists of a sparse, rank-one matrix observed through a Gaussian noise. In [166], the authors proved that, for $\rho > 0.041$, AMP is a computationally efficient algorithm that asymptotically achieves the information theoretically optimal mean-square error for *any* value of the noise Δ.

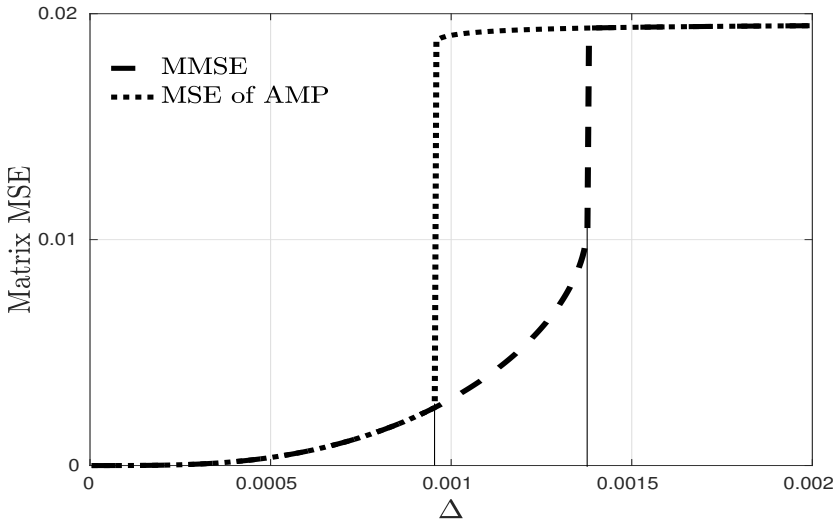

Figure 4.2: Phase transition diagram for spiked Wigner model with $\rho = 0.02$. The matrix MSE is shown as a function of the noise variance Δ. AMP provably achieves the MMSE except in the region $\Delta_{\rm AMP} < \Delta < \Delta_{\rm opt}$. We conjecture that no polynomial-time algorithm will do better than AMP in this region.

For very small densities (i.e. when ρ is $o(1)$), there is a well known large gap between what is information theoretically possible and what is tractable with current algorithms in support recovery [177]. This gap is actually related to the planted clique problem [178, 179], where it is believed that no polynomial algorithm is able to achieve information theoretic performances. It is thus perhaps not surprising that the situation for $\rho < 0.041$ becomes a bit more complicated. This is discussed in details in [167] on the basis of statistical physics consideration which we now prove.

For such values of ρ, as Δ changes there is a region where two local minima appears in $i_{\rm RS}(E;\Delta)$ (see Fig. 4.1 and the RS formula (4.14)). In particular for $\Delta_{\rm AMP} < \Delta < \Delta_{\rm opt}$, the global minimum differs from the AMP one and a computational gap appears (see Fig. 4.2). Interestingly, in this problem, the region where AMP is Bayes optimal is still quite large. Moreover, AMP algorithm was shown to outperform the spectral method [102, 153].

The region where AMP is not Bayes optimal is perhaps the most interesting one. While this is by no means evident, statistical physics analogies with actual phase transition in nature suggest that this region will be hard for a very large class of algorithms. A fact that add credibility to this prediction is the following: when looking to small ρ regime, we find that both the information theoretic threshold *and* the AMP one corresponds to what has been predicted in sparse PCA for sub-extensive values of ρ [177].

Finally, another interesting line of work for such probabilistic models has appeared in the context of random matrix theory (see for instance [44] and references therein). The focus is to analyze the limiting distribution of the eigenvalues of the observed matrix. The typical picture that emerges from

this line of work is that a sharp phase transition occurs at a well-defined critical value of the noise. Above the threshold an outlier eigenvalue (and the principal eigenvector corresponding to it) has a positive correlation with the hidden signal. Below the threshold, however, the spectral distribution of the observation is indistinguishable from that of the pure random noise. In this model, this happens at $\Delta_{\text{spectral}} = \rho^2$. Note that for $\Delta > \Delta_{\text{spectral}}$ spectral methods are not able to distinguish data coming from the model from random ones, while AMP is able to sort (partly) data from noise for any values of Δ and ρ.

4.3.2 Asymmetric Community Detection

The second model is a problem of detecting two communities (groups) with different sizes ρn and $(1-\rho)n$, that generalizes the one considered in [42]. One is given a graph where the probability to have a link between nodes in the first group is $p+\mu(1-\rho)/(\rho\sqrt{n})$, between those in the second group is $p+\mu\rho/(\sqrt{n}(1-\rho))$, while interconnections appear with probability $p-\mu/\sqrt{n}$. With this peculiar "balanced" setting, the nodes in each group have the same degree distribution with mean pn, making them harder to distinguish.

According to the universality property described in Section 4.2, this is equivalent to the AWGN model (4.1) with variance $\Delta = p(1-p)/\mu^2$ where each variable s_i is chosen according to

$$P_0(s) = \rho\delta(s - \sqrt{(1-\rho)/\rho}) + (1-\rho)\delta(s + \sqrt{\rho/(1-\rho)}). \tag{4.39}$$

Our results for this problem[5] are summarized in [102, 153] where a phase transition behavior similar to that of Fig. 4.2 appears. Moreover, it was shown that for ρ grater than a critical value $\rho_c = 1/2 - \sqrt{1/12}$, it is asymptotically information theoretically *possible* to get an estimation better than chance if and only if $\Delta < 1$. When $\rho < \rho_c$, however, it becomes possible for much larger values of the noise. Interestingly, AMP and spectral methods have the same transition and can find a positive correlation with the hidden communities for $\Delta < 1$, regardless of the value of ρ.

4.4 Threshold Saturation

The main goal of this section is to prove that for a proper spatially coupled (SC) system, threshold saturation occurs (Theorem 4.3), that is $\Delta_{\text{RS}} \leq \Delta^{\text{c}}_{\text{AMP}}$. We begin with some preliminary formalism in Sections 4.4.1 and 4.4.2 on state evolution for the underlying and coupled systems. Note that the proof of threshold saturations done in this section is similar to that of sparse superposition codes in Chapter 2.

[5]Note that here since $E = v = 1$ is an extremum of $i_{\text{RS}}(E;\Delta)$, one must introduce a small bias in P_0 and let it then tend to zero at the end of the proofs.

4.4.1 State Evolution of the Underlying System

First, define the following posterior average

$$\langle A \rangle := \frac{\int dx A(x) P_0(x) e^{-\frac{x^2}{2\Sigma(E,\Delta)^2} + x\left(\frac{s}{\Sigma(E,\Delta)^2} + \frac{z}{\Sigma(E,\Delta)}\right)}}{\int dx P_0(x) e^{-\frac{x^2}{2\Sigma(E,\Delta)^2} + x\left(\frac{s}{\Sigma(E,\Delta)^2} + \frac{z}{\Sigma(E,\Delta)}\right)}}, \tag{4.40}$$

where $S \sim P_0$, $Z \sim \mathcal{N}(0,1)$. The dependence on these variables, as well as on Δ and E is implicit and dropped from the notation of $\langle A \rangle$. Let us define the following operator.

Definition 4.5 (SE operator)**.** *The state evolution operator associated with the underlying system is*

$$T_{\mathrm{u}}(E) := \mathrm{mmse}(\Sigma(E)^{-2}) = \mathbb{E}_{S,Z}[(S - \langle X \rangle)^2], \tag{4.41}$$

where $S \sim P_0$, $Z \sim \mathcal{N}(0,1)$.

The fixed points of this operator play an important role. They can be viewed as the stationary points of the replica symmetric potential function $E \in [0, v] \mapsto i_{\mathrm{RS}}(E; \Delta) \in \mathbb{R}$ or equivalently of $f^{\mathrm{u}}_{\mathrm{RS}} : E \in [0, v] \mapsto f^{\mathrm{u}}_{\mathrm{RS}}(E) \in \mathbb{R}$ where

$$f^{\mathrm{u}}_{\mathrm{RS}}(E) := i_{\mathrm{RS}}(E; \Delta) - \frac{v^2}{4\Delta}. \tag{4.42}$$

It turns out to be more convenient to work with $f^{\mathrm{u}}_{\mathrm{RS}}$ instead of i_{RS}. We have

Lemma 4.1. *Any fixed point of the SE corresponds to a stationary point of $f^{\mathrm{u}}_{\mathrm{RS}}$:*

$$E = T_{\mathrm{u}}(E) \Leftrightarrow \frac{\partial f^{\mathrm{u}}_{\mathrm{RS}}(E; \Delta)}{\partial E}\Big|_E = 0. \tag{4.43}$$

Proof. See Appendix 4.8.4. □

The asymptotic performance of the AMP algorithm can be tracked by iterating the SE recursion as follows (this is the same as equation (4.10) expressed here with the help of T_{u})

$$E^{(t+1)} = T_{\mathrm{u}}(E^{(t)}), \quad t \geq 0, \quad E^{(0)} = v, \tag{4.44}$$

where the iteration is initialized without any knowledge about the signal other than its prior distribution (in fact, both the asymptotic vector and matrix MSE of the AMP are tracked by the previous recursion as reviewed in Section 4.2.2). Let $E_{\mathrm{good}}(\Delta) = T^{(\infty)}_{\mathrm{u}}(0)$, the fixed point reached by initializing iterations at $E = 0$. With our hypothesis on P_0 it is not difficult to see that definition 4.2 is equivalent to

$$\Delta_{\mathrm{AMP}} := \sup \left\{ \Delta > 0 \,|\, T^{(\infty)}_{\mathrm{u}}(v) = E_{\mathrm{good}}(\Delta) \right\}. \tag{4.45}$$

The following definition is handy

Definition 4.6 (Bassin of attraction)**.** *The basin of attraction of the good solution* $E_{\text{good}}(\Delta)$ *is* $\mathcal{V}_{\text{good}} := \{E \,|\, T_{\text{u}}^{(\infty)}(E) = E_{\text{good}}(\Delta)\}$.

Finally, we introduce the notion of *potential gap*. This is a function $\delta f_{\text{RS}}^{\text{u}} : \Delta \in \mathbb{R}_+ \mapsto \delta f_{\text{RS}}^{\text{u}}(\Delta) \in \mathbb{R}$ defined as follows:

Definition 4.7 (Potential gap)**.** *Define*

$$\delta f_{\text{RS}}^{\text{u}}(\Delta) := \inf_{E \notin \mathcal{V}_{\text{good}}} (f_{\text{RS}}^{\text{u}}(E) - f_{\text{RS}}^{\text{u}}(E_{\text{good}})) \tag{4.46}$$

as the potential gap, with the convention that the infimum over the empty set is ∞ *(this happens for* $\Delta < \Delta_{\text{AMP}}$ *where the complement of* $\mathcal{V}_{\text{good}}$ *is the empty set).*

Our hypothesis on P_0 imply that

$$\Delta_{\text{RS}} := \sup \{\Delta > 0 \,|\, \delta f_{\text{RS}}^{\text{u}}(\Delta) > 0\} \tag{4.47}$$

4.4.2 State Evolution of the Coupled System

For the SC system, the performance of the AMP decoder is tracked by an *MSE profile* (or just *profile*) $\mathbf{E}^{(t)}$, defined componentwise by

$$E_\mu^{(t)} = \lim_{n \to +\infty} \frac{1}{n} \mathbb{E}_{\mathbf{S},\mathbf{W}} \|\mathbf{S}_\mu - \hat{\mathbf{S}}_\mu^{(t)}(\mathbf{W})\|_2^2. \tag{4.48}$$

It has $L+1$ components and describes the scalar MSE in each block μ. Let us introduce the SE associated with the AMP algorithm for the inference over this SC system. First, denote the following posterior average at fixed s, z and Δ.

$$\langle A \rangle_\mu := \frac{\int dx A(x) P_0(x) e^{-\frac{x^2}{2\Sigma_\mu(\mathbf{E},\Delta)^2} + x\left(\frac{s}{\Sigma_\mu(\mathbf{E},\Delta)^2} + \frac{z}{\Sigma_\mu(\mathbf{E},\Delta)}\right)}}{\int dx P_0(x) e^{-\frac{x^2}{2\Sigma_\mu(\mathbf{E},\Delta)^2} + x\left(\frac{s}{\Sigma_\mu(\mathbf{E},\Delta)^2} + \frac{z}{\Sigma_\mu(\mathbf{E},\Delta)}\right)}}. \tag{4.49}$$

where the *effective noise variance* of the SC system is defined as

$$\Sigma_\mu(\mathbf{E})^{-2} := \frac{v - \sum_{\nu \in \mathcal{S}_\mu} \Lambda_{\mu\nu} E_\nu}{\Delta}, \tag{4.50}$$

where we recall $\mathcal{S}_\mu := \{\nu \,|\, \Lambda_{\mu\nu} \neq 0\}$ is the set of $2w+1$ blocks coupled to block μ.

Definition 4.8 (SE operator of the coupled system)**.** *The state evolution operator associated with the coupled system (4.12) is defined component-wise as*

$$[T_{\text{c}}(\boldsymbol{E})]_\mu := \mathbb{E}_{S,Z}[(S - \langle X \rangle_\mu)^2]. \tag{4.51}$$

$T_{\text{c}}(\boldsymbol{E})$ *is vector valued and here we have written its* μ*-th component.*

We assume perfect knowledge of the variables $\{s_{i_\mu}\}$ inside the blocks $\mu \in \mathcal{B} := \{0 : w-1\} \cup \{L-w : L\}$ as mentioned in Section 4.2.3, that is $x_{i_\mu} = s_{i_\mu}$ for all i_μ such that $\mu \in \mathcal{B}$. This implies $E_\mu = 0 \; \forall \; \mu \in \mathcal{B}$. We refer to this as the *pinning condition*. The SE iteration tracking the scalar MSE profile of the SC system reads for $\mu \notin \mathcal{B}$

$$E_\mu^{(t+1)} = [T_{\rm c}(\mathbf{E}^{(t)})]_\mu \quad \forall \; t \geq 0, \tag{4.52}$$

with the initialization $E_\mu^{(0)} = v$. For $\mu \in \mathcal{B}$, the pinning condition forces $E_\mu^{(t)} = 0 \; \forall \; t$. This equation is the same as (4.31) but is expressed here in terms of the operator Tc.

Let us introduce a suitable notion of degradation that will be very useful for the analysis.

Definition 4.9 (Degradation)**.** *A profile $\boldsymbol{E}$ is degraded (resp. strictly degraded) w.r.t another one $\boldsymbol{G}$, denoted as $\boldsymbol{E} \succeq \boldsymbol{G}$ (resp. $\boldsymbol{E} \succ \boldsymbol{G}$), if $E_\mu \geq G_\mu \; \forall \; \mu$ (resp. if $\boldsymbol{E} \succeq \boldsymbol{G}$ and there exists some μ such that $E_\mu > G_\mu$).*

Define an error profile $\mathbf{E}_{\rm good}(\Delta)$ as the vector with all $L+1$ components equal to $E_{\rm good}(\Delta)$.

Definition 4.10 (AMP threshold of coupled ensemble)**.** *The AMP threshold of the coupled system is defined as*

$$\Delta_{\rm AMP}^{\rm c} := \liminf_{w,L\to\infty} \sup \{\Delta > 0 \,|\, T_{\rm c}^{(\infty)}(\boldsymbol{v}) \prec \boldsymbol{E}_{\rm good}(\Delta)\} \tag{4.53}$$

where $\boldsymbol{v}$ is the all v vector. The $\liminf_{w,L\to\infty}$ is taken along sequences where first $L \to \infty$ and then $w \to \infty$. We also set for a finite system $\Delta_{{\rm AMP},w,L} := \sup \{\Delta > 0 \,|\, T_{\rm c}^{(\infty)}(\boldsymbol{v}) \prec \boldsymbol{E}_{\rm good}(\Delta)\}$.

The proof presented in the next subsection uses extensively the following monotonicity properties of the SE operators.

Lemma 4.2. *The SE operator of the SC system maintains degradation in space, i.e. $\boldsymbol{E} \succeq \boldsymbol{G} \Rightarrow T_{\rm c}(\boldsymbol{E}) \succeq T_{\rm c}(\boldsymbol{G})$. This property is verified by $T_{\rm u}(E)$ for a scalar error as well.*

Proof. From (4.50) one can immediately see that $\mathbf{E} \succeq \mathbf{G} \Rightarrow \Sigma_\mu(\mathbf{E}) \geq \Sigma_\mu(\mathbf{G}) \; \forall \; \mu$. Now, the SE operator (4.51) can be interpreted as the mmse function associated to the Gaussian channel $y = s + \Sigma_\mu(\mathbf{E}, \Delta) z$. This is an increasing function of the noise intensity Σ_μ^2: this is intuitively clear but we provide an explicit formula for the derivative below. Thus $[T_{\rm c}(\mathbf{E})]_\mu \geq [T_{\rm c}(\mathbf{G})]_\mu \; \forall \; \mu$, which means $T_{\rm c}(\mathbf{E}) \succeq T_{\rm c}(\mathbf{G})$.

The derivative of the mmse function of the Gaussian channel can be computed as

$$\frac{d \, {\rm mmse}(\Sigma^{-2})}{d \, \Sigma^{-2}} = -2 \, \mathbb{E}_{X,Y} \Big[\|X - \mathbb{E}[X|Y]\|_2^2 \, {\rm Var}[X|Y] \Big]. \tag{4.54}$$

This formula explicitly confirms that $T_{\rm u}(E)$ (resp. $[T_{\rm c}(\mathbf{E})]_\mu$) is an increasing function of Σ^2 (resp. Σ_μ^2). □

Corollary 4.3. *The SE operator of the coupled system maintains degradation in time, i.e.,* $T_{\mathrm{c}}(\boldsymbol{E}^{(t)}) \preceq \boldsymbol{E}^{(t)} \Rightarrow T_{\mathrm{c}}(\boldsymbol{E}^{(t+1)}) \preceq \boldsymbol{E}^{(t+1)}$. *Similarly* $T_{\mathrm{c}}(\boldsymbol{E}^{(t)}) \succeq \boldsymbol{E}^{(t)} \Rightarrow T_{\mathrm{c}}(\boldsymbol{E}^{(t+1)}) \succeq \boldsymbol{E}^{(t+1)}$. *Furthermore, the limiting error profile* $\boldsymbol{E}^{(\infty)} := T_{\mathrm{c}}^{(\infty)}(\boldsymbol{E}^{(0)})$ *exists. These properties are verified by* $T_{\mathrm{u}}(E)$ *as well.*

Proof. The degradation statements are a consequence of Lemma 4.2. The existence of the limits follows from the monotonicity of the operator and boundedness of the scalar MSE. □

Finally we will also need the following generalization of the (replica symmetric) potential function to a spatially coupled system:

$$\begin{aligned} f_{\mathrm{RS}}^{\mathrm{c}}(\mathbf{E}) = &\sum_{\mu=0}^{L} \sum_{\nu \in \mathcal{S}_\mu} \frac{\Lambda_{\mu\nu}}{4\Delta}(v - E_\mu)(v - E_\nu) \\ &- \sum_{\mu=0}^{L} \mathbb{E}_{S,Z}\Big[\ln\Big(\int dx P_0(x) e^{-\frac{1}{2\Sigma_\mu(\mathbf{E})^2}\left(x^2 - 2xS + xZ\Sigma_\mu(\mathbf{E},\Delta)\right)}\Big)\Big], \end{aligned} \tag{4.55}$$

where $Z \sim \mathcal{N}(z|0,1)$ and $S \sim P_0(s)$. As for the underlying system, the following Lemma links the SE and RS formulations.

Lemma 4.3. *If* $\boldsymbol{E}$ *is a fixed point of* (4.52)*, i.e.* $E_\mu = [T_{\mathrm{c}}(\boldsymbol{E})]_\mu \Rightarrow \frac{\partial f_{\mathrm{RS}}^{\mathrm{c}}(\boldsymbol{E})}{\partial E_\mu}\big|_{\boldsymbol{E}} = 0 \ \forall \ \mu \in \mathcal{B}^{\mathrm{c}} = \{w : L - w - 1\}$.

Proof. The proof is similar to the proof of Lemma 4.1 in Appendix 4.8.4. We skip the details for brevity. □

Now that we have settled the required definitions and properties, we can prove threshold saturation.

4.4.3 Proof of Threshold Saturation (Theorem 4.3)

The proof will proceed by contradiction. Let $\mathbf{E}^*$ a fixed point profile of the SE iteration (4.52). We suppose that $\mathbf{E}^*$ *does not* satisfy $\mathbf{E}^* \prec \mathbf{E}_{\mathrm{good}}(\Delta)$, and exhibit a contradiction for $\Delta < \Delta_{\mathrm{RS}}$ and w large enough (but independent of L). Thus we must have $\mathbf{E}^* \prec \mathbf{E}_{\mathrm{good}}(\Delta)$. This is the statement of Theorem 4.4 in Section 4.4.3 and directly implies Theorem 4.3.

The pinning condition together with the monotonicity properties of the coupled SE operator (Lemma 4.2 and Corollary 4.3) ensure that any fixed point profile $\mathbf{E}^*$ which does not satisfy $\mathbf{E}^* \prec \mathbf{E}_{\mathrm{good}}(\Delta)$ necessarily has a shape as described in Fig. 4.3. We construct an associated *saturated profile* $\mathbf{E}$ as described in Fig. 4.3. From now on we work with a *saturated profile* $\mathbf{E}$ which verifies $\mathbf{E} \succeq \mathbf{E}^*$ and $\mathbf{E} \succeq \mathbf{E}_{\mathrm{good}}(\Delta)$. In the following we will need the following operator.

Definition 4.11 (Shift operator)**.** *The shift operator* S *is defined componentwise as* $[\mathrm{S}(\boldsymbol{E})]_\mu := E_{\mu-1}$.

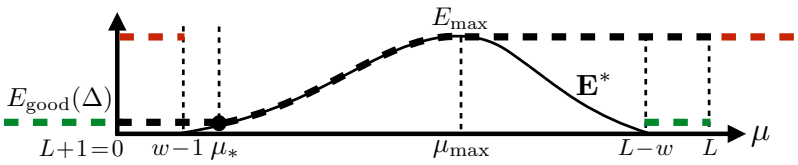

Figure 4.3: A fixed point profile $\mathbf{E}^*$ of the coupled SE iteration (solid line) is necessarily null $\forall\ \mu \in \{0 : w-1\}$ because of the pinning condition, and then it increases to reach $E_{\max} \in [0, v]$ at some $\mu_{\max} \in \{w : L-w-1\}$ (for a symmetric coupling matrix $\mu_{\max} = L/2$). Then, it starts decreasing and it is again null $\forall\ \mu \in \{L-w : L\}$. By definition, the associated *saturated* profile $\mathbf{E}$ (dashed line) starts at $E_{\text{good}}(\Delta)\ \forall\ \mu \leq \mu_*$, where μ_* is such that: $E^*_\mu \leq E_{\text{good}}(\Delta)\ \forall\ \mu \in \{0 : \mu_*\}$ and $E^*_{\mu'} > E_{\text{good}}(\Delta)\ \forall\ \mu' \in \{\mu_*+1 : L\}$. Then, $\mathbf{E}$ matches $\mathbf{E}^*\ \forall\ \mu \in \{\mu_* : \mu_{\max}\}$ and saturates at $E_{\max}\ \forall\ \mu \geq \mu_{\max}$. The saturated profile is extended for $\mu < 0$ and $\mu > L$ indices. The green (resp. red) branch shows that when the block indices of $\mathbf{E}$ are $\mu < 0$ (resp. $\mu > L$), then E_μ equals to $E_{\text{good}}(\Delta)$ (resp. $E_{\max}$). By construction, $\mathbf{E}$ is non decreasing in μ and is degraded w.r.t the fixed point profile $\mathbf{E} \succeq \mathbf{E}^*$.

Upper bound on the potential variation under a shift

The first step in the proof of threshold saturation is based on the Taylor expansion of the RS free energy of the SC system.

Lemma 4.4. *Let $\boldsymbol{E}$ be a saturated profile. Set $\boldsymbol{E}_\lambda := (1-\lambda)\boldsymbol{E} + \lambda \mathrm{S}(\boldsymbol{E})$ for $\lambda \in [0,1]$ and $\delta E_\mu := E_\mu - E_{\mu-1}$. There exists some $\lambda \in [0,1]$ such that*

$$f^{\mathrm{c}}_{\mathrm{RS}}(\mathrm{S}(\boldsymbol{E})) - f^{\mathrm{c}}_{\mathrm{RS}}(\boldsymbol{E}) = \frac{1}{2}\sum_{\mu,\mu'=0}^{L} \delta E_\mu \delta E_{\mu'} \frac{\partial^2 f^{\mathrm{c}}_{\mathrm{RS}}(\boldsymbol{E})}{\partial E_\mu \partial E_{\mu'}}\Big|_{\boldsymbol{E}_\lambda}. \tag{4.56}$$

Proof. Using the remainder Theorem, the free energy difference can be expressed as

$$f^{\mathrm{c}}_{\mathrm{RS}}(\mathrm{S}(\mathbf{E})) - f^{\mathrm{c}}_{\mathrm{RS}}(\mathbf{E}) = -\sum_{\mu=0}^{L} \delta E_\mu \frac{\partial f^{\mathrm{c}}_{\mathrm{RS}}(\mathbf{E})}{\partial E_\mu}\Big|_{\mathbf{E}} + \frac{1}{2}\sum_{\mu,\mu'=0}^{L} \delta E_\mu \delta E_{\mu'} \frac{\partial^2 f^{\mathrm{c}}_{\mathrm{RS}}(\mathbf{E})}{\partial E_\mu \partial E_{\mu'}}\Big|_{\mathbf{E}_\lambda}. \tag{4.57}$$

for some $\lambda \in [0,1]$. By definition of the saturated profile $\mathbf{E}$, we have $\delta E_\mu = 0\ \forall\ \mu \in \mathcal{A} := \{0 : \mu_*\} \cup \{\mu_{\max}+1 : L\}$ and $E_\mu = [T_{\mathrm{c}}(\mathbf{E})]_\mu$ for $r \notin \mathcal{A}$. Recalling Lemma 4.3 we see that the derivative in the first sum cancels for $r \notin \mathcal{A}$. Hence the first sum in (4.57) vanishes. □

Lemma 4.5. *The saturated profile $\boldsymbol{E}$ is smooth, i.e. $\delta E^* := \max_\mu |\delta E_\mu| = \mathcal{O}(1/w)$ uniformly in L.*

Proof. By definition of the saturated profile $\mathbf{E}$ we have $\delta E_\mu = 0 \; \forall \; \mu \in \mathcal{A} := \{0 : \mu_*\} \cup \{\mu_{\max} + 1 : L\}$. For $\mu \notin \mathcal{A}$, we can replace the fixed point profile component E_μ by $[T_{\rm c}(\mathbf{E})]_\mu$ so that $\delta E_\mu = [T_{\rm c}(\mathbf{E})]_\mu - [T_{\rm c}(\mathbf{E})]_{\mu-1}$. We will Taylor expand the SE operator. To this end, we define $\delta\Sigma_\mu^{-2} := \Sigma_\mu(\mathbf{E})^{-2} - \Sigma_{\mu-1}(\mathbf{E})^{-2}$ for $\mu \in \{\mu^* + 1 : \mu_{\max}\}$. Recall that $\Lambda_{\mu-1,\nu-1} = \Lambda_{\mu\nu}$, $\Lambda_{\mu\nu} \geq 0$ and $\Lambda^* := \sup_{\mu,\nu} \Lambda_{\mu\nu} = \mathcal{O}(1/w)$. Thus from (4.50) we get

$$\begin{aligned} |\delta\Sigma_\mu^{-2}| =& \frac{1}{\Delta}\Big| \sum_{\nu\in\mathcal{S}_\mu} \Lambda_{\mu\nu}(E_\nu - E_{\nu-1})\Big| \\ &\leq \frac{\Lambda^*}{\Delta} \sum_{\nu\in\mathcal{S}_\mu} (E_\nu - E_{\nu-1}) \\ &\leq \frac{2v\Lambda^*}{\Delta} = \mathcal{O}(\frac{1}{w}) \end{aligned} \tag{4.58}$$

where we have used $E_\nu - E_{\nu-1} \geq 0$ to get rid of the absolute value. Note that the first and second derivatives of the SE operator (4.51) w.r.t Σ_μ^{-2} are bounded as long as the five first moments of the posterior (4.49) exist and are bounded (which is true under our assumptions). Then by Taylor expansion at first order in $\delta\Sigma_\mu^{-2}$ and using the remainder theorem, we obtain

$$|\delta E_\mu| = |[T_{\rm c}(\mathbf{E})]_\mu - [T_{\rm c}(\mathbf{E})]_{\mu-1}| \leq |\delta\Sigma_\mu^{-2}|\Big|\frac{\partial[T_{\rm c}(\mathbf{E})]_\mu}{\partial\Sigma_\mu^{-2}}\Big| + \mathcal{O}(\delta\Sigma_\mu^{-4}) \leq \mathcal{O}(\frac{1}{w}), \tag{4.59}$$

where the last inequality follows from (4.58). □

Proposition 4.1. *Let $\boldsymbol{E}$ be a saturated profile. Then for all $\Delta > 0$ there exists a constant $0 < C(\Delta) < +\infty$ independent of L such that*

$$|f_{\rm RS}^{\rm c}(\mathrm{S}(\boldsymbol{E})) - f_{\rm RS}^{\rm c}(\boldsymbol{E})| \leq \frac{C(\Delta)}{w}. \tag{4.60}$$

Proof. From Lemma 4.4, in order to compute the free energy difference between the shifted and non-shifted profiles, we need to compute the Hessian associated with this free energy. We have

$$\begin{aligned} \frac{\partial f_{\rm RS}^{\rm c}(\mathbf{E})}{\partial E_\mu} &= \sum_{\nu\in\mathcal{S}_\mu} \frac{\Lambda_{\mu,\nu}E_\nu}{2\Delta} - \frac{1}{2}\sum_\nu \frac{\partial\Sigma_\nu(\mathbf{E})^{-2}}{\partial E_\mu}[T_{\rm c}(\mathbf{E})]_\nu - \frac{v}{2\Delta} \\ &= \frac{1}{2\Delta}\Big(\sum_{\nu\in\mathcal{S}_\mu} \Lambda_{\mu,\nu}E_\nu + \sum_{\nu\in\mathcal{S}_\mu} \Lambda_{\mu,\nu}[T_{\rm c}(\mathbf{E})]_\nu - v\Big) \end{aligned} \tag{4.61}$$

and

$$\frac{\partial^2 f_{\rm RS}^{\rm c}(\mathbf{E})}{\partial E_\mu \partial E_{\mu'}} = \frac{1}{2\Delta}\Big(\Lambda_{\mu,\mu'}\mathbb{I}(\mu' \in \mathcal{S}_\mu) - \frac{1}{\Delta}\sum_{\nu\in\mathcal{S}_\mu\cap\mathcal{S}_{\mu'}} \Lambda_{\mu,\nu}\Lambda_{\mu',\nu}\frac{\partial[T_{\rm c}(\mathbf{E})]_\nu}{\partial\,\Sigma_\nu^{-2}}\Big). \tag{4.62}$$

We can now estimate the sum in the Lemma 4.4. The contribution of the first term on the r.h.s of (4.62) can be bounded as

$$\frac{1}{2\Delta}\Big|\sum_{\mu=0}^{L}\sum_{\mu'\in\mathcal{S}_\mu}\Lambda_{\mu,\mu'}\delta E_\mu\delta E_{\mu'}\Big| \leq \frac{\delta E^*\Lambda^*(2w+1)}{2\Delta}\Big|\sum_{\mu=0}^{L}\delta E_\mu\Big| \leq \mathcal{O}(\frac{1}{w}), \quad (4.63)$$

where we used the facts: $\delta E_\mu \geq 0$, the sum over $\mu = 0, \cdots, L$ is telescopic, $E_\mu \in [0, v]$, $\Lambda^* = \mathcal{O}(1/w)$ and $\delta E^* = (w^{-1})$ (Lemma 4.5). We now bound the contribution of the second term on the r.h.s of (4.62). Recall the first derivative w.r.t Σ_ν^{-2} of the SE operator is bounded uniformly in L. Call this bound $K = \mathcal{O}(1)$. We obtain

$$\begin{aligned}&\frac{1}{2\Delta^2}\Big|\sum_{\mu,\mu'=1}^{L}\delta E_\mu\delta E_{\mu'}\sum_{\nu\in\mathcal{S}_\mu\cap\mathcal{S}_{\mu'}}\Lambda_{\mu,\nu}\Lambda_{\mu',\nu}\frac{\partial[T_{\rm c}(\mathbf{E})]_\nu}{\partial\Sigma_\nu^{-2}}\Big| \\ &\leq \frac{K\Lambda^{*2}\delta E^*}{2\Delta^2}\Big|\sum_{\mu=1}^{L}\delta E_\mu\sum_{\mu'\in\{\mu-2w:\mu+2w\}}\mathrm{card}(\mathcal{S}_\mu\cap\mathcal{S}_{\mu'})\Big| \leq \mathcal{O}(\frac{1}{w}). \quad (4.64)\end{aligned}$$

The last inequality follows from the following facts: the sum over $\mu = 1, \cdots, L$ is telescopic, $\Lambda^* = \mathcal{O}(1/w)$, Lemma 4.5, and for any fixed μ the following holds

$$\sum_{\mu'\in\{\mu-2w:\mu+2w\}}\mathrm{card}(\mathcal{S}_\mu\cap\mathcal{S}_{\mu'}) = (2w+1)^2. \quad (4.65)$$

Finally, from (4.63), (4.64) and the triangle inequality we obtain

$$\frac{1}{2}\Big|\sum_{\mu,\mu'=1}^{L}\delta E_\mu\delta E_{\mu'}\frac{\partial^2 f_{\rm RS}^{\rm c}(\mathbf{E})}{\partial E_\mu\partial E_{\mu'}}\Big|_{\mathbf{E}_\lambda}\Big| = \mathcal{O}(\frac{1}{w}) \quad (4.66)$$

uniformly in L. Combining this result with Lemma 4.4 ends the proof. □

Lower bound on the potential variation under a shift

The second step in the proof is based on a direct evaluation of $f_{\rm RS}^{\rm c}(\mathrm{S}(E)) - f_{\rm RS}^{\rm c}(E)$. We first need the Lemma:

Lemma 4.6. *Let $\boldsymbol{E}$ be a saturated profile such that $\boldsymbol{E} \succ \boldsymbol{E}_{\rm good}(\Delta)$. Then $E_{\max} \notin \mathcal{V}_{\rm good}$.*

Proof. The fact that the error profile is non decreasing and the assumption that $\mathbf{E} \succ \mathbf{E}_{\rm good}(\Delta)$ imply that $E_{\max} > E_0 = E_{\rm good}(\Delta)$. Moreover, $E_{\max} \leq [T_{\rm c}(\mathbf{E})]_{\mu_{\max}} \leq T_{\rm u}(E_{\max})$ where the first inequality follows from $\mathbf{E} \succ \mathbf{E}^*$ and the monotonicity of $T_{\rm c}$, while the second comes from the fact that $\mathbf{E}$ is non decreasing. Combining these with the monotonicity of $T_{\rm u}$ gives $T_{\rm u}(E_{\max}) \geq E_{\max}$ which implies $T_{\rm u}^{(\infty)}(E_{\max}) \geq E_{\max} > E_{\rm good}(\Delta)$ which means $E_{\max} \notin \mathcal{V}_{\rm good}$. □

Proposition 4.2. *Fix $\Delta < \Delta_{\rm RS}$ and let $\boldsymbol{E}$ be a saturated profile such that $\boldsymbol{E} \succ \boldsymbol{E}_{\rm good}(\Delta)$. Then*

$$|f^{\rm c}_{\rm RS}(S(\boldsymbol{E})) - f^{\rm c}_{\rm RS}(\boldsymbol{E})| \geq \delta f^{\rm u}_{\rm RS}(\Delta) \tag{4.67}$$

where $\delta f^{\rm u}_{\rm RS}(\Delta)$ is the potential gap (Definition 4.7).

Proof. Set

$$\mathcal{I}(\Sigma) := \mathbb{E}_{S,Z}\Big[\ln\Big(\int dx P_0(x) e^{-\frac{1}{2\Sigma^2}\left(x^2 - 2Sx - 2\Sigma Z x\right)}\Big)\Big].$$

By (4.55)

$$\begin{aligned} f^{\rm c}_{\rm RS}(\mathrm{S}(\mathbf{E})) - f^{\rm c}_{\rm RS}(\mathbf{E}) &= \sum_{\mu=-1}^{L-1}\sum_{\nu\in\mathcal{S}_\mu}\frac{\Lambda_{\mu+1\nu+1}}{4\Delta}(v-E_\mu)(v-E_\nu) \\ &\quad - \sum_{\mu=0}^{L}\sum_{\nu\in\mathcal{S}_\mu}\frac{\Lambda_{\mu\nu}}{4\Delta}(v-E_\mu)(v-E_\nu) \\ &\quad - \sum_{\mu=0}^{L}\mathcal{I}(\Sigma_\mu(\mathrm{S}(\mathbf{E}))) + \sum_{\mu=0}^{L}\mathcal{I}(\Sigma_\mu(\mathbf{E})) \\ &= \sum_{\nu\in\mathcal{S}_{-1}}\frac{\Lambda_{\mu\nu}}{4\Delta}(v-E_\mu)(v-E_\nu) - \sum_{\nu\in\mathcal{S}_L}\frac{\Lambda_{\mu\nu}}{4\Delta}(v-E_\mu)(v-E_\nu) \\ &\quad - \mathcal{I}(\Sigma_{-1}(\mathbf{E})) + \mathcal{I}(\Sigma_L(\mathbf{E})), \end{aligned} \tag{4.68}$$

where we used $\Lambda_{\mu+1\nu+1} = \Lambda_{\mu\nu}$ implying also $\Sigma_\mu(\mathrm{S}(\mathbf{E})) = \Sigma_{\mu-1}(\mathbf{E})$ as seen from (4.50). Recall $\Sigma(E)^{-2} = (v-E)/\Delta$. Now looking at (4.50), one notices that thanks to the saturation of $\mathbf{E}$, $\Sigma_{-1}(\mathbf{E}) = \Sigma(E_0)$ where $E_0 = E_{\rm good}(\Delta)$ (see the green branch in Fig. 4.3), while $\Sigma_L(\mathbf{E}) = \Sigma(E_L)$ where $E_L = E_{\max}$ (see the red branch Fig. 4.3). Finally from (4.68), using that the coupling matrix is (doubly) stochastic and the saturation of $\mathbf{E}$

$$\begin{aligned} f^{\rm c}_{\rm RS}(\mathrm{S}(\mathbf{E})) - f^{\rm c}_{\rm RS}(\mathbf{E}) &= \Big[\frac{(v-E_0)^2}{4\Delta} - \mathcal{I}(\Sigma(E_{\rm good}(\Delta)))\Big] \\ &\quad - \Big[\frac{(v-E_L)^2}{4\Delta} - \mathcal{I}(\Sigma(E_L))\Big] \\ &= f^{\rm u}_{\rm RS}(E_{\rm good}) - f^{\rm u}_{\rm RS}(E_{\max}) \leq -\delta f^{\rm u}_{\rm RS}(\Delta), \end{aligned} \tag{4.69}$$

where we recognized the potential function of the underlying system since $f^{\rm u}_{\rm RS}(E;\Delta) = i_{\rm RS}(E;\Delta) - \frac{v^2}{4\Delta}$, whereas the last inequality is a direct application of Lemma 4.6 and Definition 4.7. Finally, using the positivity of $\delta f^{\rm u}_{\rm RS}(\Delta)$ for $\Delta < \Delta_{\rm RS}$, we obtain the desired result. □

End of proof of threshold saturation

We now have the necessary ingredients in order to prove threshold saturation.

Theorem 4.4 (Asymptotic performance of AMP for the coupled system)**.** *Fix $\Delta < \Delta_{\rm RS}$. Take a spatially coupled system with $w > C(\Delta)/\delta f^{\rm u}_{\rm RS}(\Delta)$ where $C(\Delta)$ is the constant in Proposition 4.1. Then any fixed point profile $\boldsymbol{E}^*$ of the coupled state evolution iteration* (4.52) *must satisfy $\boldsymbol{E}^* \prec \boldsymbol{E}_{\rm good}(\Delta)$.*

Proof. The proof is by contradiction. Fix $\Delta < \Delta_{\rm RS}$ and $w \geq C(\Delta)/\delta f^{\rm u}_{\rm RS}(\Delta)$. We assume there exists a fixed point profile which does not satisfy $\mathbf{E}^* \prec \mathbf{E}_{\rm good}(\Delta)$. Then we construct the associated saturated profile $\mathbf{E}$. This profile satisfies both statements of Propositions 4.1 and 4.2. Therefore we must have $\delta f^{\rm u}_{\rm RS}(\Delta) \leq C(\Delta)/w$ which contradicts the choice $w > C(\Delta)/\delta f^{\rm u}_{\rm RS}(\Delta)$. We conclude that $\mathbf{E}^* \prec \mathbf{E}_{\rm good}(\Delta)$ must be true. □

Theorem 4.3 is a direct corollary of Theorem 4.4 and Definition 4.10. Take some $\Delta_* < \Delta_{\rm RS}$ and choose $w > C(\Delta_*)/\delta f^{\rm u}_{\rm RS}(\Delta_*)$. Then we have $\Delta_{{\rm AMP},w,L} \geq \Delta_*$. Note that $\delta f^{\rm u}_{\rm RS}(\Delta_*) \to 0_+$ for $\Delta_* \to \Delta_{\rm RS}$. Thus Taking $L \to +\infty$ first and $w \to +\infty$ second we can make Δ_* as close to $\Delta_{\rm RS}$ as we wish. Therefore we obtain $\Delta^{\rm c}_{\rm AMP} := \liminf_{L,w\to+\infty} \Delta_{{\rm AMP},w,L} \geq \Delta_{\rm RS}$ where the limit is taken in the specified order.

4.5 Invariance of the Mutual Information Under Spatial Coupling

In this section we prove that the mutual information remains unchanged under spatial coupling in a suitable asymptotic limit (Theorem 4.2). We will compare the mutual informations of the four following variants of (4.12). In each case, the signal $\mathbf{s}$ has $n(L+1)$ i.i.d components.

- *The fully connected:* If we choose $w = L/2$ and a homogeneous coupling matrix with elements $\Lambda_{\mu,\nu} = (L+1)^{-1}$ in (4.12). This yields a homogeneous fully connected system equivalent to (4.1) with $n(L+1)$ instead of n variables. The associated mutual information per variable for fixed L and n is denoted by $i^{\rm con}_{n,L}$.

- *The SC pinned system:* This is the system studied in Section 4.4 to prove threshold saturation, with the pinning condition. In this case we choose $0 < w < L/2$. The coupling matrix $\mathbf{\Lambda}$ is any matrix that fulfills the requirements in Section 4.2.3 (the concrete example given there will do). The associated mutual information per variable is here denoted $i^{\rm cou}_{n,w,L}$. Note that $i^{\rm cou}_{n,w,L} = (n(L+1))^{-1} I_{w,L}(\mathbf{S};\mathbf{W})$

- *The periodic SC system:* This is the same SC system (with same coupling window and coupling matrix) but without the pinning condition. The

associated mutual information per variable at fixed L, w, n is denoted $i^{\text{per}}_{n,w,L}$.

- *The decoupled system:* This corresponds simply to $L+1$ identical and independent systems of the form (4.1) with n variables each. This is equivalent to periodic SC system with $w = 0$. The associated mutual information per variable is denoted $i^{\text{dec}}_{n,L}$. Note that $i^{\text{dec}}_{n,L} = n^{-1} I(\mathbf{S}; \mathbf{W})$.

Let us outline the proof strategy. In a first step, we use an interpolation method twice: first interpolating between the fully connected and periodic SC systems, and then between the decoupled and periodic SC systems. This will allow to sandwich the mutual information of the periodic SC system by those of the fully connected and decoupled systems respectively (see Lemma 4.7). In the second step, using again a similar interpolation and Fekete's theorem for superadditive sequences, we prove that the decoupled and fully connected systems have asymptotically the same mutual information (see Lemma 4.8 for the existence of the limit). From these results we deduce the proposition:

Proposition 4.3. *For any* $0 \leq w \leq L/2$

$$\lim_{n\to+\infty} i^{\text{per}}_{n,w,L} = \lim_{n\to+\infty} \frac{1}{n} I(\boldsymbol{S}; \boldsymbol{W}) \tag{4.70}$$

Proof. Lemma 4.8 implies that $\lim_{n\to+\infty} i^{\text{con}}_{n,L} = \lim_{n\to+\infty} i^{\text{dec}}_{n,L}$. One also notes that $i^{\text{dec}}_{n,L} = \frac{1}{n} I(\mathbf{S}; \mathbf{W})$. Thus the result follows from Lemma 4.7. □

In a third step an easy argument shows

Proposition 4.4. *Assume* P_0 *has finite first four moments. For any* $0 \leq w \leq L/2$

$$|i^{\text{per}}_{n,w,L} - i^{\text{cou}}_{n,w,L}| = \mathcal{O}(\frac{w}{L}) \tag{4.71}$$

Proof. See Appendix 4.8.7. □

Since $i^{\text{cou}}_{n,w,L} = (n(L+1))^{-1} I_{w,L}(\mathbf{S}; \mathbf{W})$, Theorem 4.2 is an immediate consequence of Propositions 4.3 and 4.4.

4.5.1 A Generic Interpolation

Let us consider two systems of same total size $n(L+1)$ with coupling matrices $\mathbf{\Lambda}^{(1)}$ and $\mathbf{\Lambda}^{(0)}$ supported on coupling windows w_1 and w_0 respectively. Moreover, we assume that the observations associated with the first system are corrupted by an AWGN equals to $\sqrt{\Delta/t}\mathbf{z}$ while the AWGN corrupting the second system is $\sqrt{\Delta/(1-t)}\mathbf{z}'$, where Z_{ij} and Z'_{ij} are two i.i.d. standard Gaussians and $t \in [0,1]$ is the *interpolation parameter*. The interpolating

inference problem has the form

$$\begin{cases} w_{i_\mu j_\nu} &= s_{i_\mu} s_{j_\nu} \sqrt{\frac{\Lambda^{(1)}_{\mu\nu}}{n}} + z_{i_\mu j_\nu} \sqrt{\frac{\Delta}{t}}, \\ w_{i_\mu j_\nu} &= s_{i_\mu} s_{j_\nu} \sqrt{\frac{\Lambda^{(0)}_{\mu\nu}}{n}} + z'_{i_\mu j_\nu} \sqrt{\frac{\Delta}{1-t}} \end{cases} \tag{4.72}$$

In this setting, at $t = 1$ the interpolated system corresponds to the first system as the noise is infinitely large in the second one and no information is available about it, while at $t = 0$ the opposite happens. The associated interpolating posterior distribution can be expressed as

$$P_t(\mathbf{x}|\mathbf{s}, \mathbf{z}, \mathbf{z}') = \frac{1}{\mathcal{Z}_{\text{int}}(t)} e^{-\mathcal{H}(t, \mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)}))} \prod_{\mu=0}^{L} \prod_{i_\mu=1}^{n} P_0(x_{i_\mu}) \tag{4.73}$$

where the "Hamiltonian" is $\mathcal{H}_{\text{int}}(t, \mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)}) := \mathcal{H}(t, \mathbf{\Lambda}^{(1)}) + \mathcal{H}(1-t, \mathbf{\Lambda}^{(0)})$ with[6]

$$\begin{aligned} \mathcal{H}(t, \mathbf{\Lambda}) := \frac{t}{\Delta} \sum_{\mu=0}^{L} \Lambda_{\mu,\mu} \sum_{i_\mu \le j_\mu} \left(\frac{x^2_{i_\mu} x^2_{j_\mu}}{2n} - \frac{s_{i_\mu} s_{j_\mu} x_{i_\mu} x_{j_\mu}}{n} - \frac{x_{i_\mu} x_{j_\mu} z_{i_\mu j_\mu} \sqrt{\Delta}}{\sqrt{nt\Lambda_{\mu,\mu}}} \right) \\ + \frac{t}{\Delta} \sum_{\mu=0}^{L} \sum_{\nu=\mu+1}^{\mu+w} \Lambda_{\mu,\nu} \sum_{i_\mu, j_\nu=1}^{n} \left(\frac{x^2_{i_\mu} x^2_{j_\nu}}{2n} - \frac{s_{i_\mu} s_{j_\nu} x_{i_\mu} x_{j_\nu}}{n} - \frac{x_{i_\mu} x_{j_\nu} z_{i_\mu j_\nu} \sqrt{\Delta}}{\sqrt{nt\Lambda_{\mu,\nu}}} \right). \end{aligned} \tag{4.74}$$

and $\mathcal{Z}_{\text{int}}(t)$ is the obvious normalizing factor, the "partition function". The posterior average with respect to (4.73) is denoted by the bracket notation $\langle - \rangle_t$. It is easy to see that the mutual information per variable (for the interpolating inference problem) can be expressed as

$$i_{\text{int}}(t) := -\frac{1}{n(L+1)} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\ln \mathcal{Z}_{\text{int}}(t)] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n(L+1)} (2\mathbb{E}[S^4] - v^2) \tag{4.75}$$

The aim of the interpolation method in the present context is to compare the mutual informations of the systems at $t = 1$ and $t = 0$. To do so, one uses the fundamental theorem of calculus

$$i_{\text{int}}(1) - i_{\text{int}}(0) = \int_0^1 dt \, \frac{d i_{\text{int}}(t)}{dt} \, . \tag{4.76}$$

and tries to determine the sign of the integral term.

[6]Note that since the SC system is defined on a ring, we can express the Hamiltonian in terms of forward coupling only.

We first prove that

$$4\Delta(L+1)\frac{di_{\text{int}}(t)}{dt} =$$
$$\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\Big[\Big\langle -\frac{1}{n^2}\Big(\sum_{\mu=0}^{L}\sum_{\nu=\mu-w_1}^{\mu+w_1}\Lambda^{(1)}_{\mu\nu}\sum_{i_\mu,j_\nu=1}^{n}X_{i_\mu}X_{j_\nu}S_{i_\mu}S_{j_\nu}+\sum_{\mu=0}^{L}\Lambda^{(1)}_{\mu\mu}\sum_{i_\mu=1}^{n}X^2_{i_\mu}S^2_{i_\mu}\Big)$$
$$+\frac{1}{n^2}\Big(\sum_{\mu=0}^{L}\sum_{\nu=\mu-w_0}^{\mu+w_0}\Lambda^{(0)}_{\mu\nu}\sum_{i_\mu,j_\nu=1}^{n}X_{i_\mu}X_{j_\nu}S_{i_\mu}S_{j_\nu}+\sum_{\mu=0}^{L}\Lambda^{(0)}_{\mu\mu}\sum_{i_\mu=1}^{n}X^2_{i_\mu}S^2_{i_\mu}\Big)\Big\rangle_t\Big], \quad (4.77)$$

where $\langle - \rangle_t$ denotes the expectation over the posterior distribution associated with the interpolated Hamiltonian $\mathcal{H}_{\text{int}}(t, \mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)})$. We start with a simple differentiation of the Hamiltonian w.r.t. t which yields

$$\frac{d}{dt}\mathcal{H}_{\text{int}}(t, \mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)}) = \frac{1}{\Delta}\big(\mathcal{A}(t, \mathbf{\Lambda}^{(\mathbf{1})}) - \mathcal{B}(t, \mathbf{\Lambda}^{(0)})\big),$$

where

$$\mathcal{A}(t, \mathbf{\Lambda}^{(1)}) = \sum_{\mu=0}^{L}\Lambda^{(1)}_{\mu\mu}\sum_{i_\mu\le j_\mu}\left(\frac{x^2_{i_\mu}x^2_{j_\mu}}{2n} - \frac{s_{i_\mu}s_{j_\mu}x_{i_\mu}x_{j_\mu}}{n} - \frac{x_{i_\mu}x_{j_\mu}z_{i_\mu j_\mu}\sqrt{\Delta}}{2\sqrt{nt\Lambda^{(1)}_{\mu\mu}}}\right)$$
$$+\sum_{\mu=0}^{L}\sum_{\nu=\mu+1}^{\mu+w_1}\Lambda^{(1)}_{\mu\nu}\sum_{i_\mu,j_\nu=1}^{n}\left(\frac{x^2_{i_\mu}x^2_{j_\nu}}{2n} - \frac{s_{i_\mu}s_{j_\nu}x_{i_\mu}x_{j_\nu}}{n} - \frac{x_{i_\mu}x_{j_\nu}z_{i_\mu j_\nu}\sqrt{\Delta}}{2\sqrt{nt\Lambda^{(1)}_{\mu\nu}}}\right)$$
$$\mathcal{B}(t, \mathbf{\Lambda}^{(0)}) = \sum_{\mu=0}^{L}\Lambda^{(0)}_{\mu\mu}\sum_{i_\mu\le j_\mu}\left(\frac{x^2_{i_\mu}x^2_{j_\mu}}{2n} - \frac{s_{i_\mu}s_{j_\mu}x_{i_\mu}x_{j_\mu}}{n} - \frac{x_{i_\mu}x_{j_\mu}z'_{i_\mu j_\mu}\sqrt{\Delta}}{2\sqrt{n(1-t)\Lambda^{(0)}_{\mu\mu}}}\right)$$
$$\sum_{\mu=0}^{L}\sum_{\nu=\mu+1}^{\mu+w_0}\Lambda^{(0)}_{\mu\nu}\sum_{i_\mu,j_\nu=1}^{n}\left(\frac{x^2_{i_\mu}x^2_{j_\nu}}{2n} - \frac{s_{i_\mu}s_{j_\nu}x_{i_\mu}x_{j_\nu}}{n} - \frac{x_{i_\mu}x_{j_\nu}z'_{i_\mu j_\nu}\sqrt{\Delta}}{2\sqrt{n(1-t)\Lambda^{(0)}_{\mu\nu}}}\right).$$

Using integration by parts with respect to the Gaussian variables Z_{ij}, Z'_{ij}, one gets

$$\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[Z_{i_\mu j_\nu}\langle X_{i_\mu}X_{j_\nu}\rangle_t] = \sqrt{\frac{t\Lambda_{\mu,\nu}}{n\Delta}}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\Big[\langle X^2_{i_\mu}X^2_{j_\nu}\rangle_t - \langle X_{i_\mu}X_{j_\nu}\rangle^2_t\Big] \quad (4.78)$$
$$\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[Z'_{i_\mu j_\nu}\langle X_{i_\mu}X_{j_\nu}\rangle_t] = \sqrt{\frac{(1-t)\Lambda^0_{\mu,\nu}}{n\Delta}}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\Big[\langle X^2_{i_\mu}X^2_{j_\nu}\rangle_t - \langle X_{i_\mu}X_{j_\nu}\rangle^2_t\Big]. \quad (4.79)$$

Moreover an application of the Nishimori identity (4.165) shows

$$\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_{i_\mu}X_{j_\nu}\rangle^2_t] = \mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_{i_\mu}X_{j_\nu}S_{i_\mu}S_{j_\nu}\rangle_t]. \quad (4.80)$$

Combining (4.77)-(4.80) and using the fact that the SC system defined on a ring satisfies

$$\sum_{\mu=0}^{L}\Lambda_{\mu\mu}\sum_{i_\mu\le j_\mu}x_{i_\mu}x_{j_\mu}s_{i_\mu}s_{j_\mu}+\sum_{\mu=0}^{L}\sum_{\nu=\mu+1}^{\mu+w}\Lambda_{\mu\nu}\sum_{i_\mu,j_\nu=1}^{n}x_{i_\mu}x_{j_\nu}s_{i_\mu}s_{j_\nu}=$$
$$\frac{1}{2}\sum_{\mu=0}^{L}\sum_{\nu=\mu-w}^{\mu+w}\Lambda_{\mu\nu}\sum_{i_\mu,j_\nu=1}^{n}x_{i_\mu}x_{j_\nu}s_{i_\mu}s_{j_\nu}+\frac{1}{2}\sum_{\mu=0}^{L}\Lambda_{\mu\mu}x_{i_\mu}^2s_{i_\mu}^2,$$

we obtain (4.77).

Now, define the *overlaps* associated to each block μ as

$$q_\mu:=\frac{1}{n}\sum_{i_\mu=1}^{n}X_{i_\mu}S_{i_\mu},\qquad \tilde{q}_\mu:=\frac{1}{n}\sum_{i_\mu=1}^{n}X_{i_\mu}^2S_{i_\mu}^2. \tag{4.81}$$

Hence, (4.77) can be rewritten as

$$\frac{di_{\text{int}}(t)}{dt}=\frac{1}{4\Delta(L+1)}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\Big[\Big\langle\mathbf{q}^\intercal\boldsymbol{\Lambda}^{(0)}\,\mathbf{q}-\mathbf{q}^\intercal\boldsymbol{\Lambda}^{(1)}\,\mathbf{q}$$
$$+\frac{1}{n}\left(\tilde{\mathbf{q}}^\intercal\text{diag}(\boldsymbol{\Lambda}^{(0)})-\tilde{\mathbf{q}}^\intercal\text{diag}(\boldsymbol{\Lambda}^{(1)})\right)\Big\rangle_t\Big], \tag{4.82}$$

where $\mathbf{q}^\intercal=[q_0\cdots q_L]$, $\tilde{\mathbf{q}}^\intercal=[\tilde{q}_0\cdots\tilde{q}_L]$ are row vectors and $\text{diag}(\boldsymbol{\Lambda})$ represents the column vector with entries $\{\Lambda_{\mu\mu}\}_{\mu=0}^{L}$. The coupling matrices $\boldsymbol{\Lambda}^{(1)},\boldsymbol{\Lambda}^{(0)}$ are real, symmetric, circulant (due to the periodicity of the ring) and thus can be diagonalized in the same Fourier basis. We have

$$\frac{di_{\text{int}}(t)}{dt}=\frac{1}{4\Delta(L+1)}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\Big[\Big\langle\hat{\mathbf{q}}^\intercal\left(\mathbf{D}^{(0)}-\mathbf{D}^{(1)}\right)\hat{\mathbf{q}}$$
$$+\frac{1}{n}\left(\tilde{\mathbf{q}}^\intercal\text{diag}(\boldsymbol{\Lambda}^{(0)})-\tilde{\mathbf{q}}^\intercal\text{diag}(\boldsymbol{\Lambda}^{(1)})\right)\Big\rangle_t\Big], \tag{4.83}$$

where $\widehat{\mathbf{q}}$ is the discrete Fourier transfrom of $\mathbf{q}$ and $\mathbf{D}^{(1)},\mathbf{D}^{(0)}$ are the diagonal matrices with the eigenvalues of $\boldsymbol{\Lambda}^{(1)},\boldsymbol{\Lambda}^{(0)}$. Since the coupling matrices are stochastic with non-negative Fourier transform, their largest eigenvalue equals 1 (and is associated to the 0-th Fourier mode) while the remaining eigenvalues are non-negative. These properties will be essential in the following paragraphs.

4.5.2 Applications

Our first application is

Lemma 4.7. *Let the coupling matrix Λ verify the requirements (i)-(v) in Section 4.2.3. The mutual informations of the decoupled, periodic SC and fully connected systems verify*

$$i_{n,L}^{\text{dec}}\le i_{n,w,L}^{\text{per}}\le i_{n,L}^{\text{con}}. \tag{4.84}$$

Proof. We start with the second inequality. We choose $\Lambda^{(1)}_{\mu\nu} = (L+1)^{-1}$ for the fully connected system at $t = 1$. This matrix has a unique eigenvalue equal to 1 and L degenerate eigenvalues equal to 0. Therefore it is clear that $\mathbf{D}^{(0)} - \mathbf{D}^{(1)}$ is positive semi-definite and $\hat{\mathbf{q}}^\intercal \big(\mathbf{D}^{(0)} - \mathbf{D}^{(1)}\big)\hat{\mathbf{q}} \geq 0$. Moreover notice that $\Lambda^{(0)}_{\mu\mu} = \Lambda_{00}$ is independent of L. Therefore for L large enough

$$\tilde{\mathbf{q}}^\intercal \mathrm{diag}(\boldsymbol{\Lambda}^{(0)}) - \tilde{\mathbf{q}}^\intercal \mathrm{diag}(\boldsymbol{\Lambda}^{(1)}) = \Big(\Lambda_{00} - \frac{1}{L+1}\Big) \sum_{\mu=0}^{L} \tilde{q}_\mu \geq 0. \tag{4.85}$$

Therefore we conclude that (4.83) is positive and from (4.76) $i^{\mathrm{con}}_{n,L} - i^{\mathrm{per}}_{n,w,L} \geq 0$. For the first inequality we proceed similarly, but this time we choose $\Lambda^{(1)}_{\mu\nu} = \delta_{\mu\nu}$ for the decoupled system which has all eigenvalues equal to 1. Therefore $\mathbf{D}^{(0)} - \mathbf{D}^{(1)}$ is negative semidefinite so $\hat{\mathbf{q}}^\intercal \big(\mathbf{D}^{(0)} - \mathbf{D}^{(1)}\big)\hat{\mathbf{q}} \leq 0$. Moreover this time

$$\tilde{\mathbf{q}}^\intercal \mathrm{diag}(\boldsymbol{\Lambda}^{(0)}) - \tilde{\mathbf{q}}^\intercal \mathrm{diag}(\boldsymbol{\Lambda}^{(1)}) = \Big(\Lambda_{00} - 1\Big) \sum_{\mu=0}^{L} \tilde{q}_\mu \leq 0 \tag{4.86}$$

because we necessarily have $0 \leq \Lambda^{(0)}_{00} \leq 1$. We conclude that (4.83) is negative and from (4.76) $i^{\mathrm{dec}}_{n,L} - i^{\mathrm{per}}_{n,w,L} \leq 0$. □

The second application is

Lemma 4.8. *Consider the mutual information of system* (4.1) *and set* $i_n = n^{-1} I(\boldsymbol{S}; \boldsymbol{W})$. *Consider also* i_{n_1} *and* i_{n_2} *the mutual informations of two systems of size* n_1 *and* n_2 *with* $n = n_1 + n_2$. *The sequence* $n i_n$ *is superadditive in the sense that*

$$n_1 i_{n_1} + n_2 i_{n_2} \leq n i_n. \tag{4.87}$$

Fekete's lemma then implies that $\lim_{n\to+\infty} i_n$ *exists.*

Proof. This is easily proven by following the generic interpolation method of Section 4.5.1 for a coupled system with two spatial positions (i.e. $L + 1 = 2$). We choose $\Lambda^{(0)}_{\mu\nu} = \delta_{\mu\nu}$, $\mu, \nu \in 0, 1$ for the "decoupled" system and $\Lambda^{(1)}_{\mu\nu} = 1/2$ for $\mu, \nu \in 0, 1$ for the "fully connected" system. This analysis is essentially identical to [95] where the existence of the thermodynamic limit of the free energy for the Sherrington-Kirkpatrick mean-field spin glass is proven. □

4.6 Proof of the Replica Symmetric Formula (Theorem 4.1)

In this section we provide the proof of the RS formula for the mutual information of the underlying model (Theorem 4.1) for $0 < \Delta \leq \Delta_{\mathrm{opt}}$ (Proposition 4.5)

and then for $\Delta \geq \Delta_{\text{opt}}$ (Proposition 4.6). For $0 < \Delta \leq \Delta_{\text{opt}}$ the proof directly follows form the I-MMSE relation Lemma 4.9, the replica bound (4.16) and the suboptimality of the AMP algorithm. In this interval the proof doesn't require spatial coupling. For $\Delta \geq \Delta_{\text{opt}}$ the proof uses the results of Sections 4.4 and 4.5 on the spatially coupled model.

Let us start with two preliminary lemmas. The first is an I-MMSE relation [48] adapted to the current matrix estimation problem.

Lemma 4.9. *Let P_0 has finite first four moments. The mutual information and the matrix-MMSE are related by*

$$\frac{1}{n}\frac{dI(\boldsymbol{S};\boldsymbol{W})}{d\Delta^{-1}} = \frac{1}{4}\text{Mmmse}_n(\Delta^{-1}) + \mathcal{O}(1/n). \tag{4.88}$$

Proof.

$$\begin{aligned}
\frac{1}{n}\frac{dI(\mathbf{S};\mathbf{W})}{d\Delta^{-1}} =& \frac{1}{2n^2}\mathbb{E}_{\mathbf{S},\mathbf{W}}\Big[\sum_{i\leq j}\big(S_iS_j - \mathbb{E}[X_iX_j|\mathbf{W}]\big)^2\Big] \\
=& \frac{1}{4n^2}\mathbb{E}_{\mathbf{S},\mathbf{W}}\Big[\big\|\mathbf{S}\mathbf{S}^\intercal - \mathbb{E}[\mathbf{X}\mathbf{X}^\intercal|\mathbf{W}]\big\|_{\text{F}}^2\Big] \\
&+ \frac{1}{4n^2}\sum_{i=1}^{n}\mathbb{E}_{\mathbf{S},\mathbf{W}}[(S_i^2 - \mathbb{E}[X_i^2|\mathbf{W}])^2] \\
=& \frac{1}{4}\text{Mmmse}_n(\Delta^{-1}) + \mathcal{O}(1/n),
\end{aligned} \tag{4.89}$$

The proof details for first equality are in Appendix 4.8.3. The second equality is obatined by completing the sum and accounting for the diagonal terms. The last equality is obtained from

$$\begin{aligned}
\mathbb{E}_{\mathbf{S},\mathbf{W}}[(S_i^2 - \mathbb{E}[X_i^2|\mathbf{W}])^2] &= \mathbb{E}[S_i^4] - 2\mathbb{E}_{S_i,\mathbf{W}}[S_i^2\mathbb{E}[X_i^2|\mathbf{W}]] + \mathbb{E}_{\mathbf{W}}[\mathbb{E}[X_i^2|\mathbf{W}]^2] \\
&= \mathbb{E}[S_i^4] - \mathbb{E}_{\mathbf{W}}[\mathbb{E}[X_i^2|\mathbf{W}]^2] \\
&\leq \mathbb{E}[S_i^4].
\end{aligned} \tag{4.90}$$

where we used the Nishimori identity $\mathbb{E}_{S_i,\mathbf{W}}[S_i^2\mathbb{E}[X_i^2|\mathbf{W}]] = \mathbb{E}_{\mathbf{W}}[\mathbb{E}[X_i^2|\mathbf{W}]^2]$ in the second equality (Appendix 4.8.4). □

Lemma 4.10. *The limit $\lim_{n\to+\infty} n^{-1}I(\boldsymbol{S};\boldsymbol{W})$ exists and is a concave, continuous, function of Δ.*

Proof. The existence of the limit is the statement of Lemma 4.8 in Sec. 4.5. The continuity follows from the concavity of the mutual information with respect to Δ^{-1}: because the limit of a sequence of concave functions remains concave, and thus it is continuous. To see the concavity notice that the first derivative of the mutual information w.r.t Δ^{-1} equals the matrix-MMSE (Lemma 4.9) and that the later cannot increase as a function of Δ^{-1}. □

4.6.1 Proof of Theorem 4.1 in the Low Noise Regime

Lemma 4.11. *Assume P_0 is a discrete distribution. Fix $\Delta < \Delta_{\rm AMP}$. The mutual information per variable is asymptotically given by the RS formula* (4.15).

Proof. By the suboptimality of the AMP algorithm we have

$$\mathrm{Mmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) \geq \mathrm{Mmmse}_n(\Delta^{-1}). \tag{4.91}$$

Taking limits in the order $\lim_{t\to+\infty} \limsup_{n\to+\infty}$ and using (4.11) we find

$$v^2 - (v - E^{(\infty)})^2 \geq \limsup_{n\to+\infty} \mathrm{Mmmse}_n(\Delta^{-1}). \tag{4.92}$$

Furthermore, by applying Lemma 4.9 we obtain

$$\frac{v^2 - (v - E^{(\infty)})^2}{4} \geq \limsup_{n\to+\infty} \frac{1}{n} \frac{dI(\mathbf{S};\mathbf{W})}{d\Delta^{-1}}. \tag{4.93}$$

Now, for $\Delta < \Delta_{\rm AMP}$ we have $E^{(\infty)} = E_{\rm good}(\Delta)$ which is the unique and hence *global* minimum of $i_{\rm RS}(E;\Delta)$ over $E \in [0, v]$. Moreover, for $\Delta < \Delta_{\rm AMP}$ we have that $E^{(\infty)}(\Delta)$ is continuously differentiable Δ^{-1} with locally bounded derivative. Thus

$$\begin{aligned}
\frac{d}{d\Delta^{-1}}\Big(\min_{E\in[0,v]} i_{\rm RS}(E;\Delta)\Big) &= \frac{di_{\rm RS}}{d\Delta^{-1}}(E^{(\infty)};\Delta) \\
&= \frac{\partial i_{\rm RS}}{\partial E}(E^{(\infty)};\Delta)\frac{dE^{(\infty)}}{d\Delta^{-1}} + \frac{\partial i_{\rm RS}}{\partial \Delta^{-1}}(E^{(\infty)};\Delta) \\
&= \frac{\partial i_{\rm RS}}{\partial \Delta^{-1}}(E^{(\infty)};\Delta) \\
&= \frac{(v - E^{(\infty)})^2 + v^2}{4} - \frac{\partial \mathbb{E}_{S,Z}[\cdots]}{\partial \Sigma^{-2}}\bigg|_{E^{(\infty)}} \frac{\partial \Sigma^{-2}}{\partial \Delta^{-1}}\bigg|_{E^{(\infty)}} \\
&= \frac{v^2 - (v - E^{(\infty)})^2}{4},
\end{aligned} \tag{4.94}$$

where $\mathbb{E}_{S,Z}[\cdots]$ is the expectation that appears in the RS potential (4.14). The third equality is obtained from

$$\frac{\partial \Sigma^{-2}}{\partial \Delta^{-1}}\bigg|_{E^{(\infty)}} = v - E^{(\infty)} \tag{4.95}$$

and

$$\frac{\partial \mathbb{E}_{S,Z}[\cdots]}{\partial \Sigma^{-2}}\bigg|_{E^{(\infty)}} = \frac{1}{2}(v - E^{(\infty)}). \tag{4.96}$$

This last identity immediately follows from $\frac{\partial i_{\rm RS}}{\partial E}\Big|_{E^{(\infty)}} = 0$. From (4.93) and (4.94)

$$\frac{d}{d\Delta^{-1}}(\min_{E\in[0,v]} i_{\rm RS}(E;\Delta)) \geq \limsup_{n\to+\infty} \frac{1}{n}\frac{dI(\mathbf{S};\mathbf{W})}{d\Delta^{-1}}, \tag{4.97}$$

which is equivalent to

$$\frac{d}{d\Delta}(\min_{E\in[0,v]} i_{\rm RS}(E;\Delta)) \leq \liminf_{n\to+\infty} \frac{1}{n}\frac{dI(\mathbf{S};\mathbf{W})}{d\Delta}. \tag{4.98}$$

We now integrate inequality (4.98) over an interval $[0,\Delta] \subset [0,\Delta_{\rm AMP}[$

$$\begin{aligned}\min_{E\in[0,v]} i_{\rm RS}(E;\Delta) - \min_{E\in[0,v]} i_{\rm RS}(E;0) &\leq \int_0^\Delta d\tilde{\Delta}\, \liminf_{n\to+\infty} \frac{1}{n}\frac{dI(\mathbf{S};\mathbf{W})}{d\tilde{\Delta}} \\ &\leq \liminf_{n\to+\infty} \int_0^\Delta d\tilde{\Delta}\, \frac{1}{n}\frac{dI(\mathbf{S};\mathbf{W})}{d\tilde{\Delta}} \\ &= \liminf_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W}) - H(S). \end{aligned} \tag{4.99}$$

The second inequality uses Fatou's Lemma and the last equality uses that for a discrete prior

$$\lim_{\Delta\to 0_+} I(\mathbf{S};\mathbf{W}) = H(\mathbf{S}) - \lim_{\Delta\to 0_+} H(\mathbf{S}|\mathbf{W}) = nH(S). \tag{4.100}$$

In Appendix 4.8.6 an explicit calculation shows that $\min_E i_{\rm RS}(E;0) = H(S)$. Therefore

$$\min_{E\in[0,v]} i_{\rm RS}(E;\Delta) \leq \liminf_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W}). \tag{4.101}$$

The final step combines inequality (4.101) with the replica bound (4.16) to obtain

$$\min_{E\in[0,v]} i_{\rm RS}(E;\Delta) \leq \liminf_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W}) \leq \limsup_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W}) \leq \min_{E\in[0,v]} i_{\rm RS}(E;\Delta). \tag{4.102}$$

This shows that the limit of the mutual information exists and is equal to the RS formula for $\Delta < \Delta_{\rm AMP}$. Note that in this proof we did not need the a-priori existence of the limit. □

Remark 4.2. *One can try to apply the same proof idea to the regime $\Delta > \Delta_{\rm RS}$. Equations* (4.91)-(4.98) *work out exactly in the same way because the AMP fixed point $E^{(\infty)}$ is a global minimum of $i_{\rm RS}(E;\Delta)$. Then when integrating on $]\Delta,+\infty[\subset[\Delta_{\rm RS},+\infty[$, one finds*

$$\limsup_{n\to+\infty} \frac{1}{n} I(\boldsymbol{S};\boldsymbol{W}) \leq \min_{E\in[0,v]} i_{\rm RS}(E;\Delta). \tag{4.103}$$

This essentially gives an alternative proof of (4.16) *for $\Delta > \Delta_{\rm RS}$.*

Lemma 4.12. *We necessarily have* $\Delta_{\rm AMP} \leq \Delta_{\rm opt}$.

Proof. Notice first that it not possible to have $\Delta_{\rm RS} < \Delta_{\rm AMP}$ because in the range $]0, \Delta_{\rm AMP}[$, as a function of E, the function $i_{\rm RS}(E;\Delta)$ has a unique stationary point. Since $\min_{E\in[0,v]} i_{\rm RS}(E;\Delta)$ is analytic for $\Delta < \Delta_{\rm RS}$, it is analytic for $\Delta < \Delta_{\rm AMP}$. Now we proceed by contradiction: suppose we would have $\Delta_{\rm AMP} \geq \Delta_{\rm opt}$. Lemma 4.11 asserts that $\lim_{n\to+\infty} n^{-1} I(\mathbf{S};\mathbf{W}) = \min_{E\in[0,v]} i_{\rm RS}(E;\Delta)$ for $\Delta < \Delta_{\rm AMP}$ thus we would have $\lim_{n\to+\infty} n^{-1} I(\mathbf{S};\mathbf{W})$ analytic at $\Delta_{\rm opt}$. This is a contradiction by definition of $\Delta_{\rm opt}$. □

Lemma 4.13. *We necessarily have* $\Delta_{\rm RS} \geq \Delta_{\rm opt}$.

Proof. If $\Delta_{\rm RS} = +\infty$ then we are done, so we suppose it is finite. The proof proceeds by contradiction: suppose $\Delta_{\rm RS} < \Delta_{\rm opt}$. So we assume $\Delta_{\rm RS} \in [\Delta_{\rm AMP}, \Delta_{\rm opt}[$ (in the previous lemma we showed that this must be the case). For $\Delta \in]0, \Delta_{\rm RS}[$ we have $\min_{E\in[0,v]} i_{\rm RS}(E;\Delta) = i_{\rm RS}(E_{\rm good}(\Delta);\Delta)$ which is an analytic function in this interval. By definition of $\Delta_{\rm opt}$, the function $\lim_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W})$ is analytic in $]0, \Delta_{\rm opt}[$. Therefore *both functions are analytic on* $]0, \Delta_{\rm RS}[$ and since by Lemma 4.11 they are equal for $]0, \Delta_{\rm AMP}[\subset]0, \Delta_{\rm RS}[$, *they must be equal on the whole range* $]0, \Delta_{\rm RS}[$. This implies that the two functions are equal at $\Delta_{\rm RS}$ because they are continuous. Explicitly,

$$\min_E i_{\rm RS}(E;\Delta) = \lim_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W})|_\Delta \qquad \forall\ \Delta \in]0, \Delta_{\rm RS}]. \tag{4.104}$$

Now, fix some $\Delta \in]\Delta_{\rm RS}, \Delta_{\rm opt}[$. Since this Δ is greater than $\Delta_{\rm RS}$ the fixed point of state evolution $E^{(\infty)}$ is also the global minimum of $i_{\rm RS}(E;\Delta)$. Hence exactly as in (4.91)-(4.98) we can show that for $\Delta \in]\Delta_{\rm RS}, \Delta_{\rm opt}[$, (4.98) is verified. This time, combining (4.16), (4.98) and the assumption $\Delta_{\rm RS} \in [\Delta_{\rm AMP}, \Delta_{\rm opt}[$, leads to a contradiction, and hence we must have $\Delta_{\rm RS} \geq \Delta_{\rm opt}$. To see explicitly how the contradiction appears, integrate (4.98) on $]\Delta_{\rm RS}, \Delta[\subset]\Delta_{\rm RS}, \Delta_{\rm opt}[$, and use Fatou's Lemma, to obtain

$$\begin{aligned}\min_E i_{\rm RS}(E;\Delta) - \min_E i_{\rm RS}(E;\Delta_{\rm RS}) &\leq \liminf_{n\to+\infty} \Big(\frac{1}{n} I(\mathbf{S};\mathbf{W})|_\Delta - \frac{1}{n} I(\mathbf{S};\mathbf{W})|_{\Delta_{\rm RS}} \Big) \\ &= \lim_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W})|_\Delta - \lim_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W})|_{\Delta_{\rm RS}}.\end{aligned} \tag{4.105}$$

From (4.104) and (4.16) we obtain $\min_E i_{\rm RS}(E;\Delta) = \lim_{n\to+\infty} \frac{1}{n} I(\mathbf{S};\mathbf{W})$ *when* $\Delta_{\rm AMP} \leq \Delta_{\rm RS} < \Delta < \Delta_{\rm opt}$. But from (4.104), this equality is also true for $0 < \Delta \leq \Delta_{\rm RS}$. So the equality is valid in the whole interval $]0, \Delta_{\rm opt}[$ and therefore $\min_E i_{\rm RS}(E;\Delta)$ is analytic at $\Delta_{\rm RS}$. But this is impossible by the definition of $\Delta_{\rm RS}$. □

Proposition 4.5. *Assume* P_0 *is a discrete distribution. Fix* $\Delta \leq \Delta_{\rm opt}$. *The mutual information per variable is asymptotically given by the RS formula* (4.15).

Proof. Lemma 4.11 says that the two functions, $\lim_{n\to+\infty} n^{-1} I(\mathbf{S};\mathbf{W})$ and $\min_{E\in[0,v]} i_{\rm RS}(E;\Delta)$, are equal for $\Delta < \Delta_{\rm AMP}$ and Lemma 4.13 implies that *both* functions are analytic for $\Delta < \Delta_{\rm opt}$. Thus they must be equal on the whole range $\Delta < \Delta_{\rm opt}$. Since we also know they are continuous, then they are equal also at $\Delta = \Delta_{\rm opt}$. □

4.6.2 Proof of Theorem 4.1 in the High Noise Regime

We first need the following lemma where spatial coupling comes into the play.

Lemma 4.14. *The optimal threshold is given by the potential threshold:* $\Delta_{\rm opt} = \Delta_{\rm RS}$.

Proof. It suffices to see that

$$\Delta_{\rm RS} \le \Delta^{\rm c}_{\rm AMP} \le \Delta^{\rm c}_{\rm opt} = \Delta_{\rm opt} \le \Delta_{\rm RS}. \tag{4.106}$$

The first inequality is the threshold saturation result of Theorem 4.3 in Section 4.4. The second inequality is due the suboptimality of the AMP algorithm.[7] The equality is a consequence of Theorem 4.2 in Section 4.5. Indeed, equality of asymptotic mutual informations of the coupled and underlying system implies that they must be non-analytic at the same value of Δ. Finally, the last inequality is the statement of Lemma 4.13 in Section 4.6.1. □

Proposition 4.6. *Assume P_0 is a discrete distribution. Fix $\Delta \ge \Delta_{\rm opt}$. The mutual information per variable is asymptotically given by the RS formula* (4.15).

Proof. We already remarked in section 4.6.1 that for $\Delta > \Delta_{\rm RS}$,

$$\frac{d}{d\Delta}(\min_E i_{\rm RS}(E;\Delta)) \le \liminf_{n\to+\infty} \frac{1}{n}\frac{dI(\mathbf{S};\mathbf{W})}{d\Delta}. \tag{4.107}$$

Now we integrate on an interval $]\Delta_{\rm RS},\Delta]$ both sides of the inequality. Since from Lemma 4.14 we have that $\Delta_{\rm RS} = \Delta_{\rm opt}$, it is equivalent to integrate from $\Delta_{\rm opt}$ upwards[8]

$$\int_{\Delta_{\rm opt}}^{\Delta} d\tilde\Delta\, \frac{d}{d\tilde\Delta}(\min_E i_{\rm RS}(E;\tilde\Delta)) \le \int_{\Delta_{\rm opt}}^{\Delta} d\tilde\Delta\, \liminf_{n\to+\infty} \frac{1}{n}\frac{dI(\mathbf{S};\mathbf{W})}{d\tilde\Delta}. \tag{4.108}$$

By Fatou's lemma the inequality is preserved if we bring the lim inf outside of the integral, thus

$$\begin{aligned}\min_E i_{\rm RS}(E;\Delta) - \min_E i_{\rm RS}(E;\Delta_{\rm RS}) &\le \liminf_{n\to+\infty}\left\{\frac{1}{n}I(\mathbf{S};\mathbf{W})\Big|_{\Delta} - \frac{1}{n}I(\mathbf{S};\mathbf{W})\Big|_{\Delta_{\rm opt}}\right\}\\ &= \lim_{n\to+\infty}\frac{1}{n}I(\mathbf{S};\mathbf{W})\Big|_{\Delta} - \lim_{n\to+\infty}\frac{1}{n}I(\mathbf{S};\mathbf{W})\Big|_{\Delta_{\rm opt}}.\end{aligned} \tag{4.109}$$

[7] More precisely, one shows by the same methods Lemmas 4.11 and 4.12 for the spatially coupled system.

[8] This is the point we did not yet know in section 4.6.1.

To get the last line, we used the existence of the thermodynamic limit (see Lemma 4.10). We already know from Proposition 4.5 that $\min_E i_{\rm RS}(E;\Delta_{\rm opt}) = \lim_{n\to+\infty}\frac{1}{n}I(\mathbf{S};\mathbf{W})\Big|_{\Delta_{\rm opt}}$. Therefore

$$\min_E i_{\rm RS}(E;\Delta) \leq \lim_{n\to+\infty}\frac{1}{n}I(\mathbf{S};\mathbf{W}), \tag{4.110}$$

which together with (4.16) ends the proof. □

4.7 Proof of Main Corollaries

In this section, we provide the proofs of Corollary 4.1 and Corollary 4.2 concerning the MMSE formulae and the optimality of the AMP algorithm. We first show the following result about the matrix and vector MMSE's in Definition 4.4.

Lemma 4.15. *Assume the prior P_0 has finite first four moments and recall the second moment is called v. The matrix and vector MMSE verify*

$$\mathrm{Mmmse}_n \leq \big(v^2 - (v - \mathrm{Vmmse}_n)^2\big) + \mathcal{O}(\frac{1}{n}). \tag{4.111}$$

Proof. For this proof we denote $\langle\cdot\rangle$ the expectation w.r.t the posterior distribution (4.20). The matrix and vector MMSE then read

$$\mathrm{Mmmse}_n := \frac{1}{n^2}\mathbb{E}_{\mathbf{S},\mathbf{W}}\Big[\big\|\mathbf{S}\mathbf{S}^\intercal - \langle\mathbf{X}\mathbf{X}^\intercal\rangle\big\|_{\rm F}^2\Big], \tag{4.112}$$

$$\mathrm{Vmmse}_n := \frac{1}{n}\mathbb{E}_{\mathbf{S},\mathbf{W}}\Big[\big\|\mathbf{S} - \langle\mathbf{X}\rangle\big\|_2^2\Big]. \tag{4.113}$$

Expanding the Frobenius norm in (4.112) yields

$$\begin{aligned}
\mathrm{Mmmse}_n &= \frac{1}{n^2}\mathbb{E}_{\mathbf{S},\mathbf{W}}\Big[\sum_{i,j=1}^n (S_iS_j - \langle X_iX_j\rangle)^2\Big] \\
&= \frac{1}{n^2}\mathbb{E}_{\mathbf{S},\mathbf{W}}\Big[\sum_{i,j=1}^n S_i^2S_j^2 - \langle X_iX_j\rangle^2\Big] \\
&= \mathbb{E}_{\mathbf{S}}\Big[\Big(\frac{1}{n}\sum_{i=1}^n S_i^2\Big)^2\Big] - \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_iX_j\rangle^2],
\end{aligned} \tag{4.114}$$

where the second equality follows from the Nishimori identity (4.165) that yields $\mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_iX_j\rangle^2] = \mathbb{E}_{\mathbf{S},\mathbf{W}}[S_iS_j\langle X_iX_j\rangle]$. Similarly, using $\mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_i\rangle^2] = \mathbb{E}_{\mathbf{S},\mathbf{W}}[S_i\langle X_i\rangle]$ implied by the Nishimori identity, (4.113) simplifies to

$$\mathrm{Vmmse}_n = v - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_i\rangle^2]. \tag{4.115}$$

Hence,

$$\mathrm{Mmmse}_n - \left(v^2 - (v - \mathrm{Vmmse}_n)^2\right) = \mathcal{A}_n - \mathcal{B}_n, \tag{4.116}$$

with

$$\mathcal{A}_n := \mathbb{E}_{\mathbf{S}}\Big[\Big(\frac{1}{n}\sum_{i=1}^{n} S_i^2\Big)^2\Big] - v^2, \tag{4.117}$$

$$\mathcal{B}_n := \frac{1}{n^2}\sum_{i,j=1}^{n}\Big(\mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_i X_j\rangle^2] - \mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_i\rangle^2]\mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_j\rangle^2]\Big). \tag{4.118}$$

Since the signal components $\{S_i\}$ are i.i.d and P_0 has finite first four moments, $\mathcal{A}_n = \mathcal{O}(1/n)$. It remains to show that $\mathcal{B}_n \geq 0$. This is most easily seen as follows. By defining the *overlap*

$$q(\mathbf{X}, \mathbf{S}) := \frac{1}{n}\sum_{i=1}^{n} S_i X_i \tag{4.119}$$

and using the Nishimori identities

$$\begin{cases} \mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_i\rangle^2] = \mathbb{E}_{\mathbf{S},\mathbf{W}}[S_i\langle X_i\rangle] \\ \mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle X_i X_j\rangle^2] = \mathbb{E}_{\mathbf{S},\mathbf{W}}[S_i S_j\langle X_i X_j\rangle], \end{cases} \tag{4.120}$$

we observe that

$$\begin{aligned} \mathcal{B}_n &= \mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle q^2\rangle] - \mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle q\rangle]^2 \\ &= \mathbb{E}_{\mathbf{S},\mathbf{W}}[(q - \mathbb{E}_{\mathbf{S},\mathbf{W}}[\langle q\rangle])^2] \end{aligned} \tag{4.121}$$

which is non-negative. □

Remark 4.3. *Using ideas similar to [171] to prove concentration of overlaps in inference problems suggest that Lemma 4.15 holds with an equality when suitable "side observations" are added.*

4.7.1 Exact Formula for the MMSE (Corollary 4.1)

We first show how to prove the expression (4.22) for the asymptotic Mmmse_n by taking the limit $n \to +\infty$ on both sides of (4.88). First notice that since $n^{-1}I(\mathbf{S};\mathbf{W})$ is a sequence of concave functions with respect to Δ^{-1}, the limit when $n \to +\infty$ is also concave and differentiable for almost all Δ^{-1} and at all differentiability points we have (by a standard theorem of real analysis on convex functions)

$$\lim_{n\to+\infty}\frac{1}{n}\frac{d}{d\Delta^{-1}}I(\mathbf{S};\mathbf{W}) = \frac{1}{n}\frac{d}{d\Delta^{-1}}\lim_{n\to+\infty} I(\mathbf{S};\mathbf{W})\,. \tag{4.122}$$

Thus from Lemma 4.9 and Theorem 4.1 we have for all $\Delta \neq \Delta_{\mathrm{RS}}$

$$\lim_{n\to+\infty} \mathrm{Mmmse}_n(\Delta^{-1}) = 4\frac{d}{d\Delta^{-1}} \min_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta). \tag{4.123}$$

It remains to compute the right hand side. Let $E_0(\Delta)$ denote the (global) minimum of $i_{\mathrm{RS}}(E;\Delta)$. For $\Delta \neq \Delta_{\mathrm{RS}}$ this is a differentiable function of Δ with locally bounded derivative. Hence using a similar calculation to the one done in (4.94), we obtain

$$\begin{aligned}\frac{d}{d\Delta^{-1}}\Big(\min_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta)\Big) &= \frac{di_{\mathrm{RS}}}{d\Delta^{-1}}(E_0;\Delta)\\ &= \frac{\partial i_{\mathrm{RS}}}{\partial E}(E_0;\Delta)\frac{dE_0}{d\Delta^{-1}} + \frac{\partial i_{\mathrm{RS}}}{\partial \Delta^{-1}}(E_0;\Delta)\\ &= \frac{\partial i_{\mathrm{RS}}}{\partial \Delta^{-1}}(E_0;\Delta).\end{aligned} \tag{4.124}$$

To compute the partial derivative with respect to Δ we first note that $\frac{\partial i_{\mathrm{RS}}}{\partial E}\Big|_{E_0} = 0$ implies

$$0 = -\frac{v-E_0}{2\Delta} - \frac{\partial \mathbb{E}_{S,Z}[\cdots]}{\partial \Sigma^{-2}}\Big|_{E_0} \frac{\partial \Sigma^{-2}}{\partial E}\Big|_{E_0}, \tag{4.125}$$

where $\mathbb{E}_{S,Z}[\cdots]$ is the expectation that appears in the RS potential (4.14). This immediately gives

$$\frac{\partial \mathbb{E}_{S,Z}[\cdots]}{\partial \Sigma^{-2}}\Big|_{E_0} = \frac{1}{2}(v-E_0). \tag{4.126}$$

Thus

$$\begin{aligned}\frac{\partial i_{\mathrm{RS}}}{\partial \Delta^{-1}}(E_0;\Delta) &= \frac{(v-E_0)^2+v^2}{4} - \frac{\partial \mathbb{E}_{S,Z}[\cdots]}{\partial \Sigma^{-2}}\Big|_{E_0}\frac{\partial \Sigma^{-2}}{\partial \Delta^{-1}}\Big|_{E_0}\\ &= \frac{v^2-(v-E_0)^2}{4}.\end{aligned} \tag{4.127}$$

From (4.123), (4.124), (4.127) we obtain the desired result, formula (4.22).

We now turn to the proof of (4.23) for the expression of the asymptotic vector-MMSE. From Lemma 4.15 and the suboptimality of the AMP algorithm (here $E_0(\Delta)$ is the global minimum of $i_{\mathrm{RS}}(E;\Delta)$ and $E^{(\infty)}$ the fixed point of state evolution)

$$\begin{aligned}v^2-(v-E_0)^2 &= \lim_{n\to\infty} \mathrm{Mmmse}_n\\ &\leq \liminf_{n\to\infty}\big(v^2-(v-\mathrm{Vmmse}_n)^2\big)\\ &\leq \limsup_{n\to\infty}\big(v^2-(v-\mathrm{Vmmse}_n)^2\big)\\ &\leq v^2-(v-E^{(\infty)})^2,\end{aligned} \tag{4.128}$$

For $\Delta \notin [\Delta_{\mathrm{AMP}}, \Delta_{\mathrm{RS}}]$, we have that $E_0 = E^{(\infty)}$ which ends the proof.

4.7.2 Optimality of AMP (Corollary 4.2)

In view of (4.11) we have

$$\lim_{t\to+\infty}\lim_{n\to+\infty}\mathrm{Vmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) = E^{(\infty)}, \tag{4.129}$$

$$\lim_{t\to+\infty}\lim_{n\to+\infty}\mathrm{Mmse}^{(t)}_{n,\mathrm{AMP}}(\Delta^{-1}) = v^2 - (v-E^{(\infty)})^2. \tag{4.130}$$

For $\Delta \notin [\Delta_{\mathrm{AMP}}, \Delta_{\mathrm{RS}}]$, we have $E^{(\infty)} = \mathrm{argmin}_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta)$ and also the two identities of Corrolary 4.1 hold. This directly implies the identities (4.24) and (4.25).

For $\Delta \in [\Delta_{\mathrm{AMP}}, \Delta_{\mathrm{RS}}]$, we have $E^{(\infty)} > \mathrm{argmin}_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta)$, so using the monotonicity of $E^{(t)}$ leads to strict inequalities in (4.129) and (4.130) and thus to (4.26) and (4.27).

4.8 Appendix

4.8.1 Upper Bound on the Mutual Information

For the completeness of this Chapter we revisit the proof of the upper bound (4.16) on the mutual information. This result was already obtained by [47] using a Toninelli-Guerra type interpolation and is used in this Chapter, so we only sketch the main steps.

We consider the following interpolating inference problem

$$\begin{cases} w_{ij} = \frac{s_i s_j}{\sqrt{n}} + \sqrt{\frac{\Delta}{t}} z_{ij}, \\ y_i = s_i + \sqrt{\frac{\Delta}{m(1-t)}} z'_i, \end{cases} \tag{4.131}$$

with $m := v - E \in [0, v]$, and $Z'_i \sim \mathcal{N}(0,1)$. For $t = 1$ we find back the original problem (4.1) since the y_i observations become useless and for $t = 0$ we have a set of decoupled observations from a Gaussian channel. The interpolating posterior distribution associated to this set of observations is

$$P_t(\mathbf{x}|\mathbf{s},\mathbf{z},\mathbf{z}') := \frac{e^{-\mathcal{H}(t)}\prod_{i=1}^n P_0(x_i)}{\int\{\prod_{i=1}^n dx_i P_0(x_i)\} e^{-\mathcal{H}(t)}} := \frac{1}{\mathcal{Z}(t)} e^{-\mathcal{H}(t)} \prod_{i=1}^n P_0(x_i), \tag{4.132}$$

where

$$\begin{aligned}\mathcal{H}(t) = &\sum_{i\le j=1}^n \Big(\frac{t}{2\Delta n} x_i^2 x_j^2 - \frac{t}{\Delta n} x_i x_j s_i s_j - \sqrt{\frac{t}{n\Delta}} x_i x_j z_{ij}\Big) \\ &+ \sum_{i=1}^n \Big(\frac{m(1-t)}{2\Delta} x_i^2 - \frac{m(1-t)}{\Delta} x_i s_i + \sqrt{\frac{m(1-t)}{\Delta}} x_i z'_i\Big).\end{aligned}$$

can be interpreted as a "Hamiltonian" and the normalizing factor $\mathcal{Z}(t)$ is interpreted as a "partition function". We adopt the Gibbs "bracket" notation

$\langle - \rangle_t$ for the expectation with respect to the posterior (4.132). The mutual information associated to interpolating inference problem is

$$i(t) = -\frac{1}{n}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\ln \mathcal{Z}(t)] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n}(2\mathbb{E}[S^4] - v^2) \tag{4.133}$$

Note that on one hand $i(1) = \frac{1}{n}I(\mathbf{S};\mathbf{W})$ the mutual information of the original matrix factorization problem and on the other hand

$$\begin{aligned} i_0 =& -\frac{1}{n}\mathbb{E}_{\mathbf{S},\mathbf{Z}'}\Big[\ln\Big(\int \{\prod_{i=1}^{n} dx_i P_0(x_i)\} e^{-\frac{m\|\mathbf{x}\|_2^2}{2\Delta} + \mathbf{x}^\intercal\left(\frac{m\mathbf{S}}{\Delta} + \sqrt{\frac{m}{\Delta}}\mathbf{Z}'\right)}\Big)\Big] \\ &+ \frac{v^2}{4\Delta} + \frac{1}{4\Delta n}(2\mathbb{E}[S^4] - v^2) \\ =& -\mathbb{E}_{S,Z'}\Big[\ln\Big(\int dx\, P_0(x) e^{-\frac{mx^2}{2\Delta} + x\left(\frac{mS}{\Delta} + \sqrt{\frac{m}{\Delta}}Z'\right)}\Big)\Big] \\ &+ \frac{v^2}{4\Delta} + \frac{1}{4\Delta n}(2\mathbb{E}[S^4] - v^2) \\ =& i_{\mathrm{RS}}(E;\Delta) - \frac{m^2}{4\Delta} + \frac{1}{4\Delta n}(2\mathbb{E}[S^4] - v^2), \end{aligned} \tag{4.134}$$

From the fundamental theorem of calculus, we have

$$i(1) - i(0) = -\frac{1}{n}\int_0^1 dt \frac{d}{dt}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\ln \mathcal{Z}(t)], \tag{4.135}$$

so we get

$$\frac{1}{n}I(\mathbf{S};\mathbf{W}) = i_{\mathrm{RS}}(E;\Delta) - \frac{m^2}{4\Delta} + \frac{1}{4\Delta n}(2\mathbb{E}[S^4] - v^2) - \frac{1}{n}\int_0^1 dt \frac{d}{dt}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\ln \mathcal{Z}(t)]. \tag{4.136}$$

We proceed to the computation of the derivative under the integral over t. Denoting by $\langle - \rangle_t$ the expectation with respect to the posterior (4.132), we have

$$\frac{d}{dt}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\ln\big(\mathcal{Z}(t)\big)] = \mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\big[-\big\langle\frac{d\mathcal{H}(t)}{dt}\big\rangle_t\big]. \tag{4.137}$$

Hence, a simple differentiation of the Hamiltonian w.r.t. t yields

$$\begin{aligned} &\frac{d}{dt}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\ln\big(\mathcal{Z}(t)\big)] = \\ &\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\Big[\sum_{i\le j=1}^{n}\Big(-\frac{\langle X_i^2 X_j^2\rangle_t}{2\Delta n} + \frac{\langle X_i X_j\rangle_t S_i S_j}{\Delta n} + \frac{Z_{ij}\langle X_i X_j\rangle_t}{2\sqrt{n\Delta t}}\Big) \\ &+ \sum_{i=1}^{n}\Big(m\frac{\langle X_i^2\rangle_t}{2\Delta} - m\frac{\langle X_i\rangle_t S_i}{\Delta} - \frac{Z_i'\langle X_i\rangle_t}{2}\sqrt{\frac{m}{\Delta(1-t)}}\Big)\Big]. \end{aligned} \tag{4.138}$$

We now simplify this expression using integration by parts with respect to the Gaussian noises and the Nishimori identity (4.165) in Appendix 4.8.4. Integration by parts with respect to Z_{ij} and Z'_i yields

$$\begin{aligned}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[Z_{ij}\langle X_iX_j\rangle_t] &= \mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\partial_{Z_{ij}}\langle X_iX_j\rangle_t] \\ &= \sqrt{\frac{t}{n\Delta}}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\Big[\langle X_i^2X_j^2\rangle_t - \langle X_iX_j\rangle_t^2\Big] \end{aligned} \tag{4.139}$$

and

$$\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[Z'_i\langle X_i\rangle_t] = \mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\partial_{Z'_i}\langle X_i\rangle_t] = \sqrt{\frac{m(1-t)}{\Delta}}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\Big[\langle X_i^2\rangle_t - \langle X_i\rangle_t^2\Big]. \tag{4.140}$$

An application of the Nishimori identity yields

$$\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_iX_j\rangle_t S_iS_j] = \mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_iX_j\rangle_t^2] \tag{4.141}$$

and

$$\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_i\rangle_t S_i] = \mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_i\rangle_t^2] \tag{4.142}$$

Combining (4.138) - (4.142) we get

$$\begin{aligned}\frac{1}{n}\frac{d}{dt}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\ln\big(\mathcal{Z}(t)\big)] =& \frac{1}{2\Delta n^2}\sum_{i\le j=1}^{n}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_iX_jS_iS_j\rangle_t] \\ &- \frac{m}{2\Delta n}\sum_{i=1}^{n}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_iS_i\rangle_t] \\ =& \frac{1}{4\Delta}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\big[\langle q(\mathbf{S},\mathbf{X})^2\rangle_t - 2\langle q(\mathbf{S},\mathbf{X})\rangle_t m\big] \\ &+ \frac{1}{4\Delta n^2}\sum_{i=1}^{n}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_i^2\rangle_t S_i^2],\end{aligned}$$

where we have introduced the "overlap" $q(\mathbf{S},\mathbf{X}) := n^{-1}\sum_{i=1}^n S_iX_i$. Replacing this result in (4.136) we obtain the remarkable sum rule (recall $m := v - E$)

$$\begin{aligned}\frac{1}{n}I(\mathbf{S};\mathbf{W}) = i_{\rm RS}(E;\Delta) &- \frac{1}{4\Delta}\int_0^1 dt\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}\big[\langle(q(\mathbf{S},\mathbf{X})-m)^2\rangle_t\big] \\ &- \frac{1}{4\Delta n^2}\sum_{i=1}^{n}\mathbb{E}_{\mathbf{S},\mathbf{Z},\mathbf{Z}'}[\langle X_i^2\rangle_t S_i^2] + \frac{1}{4\Delta n}(2\mathbb{E}[S^4] - v^2).\end{aligned} \tag{4.143}$$

Thus for any $E \in [0, v]$ we have

$$\limsup_{n\to+\infty}\frac{1}{n}I(\mathbf{S};\mathbf{W}) \le i_{\rm RS}(E;\Delta) \tag{4.144}$$

and (4.16) follows by optimizing the right hand side over E.

4.8.2 Relating the Mutual Information to the Free Energy

The mutual information between $\mathbf{S}$ and $\mathbf{W}$ is defined as $I(\mathbf{S};\mathbf{W}) = H(\mathbf{S}) - H(\mathbf{S}|\mathbf{W})$ with

$$H(\mathbf{S}|\mathbf{W}) = -\mathbb{E}_{\mathbf{S},\mathbf{W}}\big[\ln P(\mathbf{S}|\mathbf{W})\big] = -\mathbb{E}_{\mathbf{S},\mathbf{Z}}\big[\langle \ln P(\mathbf{X}|\mathbf{S},\mathbf{Z})\rangle\big]. \tag{4.145}$$

By substituting the posterior distribution in (4.33), one obtains

$$H(\mathbf{S}|\mathbf{W}) = \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\ln \mathcal{Z}] + \mathbb{E}_{\mathbf{S},\mathbf{Z}}\big[\langle \mathcal{H}(\mathbf{X}|\mathbf{S},\mathbf{Z})\rangle\big] + H(\mathbf{S}). \tag{4.146}$$

Furthermore, using the Gaussian integration by part as in (4.139) and the Nishimori identity (4.141) yield

$$\mathbb{E}_{\mathbf{S},\mathbf{Z}}\big[\langle \mathcal{H}(\mathbf{X}|\mathbf{S},\mathbf{Z})\rangle\big] = -\frac{1}{2\Delta n}\mathbb{E}_{\mathbf{S}}\big[\sum_{i\leq j=1}^{n}(S_i^2 S_j^2)\big] = -\frac{1}{4\Delta}\big(v^2(n-1) + 2\mathbb{E}[S^4]\big). \tag{4.147}$$

Hence, the normalized mutual information is given by (4.38), which we repeat here for better referencing

$$\frac{1}{n}I(\mathbf{S};\mathbf{W}) = -\frac{1}{n}\mathbb{E}_{\mathbf{S},\mathbf{Z}}[\ln \mathcal{Z}] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n}(2\mathbb{E}[S^4] - v^2). \tag{4.148}$$

Alternatively, one can define the mutual information as $I(\mathbf{S};\mathbf{W}) = H(\mathbf{W}) - H(\mathbf{W}|\mathbf{S})$. For the AWGN, it is easy to show that

$$H(\mathbf{W}|\mathbf{S}) = \frac{n(n+1)}{4}\ln(2\pi\Delta e). \tag{4.149}$$

Furthermore, $H(\mathbf{W}) = -\mathbb{E}_{\mathbf{W}}\big[\ln P(\mathbf{W})\big]$ with

$$\begin{aligned} P(\mathbf{w}) &= \int \big\{\prod_{i=1}^{n} dx_i P_0(x_i)\big\} P(\mathbf{w}|\mathbf{x}) \\ &= \frac{1}{(2\pi\Delta)^{\frac{n(n+1)}{4}}}\int \big\{\prod_{i=1}^{n} dx_i P_0(x_i)\big\} e^{-\frac{1}{2\Delta}\sum_{i\leq j}\left(\frac{x_i x_j}{\sqrt{n}} - w_{ij}\right)^2} \\ &= \frac{\tilde{\mathcal{Z}}}{(2\pi\Delta)^{\frac{n(n+1)}{4}}}, \end{aligned} \tag{4.150}$$

where $\tilde{\mathcal{Z}}$ is the partition function with complete square (4.20). Hence, $H(\mathbf{W})$ reads

$$\begin{aligned} H(\mathbf{W}) &= \frac{n(n+1)}{4}\ln(2\pi\Delta) - \mathbb{E}_{\mathbf{W}}\big[\ln \tilde{\mathcal{Z}}\big] \\ &= \frac{n(n+1)}{4}\ln(2\pi\Delta) - \mathbb{E}_{\mathbf{S},\mathbf{Z}}\big[\ln \mathcal{Z}\big] + \frac{1}{2\Delta}\mathbb{E}_{\mathbf{S},\mathbf{Z}}\big[\sum_{i\leq j}(\frac{S_i S_j}{\sqrt{n}} + \sqrt{\Delta}Z_{ij})^2\big], \end{aligned} \tag{4.151}$$

with $\mathcal{Z}$ the simplified partition function obtained after expanding the square (4.33). A straightforward calculation yields

$$\frac{1}{2\Delta}\mathbb{E}_{\mathbf{S},\mathbf{Z}}\big[\sum_{i\leq j}(\frac{S_iS_j}{\sqrt{n}}+\sqrt{\Delta}Z_{ij})^2\big] = \frac{n(n+1)}{4}+\frac{1}{2\Delta n}\mathbb{E}_{\mathbf{S}}\big[\sum_{i\leq j}(S_i^2S_j^2)\big] \\ = \frac{n(n+1)}{4}\ln(e)+\frac{1}{4\Delta}\big(v^2(n-1)+2\mathbb{E}[S^4]\big). \tag{4.152}$$

Finally, combining (4.149), (4.151) and (4.152) yields the same identity (4.148).

4.8.3 Proof of the I-MMSE Relation

For completeness, we give a detailed proof for the I-MMSE relation of Lemma 4.9 following the lines of [48]. In the calculations below differentiations, expectations and integrations commute (see Lemma 8 in [48]). All the matrices are symmetric and $Z_{ij}\sim\mathcal{N}(0,1)$ for $i\leq j$.

Instead of (4.1) it is convenient to work with the equivalent model $w_{ij}=\frac{s_is_j}{\sqrt{n\Delta}}+z_{ij}$ and set $s_is_j=u_{ij}$. In fact, all subsequent calculations do not depend on the rank of the matrix $\mathbf{u}$ and are valid for any finite rank matrix estimation problem as long as the noise is Gaussian. The mutual information is $I(\mathbf{S};\mathbf{W})=H(\mathbf{W})-H(\mathbf{W}|\mathbf{S})$ and $H(\mathbf{W}|\mathbf{S})=\frac{n(n+1)}{2}\ln(\sqrt{2\pi e})$. Thus

$$\frac{1}{n}\frac{dI(\mathbf{S};\mathbf{W})}{d\Delta^{-1}}=\frac{1}{n}\frac{dH(\mathbf{W})}{d\Delta^{-1}}. \tag{4.153}$$

We have $H(\mathbf{W})=-\mathbb{E}_{\mathbf{W}}[\ln P(\mathbf{W})]$ where

$$P(\mathbf{w})=\mathbb{E}_{\mathbf{U}}[P(\mathbf{w}|\mathbf{U})]=\mathbb{E}_{\mathbf{U}}\Big[(2\pi)^{-\frac{n(n+1)}{4}}e^{-\frac{1}{2}\sum_{i\leq j}\left(\frac{U_{ij}}{\sqrt{n\Delta}}-w_{ij}\right)^2}\Big]. \tag{4.154}$$

Differentiating w.r.t Δ^{-1}

$$\frac{dH(\mathbf{W})}{d\Delta^{-1}}=-\mathbb{E}_{\mathbf{U}}\Big[\int d\mathbf{w}(1+\ln P(\mathbf{w}))\frac{dP(\mathbf{w}|\mathbf{U})}{d\Delta^{-1}}\Big] \tag{4.155}$$

and

$$\frac{dP(\mathbf{w}|\mathbf{U})}{d\Delta^{-1}}=\sqrt{\frac{\Delta}{4n}}\sum_{k\leq l}U_{kl}\Big(w_{kl}-\frac{U_{kl}}{\sqrt{\Delta n}}\Big)e^{-\frac{1}{2}\sum_{i\leq j}\left(\frac{U_{ij}}{\sqrt{n\Delta}}-w_{ij}\right)^2}(2\pi)^{-\frac{n(n+1)}{2}} \tag{4.156}$$

$$=-\sqrt{\frac{\Delta}{4n}}\sum_{k\leq l}U_{kl}\frac{d}{dw_{kl}}e^{-\frac{1}{2}\sum_{i\leq j}\left(\frac{U_{ij}}{\sqrt{n\Delta}}-w_{ij}\right)^2}(2\pi)^{-\frac{n(n+1)}{2}}. \tag{4.157}$$

Replacing this last expression in (4.155), using an integration by part w.r.t w_{kl} (the boundary terms can be shown to vanish), then Bayes formula, and finally (4.154), one obtains

$$\begin{aligned}
&\sqrt{\frac{4n}{\Delta}}\frac{dH(\mathbf{W})}{d\Delta^{-1}} = \\
&\sum_{k\leq l}\mathbb{E}_{\mathbf{U}}\Big[U_{kl}\int d\mathbf{w}(1+\ln P(\mathbf{w}))\frac{d}{dw_{kl}}e^{-\frac{1}{2}\sum_{i\leq j}\left(\frac{U_{ij}}{\sqrt{n\Delta}}-w_{ij}\right)^2}(2\pi)^{-\frac{n(n+1)}{2}}\Big] \\
&= -\sum_{k\leq l}\int d\mathbf{w}\,\mathbb{E}_{\mathbf{U}}\Big[U_{kl}\frac{P(\mathbf{w}|\mathbf{U})}{P(\mathbf{w})}\Big]\frac{dP(\mathbf{w})}{dw_{kl}} \\
&= -\sum_{k\leq l}\int d\mathbf{w}\,\mathbb{E}_{\mathbf{U}|\mathbf{w}}\big[U_{kl}\big]\mathbb{E}_{\mathbf{U}}\Big[\frac{dP(\mathbf{w}|\mathbf{U})}{dw_{kl}}\Big] \\
&= \sum_{k\leq l}\int d\mathbf{w}\,\mathbb{E}_{\mathbf{U}|\mathbf{w}}\big[U_{kl}\big]\mathbb{E}_{\mathbf{U}}\Big[\Big(w_{kl}-\frac{U_{kl}}{\sqrt{n\Delta}}\Big)P(\mathbf{w}|\mathbf{U})\Big] \\
&= \mathbb{E}_{\mathbf{W}}\Big[\sum_{k\leq l}\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]\Big(W_{kl}-\frac{1}{\sqrt{n\Delta}}\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]\Big)\Big]. \qquad (4.158)
\end{aligned}$$

Now we replace $\mathbf{w} = \frac{\mathbf{u}^0}{\sqrt{n\Delta}} + \mathbf{z}$, where $\mathbf{u}^0$ is an independent copy of $\mathbf{u}$. We denote $\mathbb{E}_{\mathbf{W}}[\cdot] = \mathbb{E}_{\mathbf{U}^0,\mathbf{Z}}[\cdot]$ the joint expectation. The last result then reads

$$\frac{dH(\mathbf{W})}{d\Delta^{-1}} = \frac{1}{2n}\mathbb{E}_{\mathbf{W}}\Big[\sum_{k\leq l}\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]\Big(U^0_{kl}-\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]+Z_{kl}\sqrt{n\Delta}\Big)\Big]. \qquad (4.159)$$

Now note the two Nishimori identities (see Appendix 4.8.4)

$$\mathbb{E}_{\mathbf{W}}\Big[\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]U^0_{kl}\Big] = \mathbb{E}_{\mathbf{W}}\Big[\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]^2\Big], \qquad (4.160)$$

$$\mathbb{E}_{\mathbf{W}}\Big[(U^0_{kl})^2\Big] = \mathbb{E}_{\mathbf{W}}\Big[\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U^2_{kl}\big]\Big], \qquad (4.161)$$

and the following one obtained by a Gaussian integration by parts

$$\sqrt{n\Delta}\,\mathbb{E}_{\mathbf{W}}\Big[\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]Z_{kl}\Big] = \mathbb{E}_{\mathbf{W}}\Big[\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U^2_{kl}\big]-\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]^2\Big]. \qquad (4.162)$$

Using the last three identities, equation (4.159) becomes

$$\begin{aligned}
\frac{1}{n}\frac{dH(\mathbf{W})}{d\Delta^{-1}} &= \frac{1}{2n^2}\mathbb{E}_{\mathbf{W}}\Big[\sum_{k\leq l}\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U^2_{kl}\big]-\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]^2\Big] \\
&= \frac{1}{2n^2}\mathbb{E}_{\mathbf{W}}\Big[\sum_{k\leq l}\Big(U^0_{kl}-\mathbb{E}_{\mathbf{U}|\mathbf{W}}\big[U_{kl}\big]\Big)^2\Big], \qquad (4.163)
\end{aligned}$$

which, in view of (4.153), ends the proof.

4.8.4 Nishimori Identity

Take a random vector $\mathbf{S}$ distributed according to some known prior $P_0^{\otimes n}$ and an observation $\mathbf{W}$ is drawn from some known conditional distribution $P_{\mathbf{W}|\mathbf{S}}(\mathbf{w}|\mathbf{s})$. Take $\mathbf{X}$ drawn from a posterior distribution (for example this may be (4.20))

$$P(\mathbf{x}|\mathbf{w}) = \frac{P_0^{\otimes n}(\mathbf{x})P_{\mathbf{W}|\mathbf{S}}(\mathbf{w}|\mathbf{x})}{P(\mathbf{w})}.$$

Then for any (integrable) function $g(\mathbf{s}, \mathbf{x})$ the Bayes formula implies

$$\mathbb{E}_{\mathbf{S}}\mathbb{E}_{\mathbf{W}|\mathbf{S}}\mathbb{E}_{\mathbf{X}|\mathbf{W}}[g(\mathbf{S}, \mathbf{X})] = \mathbb{E}_{\mathbf{W}}\mathbb{E}_{\mathbf{X}'|\mathbf{W}}\mathbb{E}_{\mathbf{X}|\mathbf{W}}[g(\mathbf{X}', \mathbf{X})] \tag{4.164}$$

where $\mathbf{X}, \mathbf{X}'$ are independent random vectors distributed according to the posterior distribution. Therefore

$$\mathbb{E}_{\mathbf{S},\mathbf{W}}\mathbb{E}_{\mathbf{X}|\mathbf{W}}[g(\mathbf{S}, \mathbf{X})] = \mathbb{E}_{\mathbf{W}}\mathbb{E}_{\mathbf{X}'|\mathbf{W}}\mathbb{E}_{\mathbf{X}|\mathbf{W}}[g(\mathbf{X}', \mathbf{X})]. \tag{4.165}$$

In the statistical mechanics literature this identity is sometimes called the Nishimori identity and we adopt this language here. For model (4.1) for example we can express $\mathbf{W}$ in the posterior in terms of $\mathbf{S}$ and $\mathbf{Z}$ which are independent and $\mathbb{E}_{\mathbf{X}|\mathbf{W}}[-] = \langle - \rangle$. Then the Nishimori identity reads

$$\mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle g(\mathbf{S}, \mathbf{X})\rangle] = \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle g(\mathbf{X}', \mathbf{X})\rangle]. \tag{4.166}$$

An important case for g depending only on the first argument is $\mathbb{E}_{\mathbf{S}}[g(\mathbf{S})] = \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle g(\mathbf{X})\rangle]$.

Special cases that are often used in this Chapter are

$$\begin{cases} \mathbb{E}_{\mathbf{S},\mathbf{Z}}[S_i\langle X_i\rangle] = \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle X_i\rangle^2] \\ \mathbb{E}_{\mathbf{S},\mathbf{Z}}[S_iS_j\langle X_iX_j\rangle] = \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle X_iX_j\rangle^2] \\ \mathbb{E}[S^2] = \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle X_i^2\rangle]. \end{cases} \tag{4.167}$$

A mild generalization of (4.166) which is also used is

$$\mathbb{E}_{\mathbf{S},\mathbf{Z}}[S_iS_j\langle X_i\rangle\langle X_j\rangle] = \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle X_iX_j\rangle\langle X_i\rangle\langle X_j\rangle]. \tag{4.168}$$

We remark that these identities are used with brackets $\langle - \rangle$ corresponding to various "interpolating" posteriors.

4.8.5 Relation Between State Evolution and Potential Function

We show the details for Lemma 4.1. The proof of Lemma 4.3 follows the same lines. A straightforward differentiation of $f_{\mathrm{RS}}^{\mathrm{u}}$ w.r.t. E gives

$$\frac{df_{\mathrm{RS}}^{\mathrm{u}}(E;\Delta)}{dE} = \frac{E-v}{2\Delta} + \frac{1}{2\Delta}\mathbb{E}_{Z,S}\Big[\Big\langle -X^2 + 2XS + ZX\sqrt{\frac{\Delta}{v-E}}\Big\rangle\Big]. \tag{4.169}$$

Recall that here the posterior expectation $\langle\cdot\rangle$ is defined by (4.40). A direct application of the Nishimori condition gives

$$v := \mathbb{E}_S[S^2] = \mathbb{E}_{S,Z}[\langle X^2\rangle], \tag{4.170}$$
$$\mathbb{E}_{S,Z}[S\langle X\rangle] = \mathbb{E}_{S,Z}[\langle X\rangle^2], \tag{4.171}$$

which implies

$$\mathbb{E}_{S,Z}[(S - \langle X\rangle_E)^2] = \mathbb{E}_{S,Z}[\langle X^2\rangle] - \mathbb{E}_{S,Z}[\langle X\rangle^2]. \tag{4.172}$$

Thus from (4.169) we see that stationary points of $f^{\mathrm{u}}_{\mathrm{RS}}$ satisfy

$$E = 2v - 2\mathbb{E}_{S,Z}[\langle X\rangle^2] - \sqrt{\frac{\Delta}{v-E}}\mathbb{E}_{S,Z}[Z\langle X\rangle]. \tag{4.173}$$

Now using an integration by part w.r.t Z, one gets

$$\sqrt{\frac{\Delta}{v-E}}\mathbb{E}_{S,Z}[Z\langle X\rangle] = v - \mathbb{E}_{S,Z}[\langle X\rangle^2], \tag{4.174}$$

which allows to rewrite (4.173) as

$$E = v - \mathbb{E}_{S,Z}[\langle X\rangle^2] = \mathbb{E}_{S,Z}[(S - \langle X\rangle)^2] \tag{4.175}$$

where the second equality follows from (4.170) and (4.172). Recalling the expression (4.41) of the state evolution operator we recognize the equation $E = T_{\mathrm{u}}(E)$.

4.8.6 Analysis of the Potential for Small Noise

In this appendix, we prove that $\lim_{\Delta\to 0}\min_E i_{\mathrm{RS}}(E;\Delta) = H(S)$. First, a simple calculation leads to the following relation between i_{RS} and the mutual information of the scalar denoising problem for $E\in[0,v]$

$$i_{\mathrm{RS}}(E;\Delta) = I\big(S; S + \Sigma(E)Z\big) + \frac{E^2}{4\Delta}, \tag{4.176}$$

where $Z\sim\mathcal{N}(0,1)$ and $\Sigma(E)^2 := \Delta/(v-E)$. Note that as $\Delta\to 0$, $\Sigma(E)\to 0$ (for $E\neq v$). Therefore, $\lim_{\Delta\to 0} I\big(S; S+\Sigma(E)Z\big) = H(S)$. Now let E_0 be the global minimum of $i_{\mathrm{RS}}(E;\Delta)$. By evaluating both sides of (4.176) at E_0 and taking the limit $\Delta\to 0$, it remains to show that $E_0^2/(4\Delta)\to 0$ as $\Delta\to 0$ (i.e. $E_0^2\to 0$ faster than Δ). Since E_0 is the global minimum of the RS potential, then $E_0 = T_{\mathrm{u}}(E_0) = \mathrm{mmse}(\Sigma(E_0)^{-2})$ by Lemma 4.1. Moreover, one can show, under our assumptions on P_0, that the scalar MMSE function scales as

$$\mathrm{mmse}(\Sigma^{-2}) = \mathcal{O}(e^{-c\Sigma^{-2}}), \tag{4.177}$$

with c a non-negative constant that depends on P_0 [104]. Hence, $E_0^2/(4\Delta)\to 0$ as $\Delta\to 0$, which ends the proof.

4.8.7 Opening the Chain of the Spatially Coupled System

In this appendix, we provide a proof for Proposition 4.4. Call $\mathcal{H}^{\mathrm{per}}$ and $\langle - \rangle_{\mathrm{per}}$ the Hamiltonian and posterior average associated to the periodic SC system with mutual information $i^{\mathrm{per}}_{n,w,L}$. Similarly call $\mathcal{H}^{\mathrm{c\tilde{o}u}}$ and $\langle - \rangle_{\mathrm{cou}}$ the Hamiltonian and posterior average associated to the pinned SC system with mutual information $i^{\mathrm{cou}}_{n,w,L}$. The Hamiltonians satisfy the identity $\mathcal{H}^{\mathrm{cou}} - \mathcal{H}^{\mathrm{per}} = \delta\mathcal{H}$ with

$$\begin{aligned}
\delta\mathcal{H} =& \sum_{\mu\in\mathcal{B}} \Lambda_{\mu,\mu} \sum_{i_\mu \le j_\mu} \left[\frac{x^2_{i_\mu} x^2_{j_\mu} + s^2_{i_\mu} s^2_{j_\mu}}{2n\Delta} - \frac{s_{i_\mu} s_{j_\mu} x_{i_\mu} x_{j_\mu}}{n\Delta} - \frac{(x_{i_\mu} x_{j_\mu} - s_{i_\mu} s_{j_\mu}) z_{i_\mu j_\mu}}{\sqrt{n\Delta\Lambda_{\mu,\mu}}} \right] \\
&+ \sum_{\mu\in\mathcal{B}} \sum_{\nu\in\{\mu+1:\mu+w\}\cap\mathcal{B}} \Lambda_{\mu,\nu} \sum_{i_\mu \le j_\nu} \left[\frac{x^2_{i_\mu} x^2_{j_\nu} + s^2_{i_\mu} s^2_{j_\nu}}{2n\Delta} - \frac{s_{i_\mu} s_{j_\nu} x_{i_\mu} x_{j_\nu}}{n\Delta} \right. \\
&\qquad \left. - \frac{(x_{i_\mu} x_{j_\nu} - s_{i_\mu} s_{j_\nu}) z_{i_\mu j_\nu}}{\sqrt{n\Delta\Lambda_{\mu,\nu}}} \right] \\
&+ \sum_{\mu\in\mathcal{B}} \sum_{\nu\in\{\mu-w:\mu-1\}\cap\mathcal{B}} \Lambda_{\mu,\nu} \sum_{i_\mu > j_\nu} \left[\frac{x^2_{i_\mu} x^2_{j_\nu} + s^2_{i_\mu} s^2_{j_\nu}}{2n\Delta} - \frac{s_{i_\mu} s_{j_\nu} x_{i_\mu} x_{j_\nu}}{n\Delta} \right. \\
&\qquad \left. - \frac{(x_{i_\mu} x_{j_\nu} - s_{i_\mu} s_{j_\mu}) z_{i_\mu j_\nu}}{\sqrt{n\Delta\Lambda_{\mu,\nu}}} \right].
\end{aligned}$$

It is easy to see that

$$i^{\mathrm{per}}_{n,w,L} - i^{\mathrm{cou}}_{n,w,L} = \frac{1}{n(L+1)} \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\ln\langle e^{-\delta\mathcal{H}} \rangle_{\mathrm{cou}}], \tag{4.178}$$

$$i^{\mathrm{cou}}_{n,w,L} - i^{\mathrm{per}}_{n,w,L} = \frac{1}{n(L+1)} \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\ln\langle e^{\delta\mathcal{H}} \rangle_{\mathrm{per}}]. \tag{4.179}$$

Moreover, using the convexity of the exponential, we get

$$i^{\mathrm{cou}}_{n,w,L} + \frac{\mathbb{E}_{\mathbf{S},\mathbf{Z}}\big[\langle \delta\mathcal{H} \rangle_{\mathrm{per}}\big]}{n(L+1)} \le i^{\mathrm{per}}_{n,w,L} \le i^{\mathrm{cou}}_{n,w,L} + \frac{\mathbb{E}_{\mathbf{S},\mathbf{Z}}\big[\langle \delta\mathcal{H} \rangle_{\mathrm{cou}}\big]}{n(L+1)}. \tag{4.180}$$

Due to the pinning condition we have $\mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle \delta\mathcal{H}(\mathbf{X}) \rangle_{\mathrm{cou}}] = 0$, and thus we get the upper bound $i^{\mathrm{per}}_{n,w,L} \le i^{\mathrm{cou}}_{n,w,L}$. Let us now look at the lower bound. We note that by the Nishimori identity in Appendix 4.8.4, and as long as P_0 has finite first four moments, we can find constants K_1, K_2 independent of n, w, L such that

$$\begin{aligned}
\mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle X^2_{i_\mu} X^2_{j_\nu} \rangle_{\mathrm{per}}] &\le K_1, \\
\mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle X^4_{i_\mu} \rangle_{\mathrm{per}}] &\le K_2.
\end{aligned} \tag{4.181}$$

First we use Gaussian integration by parts to eliminate $z_{i_\mu j_\nu}$ from the brackets, the Cauchy-Schwartz inequality, and the Nishimori identity of Appendix 4.8.4,

to get an upper bound where only fourth order moments of signal are involved. Thus as long as P_0 has finite first four moments we find

$$\frac{|\mathbb{E}_{\mathbf{S},\mathbf{Z}}[\langle \delta\mathcal{H}(\mathbf{X})\rangle_{\tilde{c}}]|}{n(L+1)} \leq C\frac{\Lambda^*(2w+1)^2}{L+1} = \mathcal{O}(\frac{w}{L}). \tag{4.182}$$

for some constant $C > 0$ independent of n, w, L and we recall $\Lambda_* := \sup \Lambda_{\mu\nu} = \mathcal{O}(w^{-1})$. Thus we get the lower bound $i^{\mathrm{cou}}_{n,w,L} - \mathcal{O}(\frac{w}{L}) \leq i^{\mathrm{per}}_{n,w,L}$. This completes the proof of Proposition 4.4.

Conclusion and Further Directions

5

In this thesis, we have addressed two inference problems in the fields of coding theory and machine learning. In the first problem, we have considered the generalization of a recent forward-error-correction code, namely the sparse superposition code, to a large class of noisy channels. Moreover, we have employed the sparse superposition code to perform distribution matching, an inverse source coding scheme. In the second problem, we have studied the symmetric rank-one matrix factorization, a prominent model in machine learning and statistics with many applications ranging from community detection to sparse principal component analysis. We have provided a Bayesian formulation for the problem and analyzed it using an information theoretic approach. By computing the mutual information, we have established fundamental theoretical guarantees for the problem and proven the optimality of a low-complexity message-passing algorithm.

The connection between the two problems stems from the fact that both of them can be represented on dense graphical models. This allows to devise similar algorithms, such as AMP and GAMP, and to harness recent efficient graphical constructions, such as spatial coupling. Moreover, the structure of both problems is reminiscent of spin glass models studied in statistical physics. This can help in employing some sophisticated techniques developed in statistical physics, such as the potential function predicted by the replica method, in order to perform the analysis on a rigorous mathematical basis.

In Chapter 2, we have shown that spatially coupled sparse superposition codes universally achieve capacity over any memoryless channel under GAMP decoding. In particular, we have proven that spatial coupling allows the algorithmic GAMP performance to saturate the potential threshold of the underlying code ensemble. Moreover, we have shown by analytical calculation that the potential threshold tends to capacity and the error floor vanishes in the

proper limit. The approach taken in this chapter relies on the state evolution analysis and the application of the potential method.

In Chapter 3, we have presented a novel formulation of the fixed-length distribution matching inspired from the sparse superposition codes and the compressed sensing paradigm. The proposed solution uses a low-complexity dematching based on the GAMP algorithm. We have shown that GAMP dematching along with spatial coupling yields asymptotically optimal performance. Moreover, we have investigated practical scenarios using Hadamard-based operators. A notable aspect of the proposed solution is the amenability to perform joint channel coding and matching.

In Chapter 4, we have provided an explicit single-letter expression of the asymptotic mutual information for the symmetric rank-one matrix factorization. This was made possible by proving that the heuristic predictions of the replica method are exact, a long-standing conjecture in statistical physics mean-field theory. Furthermore, we have characterized the Bayes-optimal detectability region and estimation error. Moreover, we have proven that the AMP algorithm yields the optimal performance for a large set of parameters. Spatial coupling was employed in this chapter as an auxiliary model used in the proof, and as a prototype for potential applications. In our proof technique, we have exploited three essential ingredients: the interpolation method introduced in statistical physics, the analysis of the AMP algorithm through the state evolution introduced in compressed sensing, and the theory of threshold saturation for spatially coupled systems developed in coding theory.

We end up pointing out some open problems. The AMP and GAMP algorithms [60, 110] were first introduced for the noisy compressed sensing problem. The success story of these variants of message-passing algorithms on dense graphical models stems from the fact that their performance is asymptotically tracked by the state evolution recursion, an important feature that allows for rigorous mathematical analysis. State evolution is the analogous tool of density evolution used for sparse graphical models. The justification of state evolution is based on the conditioning techniques of Bolthausen [70], which was used by Bayati and Montanari [69] to prove the exactness of state evolution for compressed sensing. Soon after that, the proof of state evolution was extended to account for general channels and spatially coupled models in [115].

The AMP and GAMP algorithms were then adapted to account for the structured sparsity in the sparse superposition codes [36, 78]. Moreover, the AMP algorithm was adapted to the symmetric rank-one matrix factorization [164]. Furthermore, the state evolution analysis was extended to both problems and proven on rigorous basis in [37], [42] and [166]. These results are valid on the underlying (uncoupled) models. An important future direction is to extend these findings on state evolution to the spatially coupled models of sparse superposition codes and matrix factorization. We believe that this is possible by extending the work of [37], [42] and[166] and following the same lines of [115].

Table 5.1: Rigorous state evolution for AMP and GAMP algorithms.

	Uncoupled	**Coupled**
Compressed Sensing (AMP)	[69]	[76]
Compressed Sensing (GAMP)	[115]	[115]
SS Codes (AMP)	[37]	–
SS Codes (GAMP)	–	–
Matrix Factorization (AMP)	[42, 166]	–

Table 5.1 summarizes the instances where state evolution was rigorously proven to track the AMP and GAMP algorithms for our problems of interest. The missing entries represent cases where state evolution was numerically verified, yet no rigorous proof exists.

Another important direction for sparse superposition codes is to analyze the finite-length effects in terms of error exponent, scaling exponent and error floor regime,[1] a direction which was recently pursued in [180], [181] and[182]. Such analysis can provide many insights on how to design practical codes by choosing the appropriate code parameters.

The use of structured operators, such as Hadamard-based matrices, for SS codes has shown an improvement in terms of the finite-length performance [77]. Moreover, the use of such operators can reduce the computational complexity and the memory need of the AMP algorithm. Although structured operators are mathematically harder to analyze compared to the Gaussian ones, it is strongly desirable to consider the rigorous analysis of such operators (e.g. Hadamard-based matrices or, more generally, row-orthogonal matrices) and to quantify the finite-length improvement incurred by their use. Of course, the employment of such operators would necessitate the adoption of other variants of AMP and state evolution that are better suited to general matrices [148, 149, 150].

In addition to that, a promising direction is to optimize the decoding schedule of AMP for spatially coupled SS codes. This could be done by extending the *windowed decoding* of [183] to the dense graphical models. Hence, one can improve the decoding complexity and the velocity of the propagation wave [83, 84].

Finally, we would like to point out that the proof strategy we have used

[1]Note that there are two notions of error floor for SS codes. The finite-length error floor (in terms of L), and the finite-alphabet error floor (in terms of B) that appears in the AWGN channel.

in Chapter 4 to assert the validity of the heuristic replica predictions is quite general. Hence, our strategy can be applied to many other open problems in statistical estimation where heuristic statistical physics predictions are available. In particular, our proof strategy can be generalized to finite-rank matrix estimation and tensor estimation, and hence provides an alternative proof for [172] and [184] using the spatial coupling technique. Note that the validity of the replica predictions for high-rank matrix estimation problems remains an open problem where we believe that our technique can be applied.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 623–656, July 1948.

[2] D. J. Costello and G. D. Forney, "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1150–1177, June 2007.

[3] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*. World Scientific, 1987.

[4] M. Mézard and A. Montanari, *Information, Physics, and Computation*. New York, NY, USA: Oxford University Press, Inc., 2009.

[5] R. W. Hamming, "Error detecting and error correcting codes," *The Bell System Technical Journal*, vol. 29, no. 2, pp. 147–160, April 1950.

[6] M. J. E. Golay, "Notes on digital coding," *The Bell System Technical Journal*, vol. 37, p. 657, June 1949.

[7] I. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *Transactions of the IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 38–49, September 1954.

[8] D. E. Muller, "Application of boolean algebra to switching circuit design and to error detection," *Transactions of the I.R.E. Professional Group on Electronic Computers*, vol. EC-3, no. 3, pp. 6–12, September 1954.

[9] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error-correcting binary group codes," *Information and Control*, vol. 3, pp. 68–79, March 1960.

[10] A. Hocquenghem, "On a class of error-correcting binary group codes," *Chiffres*, vol. 2, pp. 147–156, 1959.

[11] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[12] P. Elias, "Coding for noisy channels," *IRE Convention Record*, vol. 4, pp. 37–46, March 1955.

[13] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Communications, 1993. ICC '93 Geneva. Technical Program, Conference Record, IEEE International Conference on*, vol. 2, May 1993, pp. 1064–1070.

[14] D. A. Spielman, "Linear-time encodable and decodable error-correcting codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1723–1731, November 1996.

[15] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 399–431, March 1999.

[16] R. G. Gallager, "Low-density parity-check codes," *Cambridge:MIT Press*, 1963.

[17] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge University Press, 2008.

[18] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 599–618, February 2001.

[19] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 569–584, February 2001.

[20] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[21] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[22] A. J. Felstrom and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2181–2191, September 1999.

[23] M. Lentmaier, S. A., K. S. Zigangirov, and D. J. Costello, "Terminated ldpc convolutional codes with thresholds close to capacity," in *Information Theory Proceedings (ISIT), 2005 International Symposium on*, September 2005, pp. 1372–1376.

[24] M. Lentmaier, A. Sridharan, J. Costello, and K. Zigangirov, "Iterative decoding threshold analysis for ldpc convolutional codes," *IEEE Tansactions on Information Theory*, vol. 56, no. 10, pp. 5274–5289, October 2010.

[25] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional ldpc ensembles perform so well over the bec," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 803–834, February 2011.

[26] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Tansactions on Information Theory*, vol. 59, no. 12, pp. 7761–7813, December 2013.

[27] S. Kumar, A. J. Young, N. Macris, and H. Pfister, "Threshold saturation for spatially-coupled ldpc and ldgm codes on bms channels," *IEEE Tansactions on Information Theory*, vol. 60, no. 12, pp. 7389–7415, December 2014.

[28] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the awgn channel capacity without active shaping," in *Proceedings of IEEE International Symposium on Information Theory*, June 1997, pp. 374–.

[29] A. Barron and A. Joseph, "Toward fast reliable communication at rates near capacity with gaussian noise," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, June 2010, pp. 315–319.

[30] A. Joseph and A. R. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Tans. on Information Theory*, vol. 58, no. 5, pp. 2541–2557, 2012.

[31] Y. Takeishi, M. Kawakita, and J. Takeuchi, "Least squares superposition codes with bernoulli dictionary are still reliable at rates up to capacity," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2737–2750, May 2014.

[32] Y. Takeishi and J. Takeuchi, "An improved upper bound on block error probability of least squares superposition codes with unbiased bernoulli dictionary," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1168–1172.

[33] A. Joseph and A. R. Barron, "Fast sparse superposition codes have near exponential error probability $R < C$," *IEEE Tansactions on Information Theory*, vol. 60, no. 2, pp. 919–942, February 2014.

[34] A. R. Barron and S. Cho, "High-rate sparse superposition codes with iteratively optimal estimates," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 120–124.

[35] S. Cho and A. R. Barron, "Approximate iterative bayes optimal estimates for high-rate sparse superposition codes," in *Sixth Workshop on Information-Theoretic Methods in Science and Engineering*, 2013.

[36] J. Barbier and F. Krzakala, "Replica analysis and approximate message passing decoder for superposition codes," in *Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on*, June 2014, pp. 1494–1498.

[37] C. Rush, A. Greig, and R. Venkataramanan, "Capacity-achieving sparse superposition codes via approximate message passing decoding," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1476–1500, March 2017.

[38] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[39] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, December 2006.

[40] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physical Review E*, vol. 84, no. 6, p. 066106, 2011.

[41] L. Massoulié, "Community detection thresholds and the weak ramanujan property," in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*. ACM, 2014, pp. 694–703.

[42] Y. Deshpande, E. Abbe, and A. Montanari, "Asymptotic mutual information for the two-groups stochastic block model," *arXiv:1507.08685*, 2015.

[43] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of computational and graphical statistics*, vol. 15, no. 2, pp. 265–286, 2006.

[44] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Annals of Probability*, pp. 1643–1697, 2005.

[45] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[46] A. Saade, F. Krzakala, and L. Zdeborová, "Matrix completion from fewer entries: Spectral detectability and rank estimation," in *Advances in Neural Information Processing Systems*, 2015, pp. 1261–1269.

[47] F. Krzakala, J. Xu, and L. Zdeborová, "Mutual information in rank-one matrix estimation," in *2016 IEEE Information Theory Workshop (ITW)*, September 2016, pp. 71–75.

[48] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[49] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, February 2001.

[50] C. Measson, A. Montanari, and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5277–5307, December 2008.

[51] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[52] A. Montanari and D. Shah, "Counting good truth assignments of random k-sat formulae," in *SODA*, 2007.

[53] M. Chertkov, "Exactness of belief propagation for some graphical models with loops," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10016, 2008.

[54] "Statistical theory of superlattices," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 150, no. 871, pp. 552–575, 1935.

[55] J. Pearl, "Reverend bayes on inference engines: A distributed hierarchical approach," in *Proceedings of the Second AAAI Conference on Artificial Intelligence*, ser. AAAI'82. AAAI Press, 1982, pp. 133–136.

[56] S. B. Korada and R. L. Urbanke, "Exchange of limits: Why iterative decoding works," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2169–2187, April 2011.

[57] D. Guo and C. c. Wang, "Multiuser detection of sparsely spread cdma," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 3, pp. 421–431, April 2008.

[58] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, September 2009, pp. 52–59.

[59] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 269–280, January 2010.

[60] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," in *Proceedings of the National Academy of Sciences*, vol. 106, November 2009, pp. 18 914–18 919.

[61] ——, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, January 2010.

[62] ——, "Message passing algorithms for compressed sensing: Ii. analysis and validation," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, January 2010.

[63] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[64] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[65] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, 2004.

[66] L. Onsager, "Electric moments of molecules in liquids," *Journal of the American Chemical Society*, vol. 58, no. 8, p. 1486–1493, 1936.

[67] D. J. Thouless, P. W. Anderson, and R. G. Palmer, "Solution of 'solvable model of a spin glass'," *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics*, vol. 35, no. 3, pp. 593–601, 1977.

[68] E. Bolthausen, "An iterative construction of solutions of the tap equations for the sherrington–kirkpatrick model," *Communications in Mathematical Physics*, vol. 325, no. 1, pp. 333–366, January 2014.

[69] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," in *Information Theory*

Proceedings (ISIT), 2010 IEEE International Symposium on, June 2010, pp. 1528–1532.

[70] E. Bolthausen, "On the high-temperature phase of the sherrington-kirkpatrick model," *Seminar at EURANDOM*, September 2009.

[71] K. Takeuchi, T. Tanaka, and T. Kawabata, "Improvement of bp-based cdma multiuser detection by spatial coupling," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, July 2011, pp. 1489–1493.

[72] C. Schlegel and D. Truhachev, "Multiple access demodulation in the lifted signal graph with spatial coupling," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, July 2011, pp. 2989–2993.

[73] S. Hamed Hassani, N. Macris, and R. Urbanke, "Threshold saturation in spatially coupled constraint satisfaction problems," *Journal of Statistical Physics*, vol. 150, no. 5, pp. 807–850, 2013.

[74] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, September 2010, pp. 347–353.

[75] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, p. 021005, May 2012.

[76] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 1231–1235.

[77] J. Barbier, C. Schülke, and F. Krzakala, "Approximate message-passing with spatially coupled structured operators, with applications to compressed sensing and sparse superposition codes," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 5, p. P05013, 2015.

[78] J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity achieving sparse superposition codes," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 4894–4927, August 2017.

[79] S. H. Hassani, N. Macris, and R. Urbanke, "Coupled graphical models and their thresholds," in *Information Theory Workshop (ITW), 2010 IEEE*, August 2010, pp. 1–5.

[80] ——, "Chains of mean-field models," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 02, p. P02011, 2012.

[81] A. Yedla, Y.-Y. Jian, P. S. Nguyen, and H. Pfister, "A simple proof of threshold saturation for coupled scalar recursions," in *7th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, August 2012, pp. 51–55.

[82] A. Yedla, Y.-Y. Jian, P. Nguyen, and H. Pfister, "A simple proof of maxwell saturation for coupled scalar recursions," *IEEE Tansactions on Information Theory*, vol. 60, no. 11, pp. 6943–6965, November 2014.

[83] V. Aref, L. Schmalen, and S. ten Brink, "On the convergence speed of spatially coupled ldpc ensembles," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, October 2013, pp. 342–349.

[84] R. El-Khatib and N. Macris, "The velocity of the propagating wave for general coupled scalar systems," in *2016 IEEE Information Theory Workshop (ITW)*, September 2016, pp. 246–250.

[85] A. Giurgiu, N. Macris, and R. Urbanke, "Spatial coupling as a proof technique and three applications," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5281–5295, October 2016.

[86] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: an Introduction*. Oxford; New York: Oxford University Press, 2001.

[87] Sourlas Nicolas, "Spin-glass models as error-correcting codes," *Nature*, vol. 339, p. 693, June 1989.

[88] A. Montanari and N. Sourlas, "The statistical mechanics of turbo codes," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 18, no. 1, pp. 107–119, November 2000.

[89] T. Tanaka, "A statistical-mechanics approach to large-system analysis of cdma multiuser detectors," *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 2888–2910, November 2002.

[90] M. Mézard and R. Zecchina, "Random k-satisfiability problem: From an analytic solution to an efficient algorithm," *Physical Review E*, vol. 66, p. 056126, November 2002.

[91] D. J. Amit, H. Gutfreund, and H. Sompolinsky, "Spin-glass models of neural networks," *Physical Review A*, vol. 32, pp. 1007–1018, August 1985.

[92] A. Engel and C. P. L. V. d. Broeck, *Statistical Mechanics of Learning*. New York, NY, USA: Cambridge University Press, 2001.

[93] R. B. Griffiths, "Rigorous Results and Theorems," in *12th School of Modern Physics on Phase Transitions and Critical Phenomena Ladek Zdroj, Poland*, 1980, pp. 7–109.

[94] M. Talagrand, "The high temperature case for the random k-sat problem," *Probability Theory and Related Fields*, vol. 119, no. 2, pp. 187–212, February 2001.

[95] F. Guerra and F. L. Toninelli, "The thermodynamic limit in mean field spin glass models," *Communications in Mathematical Physics*, vol. 230, no. 1, pp. 71–79, September 2002.

[96] M. Talagrand, "The Parisi formula," *Annals of Mathematics. Second Series*, vol. 163, no. 1, pp. 221–263, 2006.

[97] E. Gardner, "The space of interactions in neural network models," *Journal of Physics A: Mathematical and General*, vol. 21, no. 1, p. 257, 1988.

[98] Y. Kabashima, N. Sazuka, K. Nakamura, and D. Saad, "Tighter decoding reliability bound for gallager's error-correcting code," *Physical Review E*, vol. 64, p. 046113, September 2001.

[99] A. Montanari, "Tight bounds for ldpc and ldgm codes under map decoding," *IEEE Transactions on Information Theory*, vol. 51, no. 9, September 2005.

[100] S. Kudekar, "Statistical physics methods for sparse graph codes," Ph.D. dissertation, EPFL, Lausanne, 2009.

[101] S. B. Korada and N. Macris, "Tight bounds on the capacity of binary input random cdma systems," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5590–5613, November 2010.

[102] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula," in *Advances in Neural Information Processing Systems 29*, 2016, pp. 424–432.

[103] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," in *Information Theory Proceedings (ISIT), 2016 IEEE International Symposium on*, July 2016, pp. 665–669.

[104] J. Barbier, N. Macris, M. Dia, and F. Krzakala, "Mutual information and optimality of approximate message-passing in random linear estimation," *arXiv preprint arXiv:1701.05823*, 2017.

[105] M. Talagrand, *Spin glasses: a challenge for mathematicians: cavity and mean field models*. Springer, 2003.

[106] J. L. van Hemmen and R. G. Palmer, "The replica method and solvable spin glass model," *Journal of Physics A: Mathematical and General*, vol. 12, no. 4, p. 563, 1979.

[107] G. Parisi, "A sequence of approximated solutions to the S-K model for spin glasses," *Journal of Physics A Mathematical General*, vol. 13, pp. L115–L121, 1980.

[108] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," *Physical Review Letters*, vol. 35, pp. 1792–1796, December 1975.

[109] F. Guerra and F. L. Toninelli, "Quadratic replica coupling in the Sherrington-Kirkpatrick mean field spin glass model," *Journal of Mathematical Physics*, vol. 43, pp. 3704–3716, January 2002.

[110] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, July 2011, pp. 2168–2172.

[111] F. Caltagirone and L. Zdeborová, "Properties of spatial coupling in compressed sensing," *CoRR*, vol. abs/1401.6380, 2014.

[112] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear gaussian channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2384–2415, October 1998.

[113] A. G. i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.

[114] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651–4665, December 2015.

[115] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *Journal of Information and Inference*, vol. 2, no. 2, pp. 115–144, 2013.

[116] J. Barbier, M. Dia, and N. Macris, "Universal sparse superposition codes with spatial coupling and gamp decoding," *arXiv preprint arXiv:1707.04203*, 2017.

[117] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, 2012.

[118] J. Barbier, M. Dia, and N. Macris, "Proof of threshold saturation for spatially coupled sparse superposition codes," in *Information Theory Proceedings (ISIT), 2016 IEEE International Symposium on*, July 2016, pp. 1173–1177.

[119] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Wave-like solutions of general 1-d spatially coupled systems," *IEEE Tansactions on Information Theory*, vol. 61, no. 8, pp. 4117–4157, August 2015.

[120] J. Barbier, M. Dia, and N. Macris, "Threshold saturation of spatially coupled sparse superposition codes for all memoryless channels," in *Information Theory Workshop (ITW), 2016 IEEE*, September 2016, pp. 76–80.

[121] E. Bıyık, J. Barbier, and M. Dia, "Generalized approximate message-passing decoder for universal sparse superposition codes," in *Information Theory Proceedings (ISIT), 2017 IEEE International Symposium on*, June 2017, pp. 1593–1597.

[122] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," in *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*, September 2016, pp. 625–632.

[123] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," *arXiv preprint arXiv:1607.02524*, 2016.

[124] C. Condo and W. J. Gross, "Sparse superposition codes: A practical approach," in *Signal Processing Systems (SiPS), 2015 IEEE Workshop on*, October 2015, pp. 1–6.

[125] P. Zegers, "Fisher information properties," *Entropy*, vol. 17, no. 7, p. 4918, 2015. [Online]. Available: http://mdpi.com/1099-4300/17/7/4918

[126] J. Barbier, "Statistical physics and approximate message-passing algorithms for sparse linear estimation problems in signal processing and coding theory," Ph.D. dissertation, Université Paris Diderot, 2015. [Online]. Available: http://arxiv.org/abs/1511.01650

[127] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," in *Proceedings of the 31st Conference On Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 75. PMLR, July 2018, pp. 728–731.

[128] M. Mondelli, R. Urbanke, and S. H. Hassani, "How to achieve the capacity of asymmetric channels," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, September 2014, pp. 789–796.

[129] M. Dia, V. Aref, and L. Schmalen, "A compressed sensing approach for distribution matching," in *Information Theory Proceedings (ISIT), 2018 IEEE International Symposium on*, June 2018.

[130] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 5, pp. 632–647, September 1984.

[131] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for gaussian channels," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[132] G. Ungerboeck, *Huffman Shaping*. Boston, MA: Springer US, 2002, pp. 299–313.

[133] G. Böcherer and R. Mathar, "Matching dyadic distributions to channels," in *Data Compression Conference (DCC 2011)*, Snowbird, USA, March 2011, pp. 23–32.

[134] R. A. Amjad and G. Böcherer, "Fixed-to-variable length distribution matching," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 1511–1515.

[135] N. Cai, S. W. Ho, and R. W. Yeung, "Probabilistic capacity and optimal coding for asynchronous channel," in *2007 IEEE Information Theory Workshop*, September 2007, pp. 54–59.

[136] N. Baur and G. Böcherer, "Arithmetic distribution matching," in *SCC 2015; 10th International ITG Conference on Systems, Communications and Coding*, February 2015, pp. 1–6.

[137] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, January 2016.

[138] F. Steiner and G. Böcherer, "Comparison of geometric and probabilistic shaping with application to atsc 3.0," in *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*, February 2017, pp. 1–6.

[139] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *arXiv preprint arXiv:1010.5141*, 2012.

[140] U. S. Kamilov, V. K. Goyal, and S. Rangan, "Message-passing dequantization with applications to compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6270–6281, December 2012.

[141] G. Caire, S. Shamai, and S. Verdu, "Lossless data compression with error correcting codes," in *2003 IEEE International Symposium on Information Theory (ISIT)*, June 2013, pp. 22–26.

[142] ——, "Noiseless data compression with low-density parity-check codes," *Advances in Network Information Theory, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, pp. 263–284, September 2004.

[143] H. S. Cronie and S. B. Korada, "Lossless source coding with polar codes," in *2010 IEEE International Symposium on Information Theory*, June 2010, pp. 904–908.

[144] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Optics Express*, vol. 21, no. 22, pp. 25 685–25 699, November 2013.

[145] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The gn-model of fiber non-linear propagation and its applications," *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 694–721, February 2014.

[146] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "Egn model of non-linear fiber propagation," *Optics Express*, vol. 22, no. 13, pp. 16 335–16 362, June 2014.

[147] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *Journal of Lightwave Technology*, vol. 34, no. 21, pp. 5063–5073, November 2016.

[148] B. Çakmak, O. Winther, and B. H. Fleury, "S-amp: Approximate message passing for general matrix ensembles," in *2014 IEEE Information Theory Workshop (ITW 2014)*, November 2014, pp. 192–196.

[149] J. Ma and L. Ping, "Orthogonal amp," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[150] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1588–1592.

[151] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Tansactions on Information Theory*, vol. 57, no. 2, pp. 764–785, February 2011.

[152] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional ldpc ensembles perform so well over the bec," in *2010 IEEE International Symposium on Information Theory*, June 2010, pp. 684–688.

[153] J. Barbier, M. Dia, N. Macris, F. Krzakala, and L. Zdeborová, "Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method," *preprint version*, 2018.

[154] I. M. Johnstone and A. Y. Lu, "Sparse principal components analysis," *Unpublished manuscript*, vol. 7, 2004.

[155] ——, "On consistency and sparsity for principal components analysis in high dimensions," *Journal of the American Statistical Association*, 2012.

[156] P. J. Bickel and A. Chen, "A nonparametric view of network models and newman–girvan and other modularities," *Proceedings of the National Academy of Sciences*, vol. 106, no. 50, pp. 21 068–21 073, 2009.

[157] B. Karrer and M. E. Newman, "Stochastic blockmodels and community structure in networks," *Physical Review E*, vol. 83, no. 1, p. 016107, 2011.

[158] A. Saade, F. Krzakala, and L. Zdeborová, "Spectral clustering of graphs with the bethe hessian," in *Advances in Neural Information Processing Systems*, 2014, pp. 406–414.

[159] F. Ricci-Tersenghi, A. Javanmard, and A. Montanari, "Performance of a community detection algorithm based on semidefinite programming," in *Journal of Physics: Conference Series*, vol. 699, no. 1. IOP Publishing, 2016, p. 012015.

[160] B. Hajek, Y. Wu, and J. Xu, "Submatrix localization via message passing," *arXiv preprint arXiv:1510.09219*, 2015.

[161] Y. Chen and J. Xu, "Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices," *arXiv preprint arXiv:1402.1267*, 2014.

[162] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[163] R. H. Keshavan, S. Oh, and A. Montanari, "Matrix completion from a few entries," in *Information Theory (ISIT), 2009 IEEE International Symposium on*, 2009, pp. 324–328.

[164] S. Rangan and A. K. Fletcher, "Iterative estimation of constrained rank-one matrices in noise," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 1246–1250.

[165] R. Matsushita and T. Tanaka, "Low-rank matrix reconstruction and clustering via approximate message passing," in *Advances in Neural Information Processing Systems*, 2013, pp. 917–925.

[166] Y. Deshpande and A. Montanari, "Information-theoretically optimal sparse pca," in *Information Theory (ISIT), 2014 IEEE International Symposium on*, June 2014, pp. 2197–2201.

[167] T. Lesieur, F. Krzakala, and L. Zdeborová, "Phase transitions in sparse pca," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 1635–1639.

[168] F. Guerra, "An introduction to mean field spin glass theory: methods and results," *Mathematical Statistical Physics*, pp. 243–271, 2005.

[169] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and message passing algorithms," *Annals of Applied Probability*, vol. 25, no. 2, pp. 753–822, April 2015.

[170] T. Lesieur, F. Krzakala, and L. Zdeborová, "Mmse of probabilistic low-rank matrix estimation: Universality with respect to the output channel," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, September 2015, pp. 680–687.

[171] S. B. Korada and N. Macris, "Exact solution of the gauge symmetric p-spin glass model on a complete graph," *Journal of Statistical Physics*, vol. 136, no. 2, pp. 205–230, 2009.

[172] M. Lelarge and L. Miolane, "Fundamental limits of symmetric low-rank matrix estimation," *Probability Theory and Related Fields*, April 2018.

[173] J. Barbier and N. Macris, "The stochastic interpolation method: A simple scheme to prove replica formulas in bayesian inference," *arXiv preprint arXiv:1705.02780*, 2017.

[174] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," *arXiv preprint arXiv:1711.01682*, 2018.

[175] S. Franz and F. L. Toninelli, "Finite-range spin glasses in the kac limit: free energy and local observables," *Journal of Physics A: Mathematical and General*, vol. 37, no. 30, p. 7433, 2004.

[176] F. Caltagirone, S. Franz, R. G. Morris, and L. Zdeborová, "Dynamics and termination cost of spatially coupled mean-field models," *Phys. Rev. E*, vol. 89, p. 012102, January 2014.

[177] A. A. Amini and M. J. Wainwright, "High-dimensional analysis of semidefinite relaxations for sparse principal components," in *Information Theory Proceedings (ISIT), 2008 IEEE International Symposium on*. IEEE, 2008, pp. 2454–2458.

[178] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. Lanckriet, "A direct formulation for sparse pca using semidefinite programming," *SIAM review*, vol. 49, no. 3, pp. 434–448, 2007.

[179] B. Barak, S. B. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. Potechin, "A nearly tight sum-of-squares lower bound for the planted clique problem," *arXiv preprint arXiv:1604.03084*, 2016.

[180] C. Rush and R. Venkataramanan, "Finite-sample analysis of approximate message passing," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 755–759.

[181] ——, "The error exponent of sparse regression codes with amp decoding," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2478–2482.

[182] A. Greig and R. Venkataramanan, "Techniques for improving the finite length performance of sparse superposition codes," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 905–917, March 2018.

[183] A. R. Iyengar, P. H. Siegel, R. L. Urbanke, and J. K. Wolf, "Windowed decoding of spatially coupled codes," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2277–2292, April 2013.

[184] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová, "Statistical and computational phase transitions in spiked tensor estimation," *arXiv preprint arXiv:1701.08010*, 2017.