

The Bethe Free Energy Allows to Compute the Conditional Entropy of Graphical Code Instances. A Proof from the Polymer Expansion

Nicolas Macris and Marc Vuffray
 Ecole Polytechnique Fédérale de Lausanne
 LTHC-IC-EPFL
 CH-1015 Lausanne, Switzerland
 nicolas.macris@epfl.ch, marc.vuffray@epfl.ch

Abstract—The main objective of this paper is to show that the Bethe free energy associated to a Low-Density Parity-Check code used over a Binary Symmetric Channel in a large noise regime is, with high probability, asymptotically exact as the block length grows. Using the loop-sum as a starting point, we develop new techniques based on the polymer expansion from statistical mechanics for general graphical models. The true free energy is computed as a series expansion containing the Bethe free energy (or entropy) as its zero-th order term plus a series of corrections. It is easily seen that convergence criteria for such expansions are satisfied for general high-temperature models. In particular, when the graphical model has large girth the Bethe free energy is asymptotically exact. We apply these general results to ensembles of Low-Density Generator-Matrix and Parity-Check codes. While the application to Generator-Matrix codes is quite straightforward, the case of Parity-Check codes requires non-trivial extra ideas because the hard constraints correspond to a low temperature regime. Nevertheless one can combine the polymer expansion with expander and counting arguments to show that the difference between the true and Bethe free energies vanishes with high probability in the large block length limit.

I. INTRODUCTION

Often one needs to compute the free energy and/or entropy of a random graphical model. For example in the theory of codes on graphs, which is our main motivation here, it is known that the conditional input-output Shannon entropy of a graphical code used over a binary memoryless symmetric channel is related by a simple formula to the free energy of the graphical model arising in Maximum Posterior decoding. The Bethe approximation and the related belief propagation equations may sometimes offer a good starting point for computing this free energy. However it is seldom a controlled approximation and even worse it is usually not clear if it yields upper or lower bounds, or even if there is any such relationship. A very interesting general result of Vontobel [1] relates the Bethe free energy of an instance of a graphical model to the average of the true free energy over all graph covers of the instance. In the special case of Ising-like graphical models with attractive pair interactions Wainwright [2] has shown that, under additional special conditions, the Bethe free energy is a bound to the true free energy. This has been extended recently to a much wider setting (for interactions satisfying log-supermodularity conditions) by Ruzzi [3]. For counting independent sets in

sparse graphs with large girth, Chandrasekaran et al. [4] show that the Bethe free energy is asymptotically exact as the size of the graph grows. References [2] and [4] use a generic representation of the partition function developed by Chertkov and Chernyak [5] and which also forms the basis of this work.

In this paper our main objective is to show that the Bethe free energy associated to a Low-Density Parity-Check (LDPC) code used over a Binary Symmetric Channel (BSC) in a large noise regime is, with high probability, asymptotically exact as the block length grows. This is done thanks to a tool from statistical mechanics, called the polymer expansion, that is somewhat new in the context of coding theory (see however the end of this introduction for related ideas). Interestingly the polymer expansion has to be combined with special features of the graphical model associated to LDPC codes (features that are not needed in the usual applications of the polymer expansion). In fact the polymer expansion has an easy application to the case of Low-Density Generator-Matrix (LDGM) codes for high noise and more generally to graphical models in a high temperature regime. Since we believe these tools are somewhat new to the coding theory community we present these applications as well. This also serves the pedagogical purpose of introducing polymer expansions.

Let us immediately mention that we develop the analysis for the BSC only to keep the technicalities to a minimal level, but the present techniques have a wider range of validity.

A few years ago Chertkov and Chernyak [5] developed a loop-sum representation for the partition function of graphical models. The virtue of this representation is that the partition function factorizes as the product of the Bethe contribution and a finite sum of terms over subgraphs (not necessarily connected) with no dangling edges. Each term of the sum involves only belief propagation messages adjacent to the subgraphs. In [5] these subgraphs are called loops.

It is tempting to use the loop-sum representation not only as a mere formal tool, but to compare the true and Bethe free energies. One of the aims of this contribution is to develop this idea from a systematic point of view. We recognize that the loop-sum is itself the partition function of a system of polymers. A loop is the union of connected subgraphs with no dangling edges, which are called polymers. Each polymer has an associated weight which depends only on

belief propagation messages adjacent to it. By definition the polymers cannot intersect. This places a constraint that can be viewed as an infinitely repulsive pair interaction. The representation of the loop-sum as the partition function of a polymer system with infinitely repulsive interactions opens the way to the computation of the logarithm of this sum via a combinatorial expansion known in statistical mechanics as the polymer expansion [6]. If this expansion converges, then we can in principle, compute corrections to the Bethe free energy (which corresponds to the zero-th order term of the expansion) to an arbitrarily high order. If the girth of the graph is large all contributions beyond the zero-th order Bethe free energy only come from large loops and, if these contributions become small as the size of a loop increases, one may expect that, provided the expansion converges uniformly in system size, the Bethe free energy is asymptotically exact. More generally this mechanism may occur for typical instances of graphs from a random ensemble of Erdős-Rényi type, because the neighborhood of a given vertex is tree like. Conversely, when the Bethe free energy is asymptotically exact one may hope that the expansion converges and is controllable. This is of course not necessarily true as cancellations between terms in the expansion may occur. On the other hand we know of systems, such as random constraint satisfaction models (e.g. K -SAT or Q -coloring) or spin glasses, where the true free energy is definitely not given by the Bethe formula (even when averaged over the graph ensemble). For these systems it is certainly not possible for the polymer expansion to converge. The local tree-like nature of the graph is not sufficient to eliminate the contributions of large loops when long ranged correlations are present.

The program outlined above is first carried out for the case of general graphical models at high temperature. The polymer expansion starts with a zero-th order term and the rest of the series is absolutely convergent provided the temperature is large enough. We show that this has an immediate application to models whose factor graph has a large girth in the sense that the girth grows logarithmically with the size of the graph. For such models the Bethe free energy is asymptotically exact in the thermodynamic limit. Another application is to irregular LDGM codes for large noise. We show that the free energy of an instance drawn at random from an irregular LDGM ensemble is, with high probability, given by the Bethe formula in the large block length limit.

Let us now describe the results concerning LDPC codes. We consider regular LDPC codes used over a BSC. Our analysis goes through essentially unchanged for irregular codes but we refrain to present it in such generality to avoid technical complications. In the case of LDPC codes we cannot prove that the polymer expansion is absolutely convergent. The reason is that the check node constraints are not of high temperature nature but rather low (even zero) temperature. It is therefore not clear a priori why the polymer expansion should be of any use, except for the fact that the zero-th order term is the Bethe free energy. However, interestingly, using expander properties of typical instances from the LDPC ensemble we can show that a truncated form of this expansion does converge absolutely (uniformly in the system size). Moreover the convergent trun-

cated expansion accounts for the biggest part of the corrections to the Bethe free energy, up to a remainder of order $O(e^{-n\epsilon})$, $\epsilon > 0$. This remainder part is not expanded but estimated by a combinatorial counting method. The final result is again that the Bethe free energy is asymptotically exact with high probability in the large size limit.

Let us briefly comment on the connections of this work with other recent approaches. For the class of graphical models that describe communication with low density parity-check and low-density generator-matrix codes over binary-symmetric memoryless channels we have plenty of evidence that the replica-symmetric solution¹ is exact. Bounds between the replica-symmetric and true free energy were derived in [7], [8], [9], and for the special case of the binary erasure channel equality was proven in [10], [11]. These results are based on specific methods such as combinatorial calculations for the binary erasure channel, and the interpolation method for the bounds on general channels. In [12] a more generic approach is taken based on cluster expansions combined with duality. The cluster expansions used in [12] are sophisticated forms of polymer expansions. It is proven that correlations between pairs of distant (with respect to graph distance) bits decay exponentially fast for LDGM codes in the regime of large noise, and LDPC codes in the regime of small noise. This also allowed to conclude that the replica symmetric formulas are exact in these regimes for general binary-symmetric memoryless channels. A case where the cluster expansions of [12] do not work is that of LDPC codes on general channels in the regime of large noise considered here.

In the next section II we give the precise definitions of the models and briefly review the associated Bethe formulas. The main results pertaining to LDGM and LDPC codes are summarized in section III. The polymer representation and expansion are developed in section IV. This expansion is then applied to the analysis of general factor graph models and LDGM codes for large noise in sections V and VI. The more involved analysis for LDPC codes is then presented in section VII. Extensions of the method presented in the paper are discussed in VIII. For the convenience of the reader, simple derivations of the loop sum identity and polymer expansion are reviewed in a streamlined fashion in appendices A and C. Other appendixes contain more technical material needed throughout the analysis.

II. PRELIMINARIES

A. Factor graphs

We begin with a few definitions and notations. Consider two vertex sets: V a set of n variable nodes and C a set of m check nodes. We think of n and m large. We consider bipartite graphs - call them Γ - connecting V and C . The set of edges is E . When we say that Γ is random we mean that we draw it uniformly from some specified ensemble. The corresponding expectation and probability are denoted by \mathbb{E}_Γ , \mathbb{P}_Γ . Letters i, j will always denote nodes in V and letters a, b nodes in C . We

¹Replica-symmetric formulas are averaged forms of the Bethe formulas, where the average is over the channel output realizations and code ensemble.

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

For a graph Γ from a standard ensemble LDGM(Λ, P) [10] the fraction of variable nodes of degree $1 \leq s \leq l_{\max}$ is $\Lambda_s \geq 0$, and the fraction of check nodes with degree $1 \leq t \leq r_{\max}$ is $P_t \geq 0$. Of course $\sum_{s=1}^{l_{\max}} \Lambda_s = \sum_{t=1}^{r_{\max}} P_t = 1$. Here Γ is the Tanner graph of an LDGM code with design rate $r/l = n/m$, where l and r are the average variable and check nodes degree respectively. The large block length limit corresponds to $n, m \rightarrow \infty$ with fixed design rate.

For LDPC codes for simplicity we will limit ourselves to regular codes. Instead of working with the standard LDPC(l, r) ensemble with variable node degree l and check node degree r , we find it more convenient to consider a different ensemble $\mathcal{B}(l, r, n)$. This is simply the set of all bipartite (l, r) regular graphs - call them Γ - connecting V and C . In other words vertices of V have degree l , vertices of C have degree r , and there are no double edges. $\Gamma \subset \mathcal{B}(l, r, n)$ is the Tanner graph of an LDPC code with design rate $1 - l/r = 1 - m/n$. The large block length limit again corresponds to $n, m \rightarrow +\infty$ with fixed design rate.

In the case of LDPC codes we will make use of expansion arguments. For the convenience of the reader we briefly review the necessary tools [10]. We will say that Γ is a (λ, κ) expander if for every subset $\mathcal{V} \subset V$ such that $|\mathcal{V}| < \lambda n$ we have

$$|\partial \mathcal{V}| \geq \kappa l |\mathcal{V}|, \quad (1)$$

where $|\partial \mathcal{V}|$ is the number of check nodes that are connected to \mathcal{V} , and λ, κ are two positive numerical constants. Take a random $\Gamma \subset \mathcal{B}(l, r, n)$. Fix $0 < \kappa < 1 - \frac{1}{l}$ and $0 < \lambda < \lambda_0$ where λ_0 is the (only) positive solution of the equation²

$$\frac{l-1}{l} h_2(\lambda_0) - \frac{1}{r} h_2(\lambda_0 \kappa r) - \lambda_0 \kappa r h_2\left(\frac{1}{\kappa r}\right) = 0. \quad (2)$$

Then we have

$$\mathbb{P}[\Gamma \text{ is a } (\lambda, \kappa) \text{ expander}] = 1 - O\left(\frac{1}{n^{l(1-\kappa)-1}}\right). \quad (3)$$

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

B. General factor graph models

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

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

$s_{\partial a}$ and $\psi_a(s_{\partial a})$. The partition function of the factor graph model (or spin system) is

$$Z = \sum_{\underline{s} \in \{-1, +1\}^n} \prod_{a \in C} \psi_a(s_{\partial a}). \quad (4)$$

The free energy is defined by

$$f = \frac{1}{n} \ln Z \quad (5)$$

and the thermodynamic limit is the limit $n \rightarrow +\infty$.

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

$$\psi_a(s_{\partial a}) = \exp\left\{\beta \sum_{I \subset \partial a} J_I \prod_{i \in I} s_i\right\}, \quad (6)$$

where $\beta > 0$ has the interpretation of an inverse temperature and $J_I \in \mathbb{R}$ have the interpretation of coupling constants³. When we speak of a high temperature regime it is meant that $\beta > 0$ is small enough so that

$$\mu \equiv 2\beta \sup_{a \in C} \sum_{I \subset \partial a} |J_I| \ll 1. \quad (7)$$

We remark for later use that in a high temperature regime

$$|\psi_a(\{s_{\partial a}\}) - 1| \leq 2\beta \sup_{a \in C} \sum_{I \subset \partial a} |J_I| = \mu. \quad (8)$$

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

C. Transmission with LDGM codes.

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

$$x_a = \bigoplus_{i \in \partial a} u_i. \quad (9)$$

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

$$p_{\underline{U}|\underline{Y}}(\underline{u}|\underline{y}) = \frac{1}{Z_{\text{LDGM}}} \prod_{a \in C} e^{h_a \prod_{i \in \partial a} (-1)^{u_i}}. \quad (10)$$

In this expression

$$h_a = (-1)^{y_a} \frac{1}{2} \ln \frac{1-p}{p} \quad (11)$$

are the half-log-likelihood variables and

$$Z_{\text{LDGM}} = \sum_{\underline{u} \in \{0, 1\}^n} \prod_{a \in C} e^{h_a \prod_{i \in \partial a} (-1)^{u_i}} \quad (12)$$

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

³Units are suitably chosen so that βJ_I is dimensionless.

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

$$|h_a| \equiv h \equiv \frac{1}{2} \ln \frac{1-p}{p}. \quad (13)$$

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

$$c(h_a) = (1-p)\delta(h_a - h) + p\delta(h_a + h). \quad (14)$$

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

Remark 1. Equ. (12) is the partition function of a spin system with one coupling constant $\beta J_I \rightarrow h_a$ per check, and the high temperature regime (8) simply corresponds to $h \ll 1$.

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

$$f_{\text{LDGM}} = \frac{1}{n} \ln Z_{\text{LDGM}} \quad (15)$$

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

$$\frac{1}{n} H_{\text{LDGM}}(\underline{U}|\underline{Y}) = \mathbb{E}_{\underline{h}} [f_{\text{LDGM}}] - \frac{l}{r} \frac{1-2p}{2} \ln \frac{1-p}{p}. \quad (16)$$

D. Transmission with LDPC codes.

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

$$p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{1}{Z_{\text{LDPC}}} \prod_{a \in C} \mathbb{I}(\oplus_{i \in \partial a} x_i = 0) \prod_{i \in V} \exp((-1)^{x_i} h_i). \quad (17)$$

In this formula

$$h_i = (-1)^{y_i} \frac{1}{2} \ln \frac{1-p}{p} \quad (18)$$

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

$$Z_{\text{LDPC}} = \sum_{\underline{x} \in \{0,1\}^n} \prod_{a \in C} \mathbb{I}(\oplus_{i \in \partial a} x_i = 0) \prod_{i \in V} \exp((-1)^{x_i} h_i). \quad (19)$$

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

Remark 2. Equ. (19) is the partition function of a spin system with two types of coupling constants $\beta J_I \rightarrow h_i$ and $+\infty$. The infinite coupling constant mimics the parity check constraints, so the high temperature condition 8 is never satisfied which makes the ensuing analysis more challenging.

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

$$f_{\text{LDPC}} = \frac{1}{n} \ln Z_{\text{LDPC}} \quad (20)$$

through the formula

$$\frac{1}{n} H_{\text{LDPC}}(\underline{X}|\underline{Y}) = \mathbb{E}_{\underline{h}} [f_{\text{LDPC}}] - \frac{1-2p}{2} \ln \frac{1-p}{p}. \quad (21)$$

E. The Bethe Approximation

The Bethe-Peierls (mean field) theory allows to compute candidate approximations, called Bethe free energies, for $f = \frac{1}{n} \ln Z$. In the case of LDPC and LDGM it allows to compute candidate approximations for the free energies f_{LDPC} and f_{LDGM} . As explained in the introduction, controlling in a rather systematic way the quality of these approximation is the object of this paper.

Let us first recall the Bethe formulas for general factor graph models. This involves a set of messages $\zeta_{i \rightarrow a}$ and $\hat{\zeta}_{a \rightarrow i}$ attached to the edges of $(ia) \in E$. The collection of all messages is denoted $(\underline{\zeta}, \underline{\hat{\zeta}})$; they satisfy the belief propagation fixed point equations

$$\begin{cases} \zeta_{i \rightarrow a} = \sum_{b \in \partial i \setminus a} \hat{\zeta}_{b \rightarrow i} \\ \tanh \hat{\zeta}_{a \rightarrow i} = \frac{\sum_{s \in \partial a} \psi_a(s \partial a) \prod_{j \in \partial a \setminus i} (1 + s_j \tanh \zeta_{j \rightarrow a})}{\sum_{s \in \partial a} \psi_a(s \partial a) \prod_{j \in \partial a \setminus i} (1 + s_j \tanh \zeta_{j \rightarrow a})}. \end{cases} \quad (22)$$

The Bethe free energy associated to a particular solution of these equations is

$$f^{\text{Bethe}}(\underline{\zeta}, \underline{\hat{\zeta}}) = \frac{1}{n} \left(\sum_{a \in C} F_a + \sum_{i \in V} F_i - \sum_{(i,a) \in E} F_{ia} \right), \quad (23)$$

where

$$\begin{cases} F_a = \ln \left\{ \sum_{s \in \partial a} \psi_a(s \partial a) \prod_{j \in \partial a} \left(\frac{1 + s_j \tanh \zeta_{j \rightarrow a}}{2} \right) \right\}, \\ F_i = \ln \left\{ \prod_{a \in \partial i} (1 + \tanh \hat{\zeta}_{a \rightarrow i}) \right. \\ \quad \left. + \prod_{a \in \partial i} (1 - \tanh \hat{\zeta}_{a \rightarrow i}) \right\}, \\ F_{ia} = \ln \{ 1 + \tanh \zeta_{i \rightarrow a} \tanh \hat{\zeta}_{a \rightarrow i} \}. \end{cases} \quad (24)$$

It is easy to check that the stationary points of $f^{\text{Bethe}}(\underline{\zeta}, \underline{\hat{\zeta}})$ considered as a function of the messages over $\mathbb{R}^E \times \mathbb{R}^E$ satisfy the Belief propagation equations.

It is immediate to specialize these formulas to LDGM codes. This yields

$$\begin{cases} \zeta_{i \rightarrow a} = \sum_{b \in \partial i \setminus a} \hat{\zeta}_{b \rightarrow i} \\ \hat{\zeta}_{a \rightarrow i} = \tanh^{-1}(\tanh h_a \prod_{j \in \partial a \setminus i} \tanh \zeta_{j \rightarrow a}). \end{cases} \quad (25)$$

and

$$\begin{cases} F_a = \ln \{ 1 + \tanh h_a \prod_{i \in \partial a} \tanh \zeta_{i \rightarrow a} \} + \ln \cosh h, \\ F_i = \ln \left\{ \prod_{a \in \partial i} (1 + \tanh \zeta_{a \rightarrow i}) \right. \\ \quad \left. + \prod_{a \in \partial i} (1 - \tanh \zeta_{a \rightarrow i}) \right\}, \\ F_{ia} = \ln \{ 1 + \tanh \zeta_{i \rightarrow a} \tanh \hat{\zeta}_{a \rightarrow i} \}. \end{cases} \quad (26)$$

The Bethe free energy given by a sum of these three type of quantities and is denoted by $f_{\text{LDGM}}^{\text{Bethe}}(\underline{\zeta}, \widehat{\underline{\zeta}})$.

Since LDPC codes will require a separate treatment, in order to avoid confusions, the messages are denoted $(\underline{\eta}, \widehat{\underline{\eta}})$. They satisfy the belief propagation fixed point equations

$$\begin{cases} \eta_{i \rightarrow a} = h_i + \sum_{b \in \partial i \setminus a} \widehat{\eta}_{b \rightarrow i} \\ \widehat{\eta}_{a \rightarrow i} = \tanh^{-1} \left(\prod_{j \in \partial a \setminus i} \tanh \eta_{j \rightarrow a} \right). \end{cases} \quad (27)$$

The Bethe free energy associated to a solution is

$$f_{\text{LDPC}}^{\text{Bethe}}(\underline{\eta}, \widehat{\underline{\eta}}) = \frac{1}{n} \left(\sum_{a \in C} P_a + \sum_{i \in V} P_i - \sum_{(i,a) \in E} P_{ia} \right), \quad (28)$$

where

$$\begin{cases} P_a = \ln \{ 1 + \prod_{i \in \partial a} \tanh \eta_{i \rightarrow a} \} - \ln 2, \\ P_i = \ln \{ e^{h_i} \prod_{a \in \partial i} (1 + \tanh \eta_{a \rightarrow i}) \\ \quad + e^{-h_i} \prod_{a \in \partial i} (1 - \tanh \eta_{a \rightarrow i}) \}, \\ P_{ia} = \ln \{ 1 + \tanh \eta_{i \rightarrow a} \tanh \widehat{\eta}_{a \rightarrow i} \}. \end{cases} \quad (29)$$

Considering the expression on the right hand side of (28) as a function of $(\underline{\eta}, \widehat{\underline{\eta}}) \in \mathbb{R}^E \times \mathbb{R}^E$, allows to check that its stationary points are solutions of the belief propagation equations.

III. MAIN RESULTS

A. Factor graphs with large girth

We begin with the high temperature regime of general factor graph models. It has been proven [13] that when the high temperature condition (8) is satisfied the belief propagation equations have a unique fixed point solution. In the next theorem the Bethe free energy is computed at this fixed point.

Theorem 3. *Let Γ_n be a sequence of Tanner graphs, with uniformly bounded degrees and, with large girth in the sense that $\text{girth}(\Gamma_n) \geq C \ln |\Gamma_n|$ where $C > 0$ is a numerical constant. Consider free energy sequences of models constructed on Γ_k . For $0 < \beta < \beta_0$ small enough we have*

$$\lim_{n \rightarrow +\infty} |f - f^{\text{Bethe}}(\underline{\zeta}, \widehat{\underline{\zeta}})| = 0. \quad (30)$$

Remark 4. Even if the individual limits of f and f^{Bethe} are not well defined their difference tends to zero. As will be seen in the proof, the order of magnitude of this difference is $O((c\beta)^{2\text{girth}/(2+r_{\max})})$ with $c > 0$ a constant depending only on the degrees of the nodes and the couplings J_I .

B. LDGM ensembles

For h small enough, an instance of an LDGM code is a high temperature graphical model with a special form of the factor weights. If the LDGM code contains no degree one check nodes then $(\underline{\zeta}, \widehat{\underline{\zeta}}) = (0, 0)$ is the trivial fixed point. However if there is a non-vanishing fraction of degree one check nodes the fixed point $(\underline{\zeta}, \widehat{\underline{\zeta}})$ is non-trivial.

Theorem 5. *Suppose that we draw Γ uniformly at random from the ensemble $\text{LDGM}(\Lambda, P, n)$. For $h < h_0$ small enough we have*

$$\lim_{n \rightarrow +\infty} \mathbb{E}_{\Gamma} \left[\left| f_{\text{LDGM}} - f_{\text{LDGM}}^{\text{Bethe}}(\underline{\zeta}, \widehat{\underline{\zeta}}) \right| \right] = 0. \quad (31)$$

C. LDPC ensemble

Let us now describe our main result which is the analogous theorem for LDPC codes. We assume that for $h < h_*$ small enough independent of n and $\epsilon > 0$ independent of n and h , there exist a high noise solution $(\underline{\eta}, \widehat{\underline{\eta}})$ of the belief propagation equations which satisfies (see appendix B)

$$|\tanh \eta_{i \rightarrow a}| \leq (1 + \epsilon) \tanh h. \quad (32)$$

The analysis does not require the unicity of this solution but only its existence. We call such solutions ‘‘high noise solutions’’. For simplicity of notations we use throughout the analysis the quantity θ define as

$$\theta = (1 + \epsilon) \tanh h. \quad (33)$$

Theorem 6. *Suppose l is odd and $3 \leq l \leq r$. There exists $\theta_0 > 0$ (small) independent of n , such that for $\theta \leq \theta_0$ and any high-noise-solution*

$$\mathbb{E}_{\Gamma} \left[\left| \frac{1}{n} \ln Z - f_{\text{LDPC}}^{\text{Bethe}}(\underline{\eta}, \widehat{\underline{\eta}}) \right| \right] = O(n^{-(l(1-\kappa)-1)}). \quad (34)$$

The $O(\cdot)$ is uniform in the channel output realizations \underline{h} .

Remark 7. We recall that $\kappa \in [1 - \frac{2(r-1)}{lr}, 1 - \frac{1}{l}]$ which implies that the expansion constant κ is such that, for $r > 2$, $0 < l(1-\kappa) - 1 < (r-2)/r$.

IV. LOOP SUM AND POLYMER EXPANSION

The formalism developed in this section is valid for general graphical models, and in particular for fixed instances of LDPC and LDGM codes. We give only the necessary information needed for the subsequent analysis in sections V, VI, VII. More details can be found in appendices A, C.

A. Polymer representation

Take a subset of edges of Γ together with the end-vertices of these edges. This forms a subgraph g of Γ . We call $d_i(g)$ (resp. $d_a(g)$) the induced degree of node i (resp. a) in a subgraph g . If $d_i(g) \geq 2$ and $d_a(g) \geq 2$ for all $i, a \in g$, we say that g is a loop. In other words a loop has no dangling edge. Note that a loop is not necessarily a cycle, and is not necessarily connected. Figure 1 shows an example.

For a finite size system, Chertkov and Cherniak [5] derived the following loop-sum identity

$$Z = \exp(n f^{\text{Bethe}}) \times \left(1 + \sum_{g \subset \Gamma} K(g) \right). \quad (35)$$

where each quantity on the right hand side is computed for a solution of the belief propagation equations. The sum on the right hand side carries over all loops included in Γ . As long as the graph is finite, this is a finite sum which is well defined. The quantities $K(g)$ can be expressed entirely in terms of

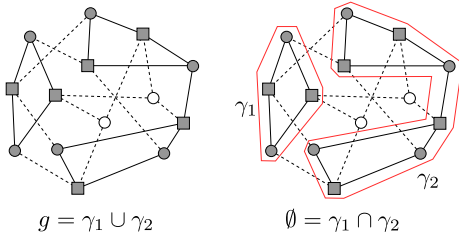


Figure 1. Example of $\Gamma \in \mathcal{B}(3, 4, 8)$. The generalized loop g has two disjoint connected parts γ_1 and γ_2 .

belief propagation messages $(\zeta_{i \rightarrow a}, \hat{\zeta}_{a \rightarrow i})$ or $(\eta_{i \rightarrow a}, \hat{\eta}_{a \rightarrow i})$ such that i or a belong to g . The explicit formulas for general models as well as for LDPC and LDGM codes are given in appendix B. For the convenience of the reader we give a short derivation of identity (35) in appendix A.

Each generalized loop can be decomposed in a unique way as a union $g = \cup_k \gamma_k$ where γ_k are connected and disjoint loops (see figure 1). The connected parts γ_k are called polymers. Remarkably each $K(g)$ can be factorized (see appendix A, eqn. (124)) in a product of contributions associated to the connected parts of g . We have⁴

$$K(g) = \prod_k K(\gamma_k). \quad (36)$$

The factorization implies

$$1 + \sum_{g \subset \Gamma} K(g) \equiv Z^{\text{polymer}}, \quad (37)$$

with

$$Z^{\text{polymer}} = \sum_{M \geq 0} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M \subset \Gamma} \prod_{k=1}^M K(\gamma_k) \times \prod_{k < k'} \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset). \quad (38)$$

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

From a physical point of view (38) interprets the loop sum in Equ. (35) as the *partition function of a gas of polymers* that can acquire any shape allowed by Γ , have *activity*⁵ $K(\gamma)$, and interact via a two body hard-core repulsion which precludes their overlap. This analogy allows us to use methods from statistical mechanics to analyze the corrections to the Bethe free energy.

⁴We note that this factorization is not necessarily unique and in practice one should choose the most natural one.

⁵“Activity” is the name used by chemists for the prior probability weight $K(\gamma)$ of an isolated polymer. Note that here $K(\gamma)$ can be negative and this analogy is at best formal. We use the name “activity” rather than “prior weight” for $K(\gamma)$ precisely because they can be negative in the present context.

B. Polymer expansion

All the corrections to the Bethe free energy are contained in the free energy of the polymer gas, namely

$$\frac{1}{n} \ln Z = f^{\text{Bethe}} + \frac{1}{n} \ln Z^{\text{polymer}}. \quad (39)$$

We start with a heuristic discussion in order to motivate the ensuing formalism. If a suitable fixed point of the belief propagation equations is chosen such that the Bethe free energy is a good approximation, then we expect that the polymer free energy is small (or negligible in an appropriate limit). One way that this may happen is if the activities of the polymers become small as the size of the polymers increase. Let us explain this point in more detail. We expect the activities to be exponentially small in $|\gamma|$ (as will become clear later for LDPC and LDGM this is true for small h). This smallness of the activities is counterbalanced by an entropic contribution that accounts for the large number of polymers of given size. This number is exponentially large in $|\gamma|$. For h small enough the smallness of the activities wins over the entropic terms and one can expand the log in a power series in $K(\gamma)$. Since the polymers have no dangling edges, on a locally tree like graph they have a typical size $|\gamma| \approx c \ln n$ for some small constant c . This means $K(\gamma) \approx O(e^{-c \ln n})$ and since the series expansion starts linearly with $K(\gamma)$, the polymer free energy is itself $O(e^{-c \ln n})$. Note that the polymer free energy could still be negligible even if the activities are not small because in general they have signs and cancellations could occur. However such cancellations would be difficult to control. The regimes investigated in this paper are those where the activities are small enough so that their weight counterbalances the entropy of the polymers and we do not need to track sign cancellations.

As just explained, with small activities it is natural to expand the logarithm on the right hand side of (39) and to control the convergence of this expansion. Such expansions are called *polymer expansions*⁶. We now describe the combinatorial structure of the polymer expansion and explain what convergence criteria are available.

We introduce the set \mathcal{G}_M of all connected graphs G with M labeled vertices $1, \dots, M$ (see figure 2). These are called *Mayer graphs*. To each Mayer graph G we associate an Ursell function defined as

$$U(G) = \prod_{(k, k') \in G} \left\{ \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset) - 1 \right\}. \quad (40)$$

This function takes values ± 1 if and only if any pair of polymers $\gamma_k, \gamma_{k'}$ labeling two vertices k, k' of G connected by an edge intersect in Γ , $\gamma_k \cap \gamma_{k'} \neq \emptyset$ (see figure 2). Expanding the \ln in the free energy of polymers in powers of the activities yields the expansion

$$\frac{1}{n} \ln Z^{\text{polymer}} = \frac{1}{n} \sum_{M=1}^{+\infty} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M \subset \Gamma} \prod_{k=1}^M K(\gamma_k) \sum_{G \in \mathcal{G}_M} U(G). \quad (41)$$

⁶They are also called Mayer expansions. another more generic name is cluster expansion.

In this formula the third sum on the right hand side runs over the set of all Mayer graphs $G \in \mathcal{G}_M$. A short check of this identity is given in appendix C.

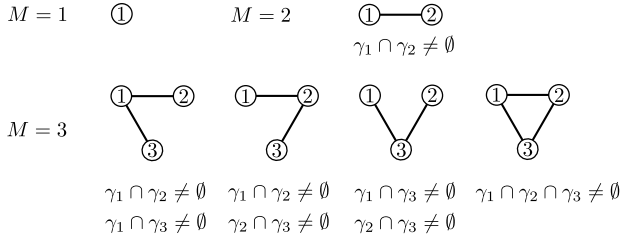


Figure 2. All the Mayer graphs for $M = 1, 2, 3$.

It is important to note that now the first two sums on the right hand side are infinite because the Ursell function forces polymers to overlap. It is therefore important to control the convergence of this formal power series. A standard criterion for uniform (with respect to n) convergence is that

$$Q \equiv \sum_{t=0}^{\infty} \frac{1}{t!} \sup_{x \in V \cup C} \sum_{\gamma \ni x} |\gamma|^t |K(\gamma)| < 1. \quad (42)$$

This implies in particular that the polymer free energy is analytic as a function of $\{K(\gamma), \gamma \subset \Gamma\}$.

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

$$\frac{1}{n} \ln Z^{\text{polymer}}(z) \quad (43)$$

and (41) becomes a series expansion in z^M , $M \geq 1$. If the convergence criterion (42) holds with $K(\gamma)$ replaced by $z_0 K(\gamma)$ we can conclude that the series is holomorphic for $|z| < z_0$. Moreover the limit $n \rightarrow +\infty$, as long as it exists, is also holomorphic for $|z| < z_0$. In practice, existence of the limit requires some regularity structure on the sequence of graphical models (which is not the case in the present formulation), and it can be checked term by term on the series expansion. We take $z_0 > 1$ in order to then apply the results to the case of interest $z = 1$.

As will be seen in sections V, VI it is fairly easy to check that (42) is satisfied for high temperature general models and also for typical instances of LDGM codes in the large noise regime. This case also serves as a pedagogical one to better understand the difficulties that arise in the case of LDPC codes. In fact for LDPC codes we are not able to satisfy this criterion as such. However the criterion holds if Γ is an expander and the sum in Z^{polymer} is restricted to small polymers of size $|\gamma| < \lambda n$, $0 < \lambda < \lambda_0$ (recall λ_0 is defined in section II-A). The contribution of “large” polymers $|\gamma| > \lambda n$ is treated differently.

V. HIGH TEMPERATURE MODELS

When the high temperature condition (8) is satisfied the unique fixed point solution of the belief propagation equations satisfies

$$|\tanh \zeta_{i \rightarrow a}| \leq 2(l_{\max} - 1)\mu, \quad |\tanh \hat{\zeta}_{a \rightarrow i}| \leq 2\mu. \quad (44)$$

Lemma 8. Consider the z dependent free energy defined in (43) computed at the fixed point (44). One can find a $\beta_0 > 0$ small enough such that for $0 < \beta < \beta_0$, such that:

- 1) $n^{-1} \ln Z^{\text{polymer}}(z)$ has an absolutely uniformly (in n) convergent power series expansion in z^M , $M \geq 1$ for $|z| < z_0$.
- 2) If one considers a sequence of factor graph models such that the thermodynamic limit $\lim_{n \rightarrow +\infty} n^{-1} \ln Z^{\text{polymer}}(z)$ exists, this limit is an analytic function of z for $|z| < z_0$.
- 3)

$$\frac{1}{n} |\ln Z^{\text{polymer}}(z)| \leq \frac{2}{n} \sum_{x \in V \cup C} \sum_{\gamma \ni x} z_0^{|\gamma|} (6e\mu)^{\frac{2|\gamma|}{2+r_{\max}}} e^{|\gamma|}. \quad (45)$$

Remark 9. Note that the second statement is an immediate consequence of the first one. Later we make use of the third statement for $z = 1$.

Proof: For the activities of the polymers computed at the fixed point we have the bounds (appendix B)

$$|K(\gamma)| \leq (6e\mu)^{\frac{2|\gamma|}{2+r_{\max}}}. \quad (46)$$

Next we use the remarkable inequality [6]

$$\left| \sum_{G \in \mathcal{G}_M} U(G) \right| \leq \sum_{T \in \mathcal{T}_M} |U(T)|, \quad (47)$$

where \mathcal{T}_M is the set of trees on M vertices labeled $1, \dots, M$. Using (46) and (47) we find that the term of order M in (41) is smaller than

$$\frac{1}{M!} \sum_{T \in \mathcal{T}_M} \sum_{\gamma_1, \dots, \gamma_M} \prod_{k=1}^M z_0^{|\gamma_k|} (6e\mu)^{\frac{2|\gamma_k|}{2+r_{\max}}} |U(T)| \quad (48)$$

Notice that in this formula,

$$|U(T)| = \prod_{(k, k') \in T} \mathbb{I}(\gamma_k \cap \gamma_{k'} \neq \emptyset) \quad (49)$$

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

$$\begin{aligned} & \sum_{\gamma_2, \dots, \gamma_M} \prod_{k=2}^M z_0^{|\gamma_k|} (6e\mu)^{\frac{2|\gamma_k|}{2+r_{\max}}} \prod_{(k, k') \in T \setminus \{1\}} \mathbb{I}(\gamma_k \cap \gamma_{k'} \neq \emptyset) \\ & \leq |\gamma_1|^{t_1-1} \prod_{k=2}^M \left\{ \sup_{x \in V \cup C} \sum_{\gamma \ni x} |\gamma|^{t_k-1} z_0^{|\gamma|} (6e\mu)^{\frac{2|\gamma|}{2+r_{\max}}} \right\}. \end{aligned} \quad (50)$$

This implies

$$\begin{aligned} & \sum_{\gamma_1, \dots, \gamma_M} \prod_{k=1}^M z_0^{|\gamma_k|} (6e\mu)^{\frac{2|\gamma_k|}{2+r_{\max}}} \prod_{(k, k') \in T} \mathbb{I}(\gamma_k \cap \gamma_{k'} \neq \emptyset) \\ & \leq \sum_{y \in VUC} \sum_{\gamma_1 \ni y} z_0^{|\gamma_1|} (6e\mu)^{\frac{2|\gamma_1|}{2+r_{\max}}} |\gamma_1|^{t_1-1} \\ & \quad \times \prod_{k=2}^M \left\{ \sup_{x \in VUC} \sum_{\gamma \ni x} |\gamma|^{t_k-1} z_0^{|\gamma|} (6e\mu)^{\frac{2|\gamma|}{2+r_{\max}}} \right\}. \end{aligned} \quad (51)$$

Now it is easy to estimate the sum over T in (48). According to the Cayley formula the number of trees with M vertices of degrees t_1, \dots, t_M is equal to

$$\frac{(M-2)!}{(t_1-1)! \dots (t_M-1)!}, \quad (52)$$

so we find that (48) is upper bounded by

$$\begin{aligned} & \sum_{y \in VUC} \sum_{\gamma_1 \ni y} (z_0 \beta J)^{|\gamma_1|} e^{|\gamma_1|} \frac{1}{M(M-1)} \\ & \quad \times \left\{ \sum_{t=0}^{+\infty} \frac{1}{t!} \sup_{x \in VUC} \sum_{\gamma \ni x} |\gamma|^t z_0^{|\gamma|} (6e\mu)^{\frac{2|\gamma|}{2+r_{\max}}} \right\}^{M-1}. \end{aligned} \quad (53)$$

We will check that in this expression the quantity in brackets

$$Q^* \equiv \sum_{t=0}^{+\infty} \frac{1}{t!} \sup_{x \in VUC} \sum_{\gamma \ni x} |\gamma|^t z_0^{|\gamma|} (6e\mu)^{\frac{2|\gamma|}{2+r_{\max}}}. \quad (54)$$

can be made smaller than $1/2$ for β_0 small enough. This implies the first statement of the lemma.

The number of polymers attached to any vertex x is bounded by $e^{A|\gamma|}$ where $A > 0$ is a numerical constant depending only on the maximal degrees of Γ . So one finds

$$\begin{aligned} Q^* & \leq \sum_{t=1}^{+\infty} z_0^t (6e\mu)^{\frac{2t}{2+r_{\max}}} e^{t(A+1)} \\ & \leq \frac{z_0 (6e\mu)^{\frac{2}{2+r_{\max}}} e^{(A+1)}}{1 - z_0 (6e\mu)^{\frac{2}{2+r_{\max}}} e^{(A+1)}} \leq \frac{1}{2} \end{aligned} \quad (55)$$

Summing (53) over $M \geq 1$ and using $Q^* \leq 1/2$ yields the third statement (45). \blacksquare

An immediate application of this lemma is the proof of theorem 3.

Proof: [Proof of theorem 3] Let n the number of variable nodes of the graph Γ_n . Proving Equ. (30) is equivalent to

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \ln Z^{\text{polymer}} = 0. \quad (56)$$

To show this we will use estimate (45) for $z = 1$. The graphs Γ_n have large girth, and since a polymer $\gamma \subset \Gamma_n$ attached at x certainly contains at least one closed cycle, we have $|\gamma| \geq C \ln n$ (for $C > 0$ not too large). Using this fact and that the number of such polymers is less than $e^{A|\gamma|}$ we find

$$\begin{aligned} \frac{1}{n} |\ln Z^{\text{polymer}}| & \leq \frac{2}{n} \sum_{x \in \Gamma_n} \sum_{\gamma \ni x} (6e\mu)^{\frac{2|\gamma|}{2+r_{\max}}} e^{|\gamma|} \\ & \leq \sum_{t \geq C \ln n}^{+\infty} (6e\mu)^{\frac{2t}{2+r_{\max}}} e^{t(A+1)}. \end{aligned} \quad (57)$$

Clearly there exist $\beta_0 > 0$ such that this estimate tends to zero as $n \rightarrow +\infty$ for $\beta < \beta_0$. In fact we have that $n^{-1} \ln |Z^{\text{polymer}}| = O(\beta^{2\text{girth}/(2+r_{\max})})$. \blacksquare

VI. ANALYSIS FOR LDGM CODES

For h small enough the fixed point solution of the belief propagation equations of an LDGM(Λ, P) ensemble satisfies

$$|\tanh \zeta_{i \rightarrow a}| \leq 4(l_{\max} - 1)h, \quad |\tanh \hat{\zeta}_{a \rightarrow i}| \leq 4h. \quad (58)$$

For the activities of the polymers computed at the fixed point we have the bounds (appendix B)

$$|K(\gamma)| \leq (12eh)^{2|\gamma|/(2+r_{\max})}. \quad (59)$$

Therefore lemma 8 applies with βJ replaced by h . This allows to prove theorem 5.

Proof: [Proof of theorem 5] For h small enough Lemma 8 implies

$$\frac{1}{n} |\ln Z_{\text{LDGM}}^{\text{polymer}}| \leq \frac{2}{n} \sum_{x \in VUC} \sum_{\gamma \ni x} (12eh)^{\frac{2|\gamma|}{2+r_{\max}}} e^{|\gamma|} \quad (60)$$

Taking the expectation of this inequality,

$$\mathbb{E}_{\Gamma} \left[\frac{1}{n} |\ln Z_{\text{LDGM}}^{\text{polymer}}| \right] \leq \mathbb{E}_{\Gamma} \left[\sum_{\gamma \ni o} (12eh)^{\frac{2|\gamma|}{2+r_{\max}}} e^{|\gamma|} \right]. \quad (61)$$

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

$$1 - O\left(\frac{C_{l_{\max}, r_{\max}, R}}{n}\right), \quad (62)$$

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

$$\begin{aligned} & \mathbb{E}_{\Gamma} \left[\sum_{\gamma \ni o} (12eh)^{\frac{2|\gamma|}{2+r_{\max}}} e^{|\gamma|} \right] \\ & \leq (1 - O\left(\frac{C_{l_{\max}, r_{\max}, R}}{n}\right)) \sum_{t \geq R} ((12eh)^{\frac{2}{2+r_{\max}}} e^{A+1})^t \\ & \quad + O\left(\frac{C_{l_{\max}, r_{\max}, R}}{n}\right) \sum_{t \geq 3} ((12eh)^{\frac{2}{2+r_{\max}}} e^{A+1})^t. \end{aligned} \quad (63)$$

Replacing this estimate in (61) and taking the limit $n \rightarrow +\infty$,

$$\lim_{n \rightarrow +\infty} \mathbb{E}_{\Gamma} \left[\frac{1}{n} |\ln Z_{\text{LDGM}}^{\text{polymer}}| \right] \leq \sum_{t \geq R} ((12eh)^{\frac{2}{2+r_{\max}}} e^{A+1})^t. \quad (64)$$

Finally, taking the limit $R \rightarrow +\infty$ ends the proof. \blacksquare

VII. ANALYSIS FOR LDPC CODES

Recall that $\theta = (1 + \epsilon) \tanh h$ (eqn. (33)). From (32) we deduce in appendix B a (qualitatively) optimal estimate (166), (167) on the activity of a polymer.

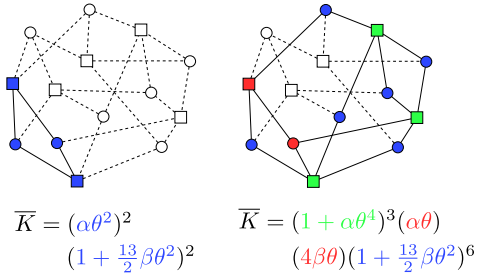


Figure 3. Example for a $\Gamma \in \mathcal{B}(3, 4, 8)$ of polymers and their associated activities. On the left a small polymer and on the right a large polymer.

A. Contribution of small polymers

Estimate (166), (167) in appendix B is quite cumbersome, so let us begin with a few remarks to understand its main qualitative features. The activity $K(\gamma)$ is not necessarily very small for graphs containing too many check nodes of maximal induced degree and too many variable nodes of even induced degree. More precisely for these “bad graphs” the rate of decay as $|\gamma|$ grows is too slow even for θ small, and it is not clear that it counterbalances the exponentially large entropic terms. However the rate of decay as $|\partial\gamma \cap C|$ grows is large for θ small. Here the boundary $\partial\gamma \cap C$ is by definition the set of check nodes in γ of non-maximal induced degree. An example is shown on figure 3.

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

Lemma 10. *Assume that Γ is a (λ, κ) expander with $\kappa \in]1 - \frac{2(r-1)}{lr}, 1 - \frac{1}{l}[$. For $|\gamma| < \lambda n$ we have for θ small enough*

$$|K(\gamma)| \leq \theta^{\frac{c}{2}|\gamma|}, \quad (65)$$

with

$$c = r - \frac{2+r}{3-l(1-\kappa)}. \quad (66)$$

Remark 11. In the process of this derivation one has to require $3-l(1-\kappa) > 0$ and $c > 0$. This imposes the condition on the expansion constant $\kappa > 1 - \frac{2(r-1)}{lr}$. Note that an expansion constant cannot be greater than $1 - 1/l$, so it is fortunate that we have $1 - \frac{1}{l} > 1 - \frac{2(r-1)}{lr}$ (for any $r > 2$).

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

$$\begin{cases} |\gamma| = \sum_{s=2}^l n_s(\gamma) + \sum_{t=2}^r m_t(\gamma) \\ \sum_{s=2}^l s n_s(\gamma) = \sum_{t=2}^r t m_t(\gamma) \end{cases} \quad (67)$$

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

$$|\partial\mathcal{V}| \geq \kappa l \sum_{s=2}^l n_s(\gamma). \quad (68)$$

On the other hand $|\partial\mathcal{V}| \leq \sum_{t=2}^r m_t(\gamma) + \sum_{s=2}^l (l-s)n_s(\gamma)$. With (67) and (68) this yields the constraint

$$\sum_{t=2}^r (r-t)m_t(\gamma) \geq -|\gamma|l(1-\kappa) + (l(1-\kappa) + r-1) \sum_{t=2}^r m_t(\gamma). \quad (69)$$

Relaxing the second constraint in (67) gives

$$\sum_{t=2}^r t m_t(\gamma) \geq 2 \sum_{s=2}^l n_s(\gamma). \quad (70)$$

Combined with the first constraint of (67) this yields

$$(r+2) \sum_{t=2}^r m_t(\gamma) \geq 2|\gamma| + \sum_{t=2}^r (r-t)m_t(\gamma). \quad (71)$$

We have by use of inequalities (69) and (71)

$$\sum_{t=2}^{r-1} (r-t)m_t(\gamma) \geq \left(r - \frac{2+r}{3-l(1-\kappa)}\right) |\gamma|. \quad (72)$$

Finally, by bounding the product over $t = 2, \dots, r-1$ in the activity bound (167) of Appendix B, we obtain (65). ■

We say that a polymer is small if $|\gamma| < \lambda n$. We define the partition function (with activities computed at the fixed point $(\underline{\eta}, \hat{\eta})$) of a gas of small polymers

$$\begin{aligned} Z_{\text{small}} &= \sum_{M \geq 0} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M \text{ s.t. } |\gamma_k| < \lambda n} \prod_{k=1}^M K(\gamma_k) \\ &\quad \times \prod_{k < k'} \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset). \end{aligned} \quad (73)$$

The free energy of the gas of small polymers $n^{-1} \ln Z_{\text{small}}$ has a polymer expansion (41) with the second sum replaced by a sum over $\gamma_1, \dots, \gamma_M$ s.t. $|\gamma_i| < \lambda n$.

Lemma 12. *Suppose $r > 2$. Fix $z_0 > 1$ and replace $K(\gamma)$ by $zK(\gamma)$, $z \in \mathbb{C}$, $|z| < z_0$, in the polymer expansion of $n^{-1} \ln Z_{\text{small}}$ which now becomes a power series in the parameter z^M , $M \geq 1$. Assume that Γ is a (λ, κ) expander with $\kappa \in]1 - \frac{2(r-1)}{lr}, 1 - \frac{1}{l}[$. One can find $\theta_0 > 0$ such that for $|\theta| < \theta_0$:*

- 1) *This power series is absolutely uniformly convergent in n and θ .*
- 2) *The following bound holds*

$$\left| \frac{1}{n} \ln Z_{\text{small}} \right| \leq \frac{2}{n} \sum_{x \in V \cup C} \sum_{\gamma \ni x, |\gamma| < \lambda n} \theta^{\frac{c}{2}|\gamma|} e^{|\gamma|}. \quad (74)$$

Proof: When Γ is an expander we can use the bound (65) on the activities of the small polymers. The proof is then almost identical to that of lemma 8. ■

Lemma 12 has the following consequence (we now take $z = 1$):

Corollary 13. *Suppose $r > 2$. Let E be the event that Γ is (λ, κ) expander. For $|\theta| < \theta_0$,*

$$\mathbb{E}_{\Gamma} \left[\frac{1}{n} \ln Z_{\text{small}} \middle| E \right] = O(n^{-(1-\chi)}) \quad (75)$$

for any $0 < \chi < 1$.

Remark 14. We stress that corollary 13 and lemma 12 hold for any (l, r) with $r > 2$. The restriction to odd l will come only when we estimate the contribution of large polymers.

Proof: Taking the conditional expectation over expander graphs (74) implies

$$\frac{1}{n} \mathbb{E}_\Gamma \left[\left| \ln Z_{\text{small}} \right| \middle| E \right] \leq \mathbb{E}_\Gamma \left[\sum_{\gamma \ni o, |\gamma| < \lambda n} \theta^{\frac{\epsilon}{2} |\gamma|} e^{|\gamma|} \middle| E \right]. \quad (76)$$

We can compute this expectation by conditioning on the first event that Γ is tree-like in a neighborhood of size $O(\ln n)$ around this vertex, and on the second complementary event. The second event has small probability $O(n^{-(1-\chi)})$ for any $0 < \chi < 1$. Besides, from (74) it is easy to show that $n^{-1} |\ln Z_{\text{small}}|$ is bounded uniformly in n . Thus the second event contributes only $O(n^{-(1-\chi)})$ to the expectation. For the first event we have that the smallest polymer is a cycle with $|\gamma| = O(\ln n)$. This implies that this event contributes $O((\theta^{\frac{\epsilon}{2}} e^{A+1})^{\ln n})$ to the expectation. For small θ it is $O(n^{-(1-\chi)})$ that dominates. ■

B. Probability estimates on graphs

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

$$1 + \sum_{g \subset \Gamma} K(g) = Z_{\text{small}} + R_{\text{large}}, \quad (77)$$

where

$$R_{\text{large}} = \sum_{g \subset \Gamma \text{ s.t. } \exists \gamma \subset g \text{ with } |\gamma| \geq \lambda n} K(g), \quad (78)$$

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

Lemma 15. Fix $\delta > 0$. Assume $l \geq 3$ odd and $l < r$. There exists a constant $C > 0$ depending only on l and r such that for θ small enough

$$\mathbb{P}_\Gamma \left[|R_{\text{large}}| \geq \delta \right] \leq \frac{1}{\delta} e^{-Cn}. \quad (79)$$

The proof which relies on counting estimates for subgraphs is presented in the appendix D. Unfortunately it breaks down for l even.

C. Proof of Theorem 6

The results of sections VII-A and VII-B allow to first prove

Proposition 16. Suppose l is odd and $3 \leq l < r$. Take Γ at random in $\mathcal{B}(l, r, n)$. There exist a small θ_0 independent of n such that for $\theta < \theta_0$, and any high-noise-solution $(\hat{\eta}, \hat{\eta})$ of the BP equations, with probability $1 - O(n^{-l(1-\kappa)-1})$ we have

$$\left| \frac{1}{n} \ln Z_{\text{LDPC}} - (f_{\text{LDPC}}^{\text{Bethe}}(\hat{\eta}, \hat{\eta}) + \frac{1}{n} \ln Z_{\text{small}}) \right| = O(e^{-n^{l(1-\kappa)-1}}). \quad (80)$$

Remark 17. We recall that $0 < l(1-\kappa) - 1 < (r-2)/r$. This proposition shows that large polymers contribute only with exponentially small corrections to the Bethe free energy. Inverse power in n corrections can be computed systematically from the polymer expansion of $n^{-1} \ln Z_{\text{small}}$.

Proof: Note that

$$\frac{1}{n} \ln \left\{ \sum_{g \subset \Gamma} K(g) \right\} = \frac{1}{n} \ln Z_{\text{small}} + \frac{1}{n} \ln \left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}} \right), \quad (81)$$

which means that the term on the left hand side of (80) is equal to

$$\frac{1}{n} \left| \ln \left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}} \right) \right|. \quad (82)$$

On one hand, from corollary 13 and the Markov bound, we have for any $\epsilon > 0$,

$$\mathbb{P} \left[e^{-n\epsilon} \leq Z_{\text{small}}^{-1} \leq e^{n\epsilon} \middle| E \right] = 1 - \frac{1}{\epsilon} O(n^{-(1-\chi)}). \quad (83)$$

On the other hand, from lemma 15

$$\mathbb{P} \left[|R_{\text{large}}| \geq \delta \middle| E \right] \mathbb{P}[E] \leq \mathbb{P} \left[|R_{\text{large}}| \geq \delta \right] \leq \frac{1}{\delta} e^{-Cn} \quad (84)$$

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

$$P \left[|R_{\text{large}}| \geq \delta \middle| E \right] \leq \frac{2}{\delta} e^{-Cn}. \quad (85)$$

Using (83) and (85), and choosing $\frac{\delta}{2} = e^{-2n\epsilon}$ it is not difficult to show that (at this point one must take $0 < 2\epsilon < C$)

$$\mathbb{P} \left[\left| \frac{R_{\text{large}}}{Z_{\text{small}}} \right| \geq e^{-n\epsilon} \middle| E \right] \leq \frac{1}{\epsilon} O(n^{-(1-\chi)}) + e^{-n(C-2\epsilon)}. \quad (86)$$

Thus

$$\mathbb{P} \left[\left| \frac{R_{\text{large}}}{Z_{\text{small}}} \right| \leq e^{-n\epsilon} \middle| E \right] \geq 1 - \frac{1}{\epsilon} O(n^{-(1-\chi)}). \quad (87)$$

This implies for n large

$$\mathbb{P} \left[\left| \ln \left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}} \right) \right| \leq 2e^{-n\epsilon} \middle| E \right] \geq 1 - \frac{1}{\epsilon} O(n^{-(1-\chi)}). \quad (88)$$

Furthermore

$$\begin{aligned} \mathbb{P} \left[\left| \ln \left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}} \right) \right| \leq 2e^{-n\epsilon} \right] \\ \geq \left(1 - \frac{1}{\epsilon} O(n^{-(1-\chi)}) \right) \left(1 - O(n^{-l(1-\kappa)-1}) \right) \\ \geq 1 - O(n^{-l(1-\kappa)-1}). \end{aligned} \quad (89)$$

The last line is obtained by choosing

$$\epsilon = \frac{n^{l(1-\kappa)-1}}{n^{1-\chi}} \leq n^{\frac{r-2}{r}-1+\chi} < \frac{C}{2}, \quad (90)$$

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

Finally from

$$n\epsilon = n^{l(1-\kappa)-1+\chi} \geq n^{l(1-\kappa)-1} \quad (91)$$

and (89) we deduce the statement of the proposition. ■

It is now possible to complete the proof of theorem 6.

Proof: [Proof of theorem 6] Consider the difference

$$\left| \frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}}(\underline{\eta}, \widehat{\underline{\eta}}) \right| \quad (92)$$

We first remark that this quantity is bounded uniformly in n because each term $n^{-1} |\ln Z_{\text{LDPC}}|$ and $|f_{\text{LDPC}}^{\text{Bethe}}|$ is bounded, as can be checked directly from their definition.

Now consider the event S - or the set of graphs - such that

$$\left| \ln \left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}} \right) \right| \leq e^{-n^{l(1-\kappa)-1}}. \quad (93)$$

Proposition (16) says that

$$\mathbb{P}[S^c] = O(n^{-(l(1-\kappa)-1)}). \quad (94)$$

Thus we have

$$\mathbb{E} \left[\left| \frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}} \right| \middle| S^c \right] \mathbb{P}[S^c] = O(n^{-(l(1-\kappa)-1)}). \quad (95)$$

We will now estimate

$$\mathbb{E} \left[\left| \frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}} \right| \middle| S \right] \mathbb{P}[S]. \quad (96)$$

Since $\mathbb{P}[S] = 1 - O(n^{-(l(1-\kappa)-1)})$ we have to show that the expectation conditioned over S is small.

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}} \right| \middle| S \right] \\ &= \mathbb{E} \left[\left| \frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}} \right| \middle| S \cap E \right] \mathbb{P}[E] \\ &+ \mathbb{E} \left[\left| \frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}} \right| \middle| S \cap E^c \right] \mathbb{P}[E^c]. \end{aligned} \quad (97)$$

Since, as remarked before, (92) is bounded and $\mathbb{P}[E^c] = O(n^{-(l(1-\kappa)-1)})$ the second term on the right hand side is $O(n^{-(l(1-\kappa)-1)})$. It remains to show that the conditional expectation in the first term of the right hand side is small. This is bounded above by two contributions. The first one is

$$\mathbb{E} \left[\left| \frac{1}{n} \ln \left(1 + \frac{R_{\text{large}}}{Z_{\text{small}}} \right) \right| \middle| S \cap E \right] \leq e^{-n^{l(1-\kappa)-1}}, \quad (98)$$

and the second (recall corollary 13)

$$\mathbb{E} \left[\left| \frac{1}{n} \ln Z_{\text{small}} \right| \middle| S \cap E \right] = O(n^{-(1-\chi)}). \quad (99)$$

Putting all contributions (95), (97), (98), (99) together we obtain the desired result

$$\begin{aligned} \mathbb{E} \left[\left| \frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}} \right| \right] &= O(n^{-(1-\chi)}) + O(n^{-(l(1-\kappa)-1)}) \\ &= O(n^{-(l(1-\kappa)-1)}). \end{aligned} \quad (100)$$

In the last step we have taken $0 < \chi < l(1-\kappa) - 1$. ■

VIII. DISCUSSION

A. LDPC: The case l even

When l is even the point $\theta = 0$ has a singular behavior. As the channel realization is trivial $\underline{h} = \underline{0}$, the infinitely high noise fixed point is simply the all zeros messages $(\eta_{i \rightarrow a}, \widehat{\eta}_{a \rightarrow i}) = (0, 0)$. The activities can be computed exactly for this BP fixed point

$$K_a(\gamma) = \begin{cases} 1 & \text{if } |\partial a \cap \gamma| = r \\ 0 & \text{otherwise} \end{cases}, \quad K_i(\gamma) = \frac{1 + (-1)^{|\partial i \cap \gamma|}}{2}. \quad (101)$$

When the graph Γ is an expander, every small polymer $|\gamma| < n\lambda$ contains at least one check node with induced degree less than r (see Lemma 10). Thus $K(\gamma) = 0$ and

$$Z_{\text{small}} = 1. \quad (102)$$

The contribution of small polymers to the free energy vanishes which is of course in adequacy with the prediction of the polymer expansion (see Lemma 12).

However for the total graph, and unlike the case l odd, we have (for l even we would have $K(\Gamma) = 0$)

$$K(\Gamma) = 1. \quad (103)$$

More generally polymers with a size of the order of the total graph have an activity close to one. This implies that the contribution coming from large polymers is non-vanishing but is growing polynomially

$$1 < \mathbb{E}_{\Gamma}(|R_{\text{large}}|) < C_{l,r} n^{4r^2}, \quad (104)$$

as can be shown using the same counting arguments as in the Appendix D. As a consequence, we find similarly to Theorem 16 that the Bethe free energy is asymptotically exact with high probability. More precisely, with probability $1 - O(n^{-(l(1-\kappa)-1)})$

$$\left| \frac{1}{n} \ln Z_{\text{LDPC}} - f_{\text{LDPC}}^{\text{Bethe}} \right| = O\left(\frac{1}{n} \ln n\right). \quad (105)$$

The notable difference with Theorem 16 is that the decay rate of the difference is not exponential.

When l is even and $\theta > 0$, the bound on the activity of the total graph predicts an exponential growth for $\overline{K}(\Gamma)$,

$$\overline{K}(\Gamma) = (1 + \alpha\theta^r)^{\frac{l}{r}n} (1 + \beta\theta^2)^n. \quad (106)$$

The contribution of the large polymer can no longer be estimated as in Appendix D. To tackle this problem, it seems necessary to have a precise control of sign cancellations in the sum R_{large} . Such a control is out of reach of the methods used in this paper.

B. LDPC: $l > r$

The constraint $r > l$ appears naturally for proving Theorem 15. If $r < l$, large polymer with an activity exponentially increasing with their size are also exponentially numerous. Therefore we found using counting arguments that the contribution of R_{large} is not negligible.

When $r < l$, the graphical model is no longer describing a code. In fact for at exactly $h = 0$, the partition function

counts the number of solution of a random linear system of equations which is overdetermined. This corresponds to an UNSAT phase of a linear constraints satisfaction problem. In this phase, it is expected that the Bethe free energy is not a good approximation of the free energy (instead one should use the RSB free energy [14]). It seems then reasonable to think that the corrections to the Bethe free energy are non vanishing even with a precise control of the sign cancellation in the activities.

C. LDPC low noise

The low noise regime is characterized by half-log likelihoods with high magnitudes $h \approx \infty$. The low noise fixed point of the Belief-Propagation equations is the trivial solution

$$(\eta_{i \rightarrow a}, \hat{\eta}_{a \rightarrow i}) = (+\infty, +\infty). \quad (107)$$

The activities can be computed exactly at the low noise fixed point

$$K(\gamma) = \prod_{a \in \gamma \cap C} \frac{1 + (-1)^{|\partial a \cap \gamma|}}{2} \times \prod_{i \in \gamma \cap V} (-1)^l e^{-2h_i \mathbb{I}(|\partial i \cap \gamma| = l)}. \quad (108)$$

According to (108), polymers are subgraphs which have check nodes with even induced degree and variable nodes with induced degree equal to l . The particularity of the low noise activities is that their intensity depends on the sign of the half log-likelihoods h_i , and a fortiori on the distribution of \underline{h} . Using Hoeffding's inequality, we see that a polymer has, with large probability, a small activity

$$\mathbb{P}_{\underline{h}} \left(|K(\gamma)| \leq e^{-2h(\tanh h - 2\epsilon)|\gamma \cap V|} \right) \geq 1 - e^{-2\epsilon^2 |\gamma \cap V|}. \quad (109)$$

However, the expected activity is dominated by rare events

$$\mathbb{E}_{\underline{h}} [|K(\gamma)|] = 1. \quad (110)$$

This prevents to use the method if this paper.

D. Lattice graphs with low dimension

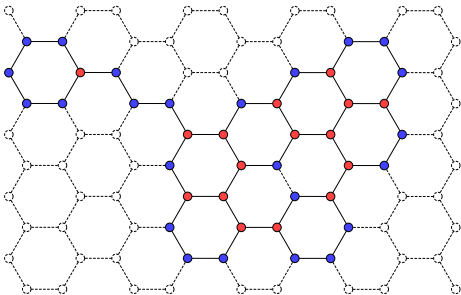


Figure 4. Polymer in a honeycomb lattice at low temperature. The activity is decreasing in the boundary size (blue nodes).

Another possible application of the polymer expansion is the study of spin systems on a regular lattice $\Lambda = (V, E)$. In

low dimension, it is known that the mean field approximations such as the Bethe free energy is not exact and thus must have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln Z_{\text{polymer}} \neq 0. \quad (111)$$

However, if the polymer expansion converges, it can be used as a systematic way of computing corrections to the mean field approximation.

Let us illustrate the question of the convergence for the Ising model on the honeycomb lattice (see Figure 4). The spins are attached to vertices and an edge represents a ferromagnetic interaction between spins

$$Z_{\text{Ising}} = \sum_{\sigma \in \{-1, 1\}^n} \prod_{(i, j) \in E} \exp(\beta J \sigma_i \sigma_j). \quad (112)$$

There are three solutions of the BP equations and they can be computed exactly. There is one fixed point which describes a high temperature phase ($\beta J \approx 0$) and two fixed points describing a low temperature phase ($\beta J \approx \infty$). The two low temperature fixed points differ only by a sign.

The high temperature activity is

$$K_{\text{Ising}}^0(\gamma) = \begin{cases} (\tanh \beta J)^{|\gamma|} & \text{if } \gamma \text{ is a cycle} \\ 0 & \text{otherwise} \end{cases}. \quad (113)$$

Thus the activity is decreasing in the polymer size. To prove the convergence of the expansion for high temperatures one can apply directly Lemma 8. Or one can see that, in the honeycomb lattice, the number of cycles attach to a vertices with length t is upper-bounded by 2^t .

For low temperatures the activity is

$$K_{\text{Ising}}^{\pm \infty}(\gamma) = \prod_{i \in V, |\partial i \cap \gamma| = 3} \frac{1 + \tanh \beta J}{2 \tanh \beta J} \sqrt{\frac{2 \tanh \beta J - 1}{\tanh \beta J}} \times \prod_{i \in V, |\partial i \cap \gamma| = 2} \mp \frac{1 - \tanh \beta J}{2 \tanh \beta J}. \quad (114)$$

The polymers have an activity which is decreasing with respect to the size of the boundary (the nodes with induced degree equal to two). This is similar to the activity of polymers for LDPC (at high noise). But unlike the LDPC case, we cannot apply an expander argument because that the boundary of a polymer is not of the same order than its size. Here the mechanism of convergence for the polymer expansion is different. Notice that on the lattice, a polymer is almost completely determined by the position of the boundary nodes. Once the induced degree two nodes are placed on the lattice, there are almost no other choice for placing the degree three nodes. They must "fill" the space between the degree two nodes (see Figure 4). Thus the entropy of polymers is increasing in the boundary size and can be controlled by the activity of polymer for a small enough temperature.

ACKNOWLEDGEMENT

The work of M.V. was supported by Swiss National Science Foundation grant No. 200020-140388. We acknowledge useful discussions with M. Chertkov, on the loop expansions.

APPENDIX A
DERIVATION OF THE LOOP SUM IDENTITY

The ‘‘loop sum identity’’ is a representation of the error term between the free energy and the Bethe free energy. It takes the form of the logarithm of a sum over sub-graphs non-necessarily connected. This identity was first derived for graphical models with binary variables by Chertkov and Chernyak in [5] and later generalized for variables on a q -ary alphabet by the same authors in [15]. This appendix contains a short derivation of the loop sum identity based on the original paper [5].

Consider the problem of computing the partition function of a factor graph model

$$Z = \sum_{\underline{s} \in \{-1, +1\}^n} \prod_{a \in C} \psi_a(s_{\partial a}). \quad (115)$$

The loop expansion takes a natural form on graphical models called vertex models, where variables are attached to edges. We introduce the auxiliary set of spins $\sigma_{ia}, \hat{\sigma}_{ai} \in \{-1, 1\}$ attached to directed edges ($i \rightarrow a$) and ($a \rightarrow i$) respectively. We denote by $\sigma_{\partial a} = \{\sigma_{ja} | j \in \partial a\}$ the collection of spins that are on edges pointing toward a and $\hat{\sigma}_{\partial i} = \{\hat{\sigma}_{bi} | b \in \partial i\}$ correspondingly. We can rewrite (115) as a partition function of a vertex model

$$Z = \sum_{\underline{\sigma}, \underline{\hat{\sigma}} \in \{-1, 1\}^{|\mathcal{E}|}} \prod_{a \in C} \psi_a(\sigma_{\partial a}) \prod_{i \in V} \phi_i(\hat{\sigma}_{\partial i}) \prod_{(a,i) \in E} \frac{1 + \sigma_{ia} \hat{\sigma}_{ai}}{2}, \quad (116)$$

where

$$\phi_i(\hat{\sigma}_{\partial i}) = \prod_{b, c \in \partial i} \frac{1 + \hat{\sigma}_{bi} \hat{\sigma}_{ci}}{2}. \quad (117)$$

Let us comment on the expression (116). The weights $\phi_i(\hat{\sigma}_{\partial i})$ ensure that all spins on edges outgoing from a variable node i take the same value s_i . As for the last product, it forces spins on the same edge to be equal. The key idea in the loop expansion is to ‘‘soften’’ the constraints on the edges before performing the expansion. Using the following identity, valid for any binary distributions ν_{ia} and $\hat{\nu}_{ai}$,

$$\frac{1 + \sigma_{ia} \hat{\sigma}_{ai}}{2} = \frac{\nu_{ia}(\sigma_{ia}) \hat{\nu}_{ai}(\hat{\sigma}_{ai}) + \sigma_{ia} \hat{\nu}_{ai}(-\sigma_{ia}) \hat{\sigma}_{ai} \nu_{ia}(-\hat{\sigma}_{ai})}{\sum_{s \in \{-1, 1\}} \nu_{ia}(s) \hat{\nu}_{ai}(s)}, \quad (118)$$

we can rewrite the partition function (116)

$$\begin{aligned} Z = & \sum_{\underline{\sigma}, \underline{\hat{\sigma}} \in \{-1, 1\}^{|\mathcal{E}|}} \prod_{a \in C} \psi_a(\sigma_{\partial a}) \prod_{j \in \partial a} \nu_{ja}(\sigma_{ja}) \\ & \prod_{i \in V} \phi_i(\hat{\sigma}_{\partial i}) \prod_{b \in \partial i} \hat{\nu}_{bi}(\hat{\sigma}_{bi}) \\ & \prod_{(a,i) \in E} \left(\sum_{s \in \{-1, 1\}} \nu_{ia}(s) \hat{\nu}_{ai}(s) \right)^{-1} \\ & \prod_{(a,i) \in E} \left(1 + \sigma_{ia} \frac{\hat{\nu}_{ai}(-\sigma_{ia})}{\nu_{ia}(\sigma_{ia})} \hat{\sigma}_{ai} \frac{\nu_{ia}(-\hat{\sigma}_{ai})}{\hat{\nu}_{ai}(\hat{\sigma}_{ai})} \right). \end{aligned} \quad (119)$$

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

$$\prod_{e \in E} (1 + f(e)) = 1 + \sum_{g \subset \Gamma} \prod_{e \in E \cap g} f(e), \quad (120)$$

where the sum runs on every non-empty subset of edges represented by subgraphs g . Expanding the last product in (119) with the generalized binomial formula leads to

$$Z = \exp(n f^{\text{Bethe}}) \times \left(1 + \sum_{g \subset \Gamma} K(g) \right). \quad (121)$$

The quantity that factorized in the expansion appears to be the Bethe free energy

$$f^{\text{Bethe}}(\underline{\nu}, \underline{\hat{\nu}}) = \frac{1}{n} \left(\sum_{a \in C} F_a + \sum_{i \in V} F_i + \sum_{(i,a) \in E} F_{ia} \right), \quad (122)$$

where

$$\begin{cases} F_a = \ln \{ \sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{j \in \partial a} \nu_{ja}(s_j) \}, \\ F_i = \ln \{ \sum_{s_i} \prod_{b \in \partial i} \hat{\nu}_{bi}(s_i) \}, \\ F_{ia} = \ln \{ \sum_{s_i} \nu_{ia}(s_i) \hat{\nu}_{ai}(s_i) \}. \end{cases} \quad (123)$$

The activities $K(g)$ associated to each subgraphs can be distributed in contributions coming from vertices in g

$$K(g) = \prod_{i \in g \cap V} K_i \prod_{a \in g \cap C} K_a, \quad (124)$$

where

$$K_i(g) = \frac{\sum_{s_i} \prod_{a \in \partial i \setminus g} \hat{\nu}_{ai}(s_i) \prod_{a \in \partial i \cap g} s_i \nu_{ia}(-s_i)}{\sum_{s_i} \prod_{a \in \partial i} \hat{\nu}_{ai}(s_i)}, \quad (125)$$

and

$$K_a(g) = \frac{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{i \in \partial a \setminus g} \nu_{ia}(s_i) \prod_{i \in \partial a \cap g} s_i \hat{\nu}_{ai}(-s_i)}{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{i \in \partial a} \nu_{ia}(s_i)}. \quad (126)$$

The sum over subgraphs in (121) is the ‘‘loop sum identity’’. Note that for the moment the binary distributions entering in (118) are completely arbitrary. The transformation (118) is crucial in that it allows to keep the correlations between neighboring spins. Indeed, if we expand the Kronecker delta in (116) directly, we would be left with an expansion accurate only in a high temperature regime. Instead, the effect of correlations are present but hidden in the choice of the distributions ν and $\hat{\nu}$.

In order for the loop sum identity to be useful one has to choose the ‘‘right’’ binary distributions. The natural requirement for sparse locally tree-like graphs is that every subgraphs g with a dangling edge must have a zero weight. Said differently, the distributions ν and $\hat{\nu}$ must be chosen such that $|\partial a \cap g| = 1$ and $|\partial i \cap g| = 1$ implies $K_a(g) = 0$

and $K_i(g) = 0$ respectively. The requirement $K_a(g) = 0$ is fulfilled by distributions $\widehat{\nu}_{ai}$ that satisfy the following equation

$$\sum_{s_i} s_i \widehat{\nu}_{ai}(-s_i) \sum_{s_{\partial a \setminus i}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j) = 0. \quad (127)$$

This is satisfied if $\widehat{\nu}_{ai}$ is a solution of the first Belief-Propagation equation

$$\widehat{\nu}_{ai}(s_i) = \frac{\sum_{s_{\partial a \setminus i}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}. \quad (128)$$

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

$$\nu_{ia}(s_i) = \frac{\prod_{b \in \partial i \setminus a} \widehat{\nu}_{bi}(s_i)}{\sum_{s_i} \prod_{b \in \partial i \setminus a} \widehat{\nu}_{bi}(s_i)}. \quad (129)$$

This is nothing else but the second belief propagation equation.

APPENDIX B ACTIVITIES OF LOOPS AND BOUNDS

A. High temperature general models

We recall that the partition function of a general factor graph model is given by (4) and the weights by (8). We take β small enough so that

$$|\psi_a(s_{\partial a}) - 1| \leq 2\beta \sup_{I \subset \partial a} \sum_{I \subset \partial a} |J_I| \equiv \mu, \quad (130)$$

As will be seen later on we will need β small enough so that $\mu < O(1/l_{\max}^2 r_{\max})$.

In order to find bounds on the activities (125) and (126), we should control the behavior of the Belief-Propagation messages. This is realized through the BP equations (128) and (129). We first choose to parametrize the BP distributions $\nu, \widehat{\nu}$ with real numbers $\zeta, \widehat{\zeta}$

$$\widehat{\nu}_{ai}(s_i) = \frac{1 + s_i \tanh \widehat{\zeta}_{a \rightarrow i}}{2} \quad \text{and} \quad \nu_{ia}(s_i) = \frac{1 + s_i \tanh \zeta_{i \rightarrow a}}{2}. \quad (131)$$

The BP equation (128) now reads

$$\tanh \widehat{\zeta}_{a \rightarrow i} = \frac{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) s_i \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}{\sum_{s_{\partial a}} \psi_a(s_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}. \quad (132)$$

Injecting the high temperature condition (130) leads to the following bound

$$\begin{aligned} |\tanh \widehat{\zeta}_{a \rightarrow i}| &\leq \frac{\sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)}{1 - \sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{j \in \partial a \setminus i} \nu_{ja}(s_j)} \\ &\leq \frac{\mu}{1 - \mu} \\ &\leq 2\mu, \end{aligned} \quad (133)$$

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

The other BP equation (129) takes the form

$$\tanh \zeta_{i \rightarrow a} = \tanh \left(\sum_{b \in \partial i \setminus a} \widehat{\zeta}_{b \rightarrow i} \right). \quad (134)$$

Using the bound (133) on messages $\widehat{\zeta}_{a \rightarrow i}$ gives

$$\begin{aligned} |\tanh \zeta_{i \rightarrow a}| &\leq \tanh \left((l_{\max} - 1) \tanh^{-1}(2\mu) \right) \\ &\leq 2(l_{\max} - 1)\mu. \end{aligned} \quad (135)$$

The inequalities (133) and (135) can be restated in terms of distributions $\widehat{\nu}, \nu$ and take the form

$$\begin{cases} \frac{1-2\mu}{2} \leq \widehat{\nu}_{ai}(s_i) \leq \frac{1+2\mu}{2} \\ \frac{1-2(l_{\max}-1)\mu}{2} \leq \nu_{ia}(s_i) \leq \frac{1+2(l_{\max}-1)\mu}{2}. \end{cases} \quad (136)$$

By noticing that $\sum_s s \widehat{\nu}_{ai}(s) = \tanh \widehat{\zeta}_{a \rightarrow i}$ and using the bound (133), we are in position to control the activity (126)

$$\begin{aligned} |K_a(g)| &\leq \frac{\sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{i \in \partial a \setminus g} \nu_{ia}(s_i) \prod_{i \in \partial a \cap g} \widehat{\nu}_{ai}(-s_i)}{1 - \sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{i \in \partial a} \nu_{ia}(s_i)} \\ &\quad + \frac{\prod_{i \in \partial a \cap g} |\sum_{s_i} s_i \widehat{\nu}_{ai}(s_i)|}{1 - \sum_{s_{\partial a}} |\psi_a(s_{\partial a}) - 1| \prod_{i \in \partial a} \nu_{ia}(s_i)} \\ &\leq \frac{\mu + (2\mu)^{|\partial a \cap g|}}{1 - \mu} \\ &\leq 6\mu, \end{aligned} \quad (137)$$

where in the last line we use the fact that subgraphs g have no dangling edges (i.e. $|\partial a \cap g| \geq 2$) and $\mu \leq 1/2$.

The second activity (125) is directly controlled using bounds on distributions $\widehat{\nu}$ and ν given by equations (136)

$$\begin{aligned} |K_i(g)| &\leq \frac{\sum_{s_i} \prod_{a \in \partial i \setminus g} |\widehat{\nu}_{ai}(s_i)| \prod_{a \in \partial i \cap g} |\nu_{ia}(-s_i)|}{|\sum_{s_i} \prod_{a \in \partial i} \widehat{\nu}_{ai}(s_i)|} \\ &\leq \left(\frac{1 + (l_{\max} - 1)2\mu}{1 - 2\mu} \right)^{l_{\max}} \\ &\leq (1 + 4l_{\max}\mu)^{l_{\max}}. \end{aligned} \quad (138)$$

The total activity of a generalized loop g , given by the relation (124), is then bounded by

$$\begin{aligned} |K(g)| &\leq \prod_{i \in g \cap V} |K_i| \prod_{a \in g \cap C} |K_a| \\ &\leq \exp \left(|g \cap V| l_{\max} \ln(1 + 4l_{\max}\mu) + |g \cap C| \ln(6\mu) \right). \end{aligned} \quad (139)$$

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

$$|g| := |g \cap C| + |g \cap V|. \quad (140)$$

In order to show that the activities in (139) are exponentially decreasing in the loop size, we should show that the number of variable nodes contained in a loop cannot be arbitrarily larger than the number of check nodes. Consider the number of edges contained in a subgraph. We can bound from above this number counted from the check node perspective and we can find a lower bound counted from the variable node perspective.

This leads to the following bound on the number of variable nodes

$$r_{\max}|g \cap C| \geq 2|g \cap V|. \quad (141)$$

Using the definition (140) and the bound (141), we find that for every non-negative numbers $p \geq 0$ and $q \geq 0$

$$\begin{aligned} |g \cap V|p - |g \cap C|q &= -(p+q)|g \cap C| + p|g| \\ &\leq |g| \frac{r_{\max}p - 2q}{2 + r_{\max}} \end{aligned} \quad (142)$$

This implies the upper bound for the exponent in (139)

$$\frac{|g|}{2 + r_{\max}} \ln \left((6e\mu)^2 (1 + 4l_{\max}\mu)^{r_{\max}l_{\max}} \right). \quad (143)$$

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

$$(6\mu)^2 (1 + 4r_{\max}\mu)^{r_{\max}l_{\max}} \leq (6e\mu)^2. \quad (144)$$

From (139), (143) and (144) we deduce the bound on the activities

$$|K(g)| \leq (6e\mu)^{2|g|/(2+r_{\max})}. \quad (145)$$

B. LDGM codes

We recall that the partition function of a LDGM code is

$$Z_{\text{LDGM}} = \sum_{\underline{s} \in \{-1,1\}^n} \prod_{a \in C} e^{h_a \prod_{i \in \partial a} s_i}. \quad (146)$$

The LDGM codes can be seen as a special case of the high temperature general models with $I \rightarrow \partial a$ and $\beta J_I \rightarrow h_a$. The high temperature condition translates into $2 \sup_a |h_a| = \mu \ll 1$. Recalling that $|h_a| = h = \frac{1}{2} \ln \frac{1-p}{p}$, we see that the high temperature condition is equivalent to taking p close to $1/2$. The bound on the activity is obtained by applying (145)

$$|K(g)| \leq \left(6e \ln \frac{1-p}{p} \right)^{2|g|/(2+r_{\max})}. \quad (147)$$

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

There is a remarkable simplification for LDGM ensembles with no degree one check nodes. In this case the BP equations admit a trivial fixed point where $\nu_{ia}(s) = 1/2$, $\hat{\nu}_{ia}(s) = 1/2$. The activities at the trivial fixed point can be computed exactly

$$K_a^{\text{trivial}}(g) = \begin{cases} \tanh h_a & \text{if } \partial a \cap g = \partial a \\ 0 & \text{otherwise,} \end{cases} \quad (148)$$

and

$$K_i^{\text{trivial}}(g) = \begin{cases} 1 & \text{if } |\partial a \cap g| \text{ is even} \\ 0 & \text{otherwise.} \end{cases} \quad (149)$$

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

$$|K(g)^{\text{trivial}}| \leq (1-2p)^{|g \cap C|} \leq (1-2p)^{\frac{2|g|}{2+r_{\max}}} \quad (150)$$

C. Regular LDPC codes

The LDPC codes cannot be seen as high temperature models. Their partition function

$$Z_{\text{LDPC}} = \sum_{\underline{x} \in \{0,1\}^n} \prod_{a \in C} \mathbb{I}(\oplus_{i \in \partial a} x_i = 0) \prod_{i \in V} \exp((-1)^{x_i} h_i). \quad (151)$$

is composed of two type of weights. The variable node weights, coming from channel observations, satisfies a high temperature condition at high noise. But the check node weights, enforcing a parity check constraint, always admit a configuration of variable which cancels the weight. Thus it make impossible that the check node weights satisfies the high temperature condition (130).

We use the standard parametrization for the BP distributions $\nu, \hat{\nu}$ in term of the real numbers $\eta, \hat{\eta}$

$$\hat{\nu}_{ai}(s_i) = \frac{1 + s_i \tanh \hat{\eta}_{a \rightarrow i}}{2} \quad \text{and} \quad \nu_{ia}(s_i) = \frac{1 + s_i \tanh \eta_{i \rightarrow a}}{2}. \quad (152)$$

With this parametrization the Belief-Propagation equations (128), (129) for the messages reads

$$\begin{cases} \tanh(\hat{\eta}_{a \rightarrow i}) = \prod_{j \in \partial a \setminus i} \tanh \eta_{j \rightarrow a} \\ \eta_{i \rightarrow a} = h_i + \sum_{b \in \partial i \setminus a} \hat{\eta}_{b \rightarrow i}. \end{cases} \quad (153)$$

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

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

$$|\tanh \eta_{i \rightarrow a}| \leq \theta. \quad (154)$$

where

$$\theta = (1 + \epsilon) \tanh h. \quad (155)$$

The condition (154) can be justified by looking at the Taylor expansion of solution at high noise. For $h = 0$, the BP equations (153) admit the simple solution $\tanh \eta_{a \rightarrow i} = 0$, $\tanh \hat{\eta}_{i \rightarrow a} = 0$. If we compute the Taylor expansion of this solution with respect to the noise parameter, we find

$$\begin{cases} \tanh \hat{\eta}_{a \rightarrow i} = \prod_{j \in \partial a \setminus i} \tanh h_j \\ \tanh \eta_{i \rightarrow a} = \tanh h_i + \sum_{b \in \partial i \setminus a} \prod_{j \in \partial a \setminus i} \tanh h_j, \end{cases} \quad (156)$$

plus some term of order $\mathcal{O}((\tanh h)^r)$. This shows that there exists a $h_0(\epsilon, n)$ such that the high noise condition (154) is satisfied for $h < h_0(\epsilon, n)$. However it does not guaranteed that $h_0(\epsilon, n)$ is uniform in the size of the graph.

By using the high noise condition (154) along with the BP equations (153), we find the reciprocal bound on messages from check nodes to variable nodes

$$|\tanh \hat{\eta}_{a \rightarrow i}| \leq \theta^{r-1}. \quad (157)$$

We recall that the induced degree of check and variable node are denoted by $d_a(g) = |\partial a \cap g|$ and $d_i(g) = |\partial i \cap g|$ respectively. The number of check nodes and variable nodes with prescribed induced degree by $n_s(g) = |\{i \in g \cap V | d_i(g) = s\}|$ and $m_t(g) = |\{a \in g \cap C | d_a(g) = t\}|$. For LDPC codes, the activities associated to check nodes (126) are

$$K_a(g) = \frac{u_a + (-1)^{d_a(g)} v_a}{1 + u_a w_a}, \quad (158)$$

where

$$\begin{cases} u_a &= \prod_{i \in \partial a \setminus g} \tanh \eta_{i \rightarrow a} \\ v_a &= \prod_{i \in \partial a \cap g} \tanh \hat{\eta}_{a \rightarrow i} \\ w_a &= \prod_{i \in \partial a \cap g} \tanh \eta_{i \rightarrow a}. \end{cases} \quad (159)$$

Using inequalities (154), (157), it is straightforward to bound the check activities

$$\begin{aligned} |K_a(g)| &\leq \frac{|u_a| + |v_a|}{1 - |u_a||w_a|} \\ &\leq \frac{\theta^{r-d_a(g)} + \theta^{(r-1)d_a(g)}}{1 - \theta^r}. \end{aligned} \quad (160)$$

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

$$|K_a(g)| \leq \begin{cases} 1 + \alpha\theta^r & \text{if } d_a(g) = r \\ \alpha\theta^{r-d_a(g)} & \text{if } d_a(g) \neq r. \end{cases} \quad (161)$$

The activities associated to variable nodes (125) reads

$$K_i(g) = \frac{e^{(u_i - v_i)} + (-1)^{d_i(g)} e^{-(u_i - v_i)}}{e^{(u_i + w_i)} + e^{-(u_i + w_i)}} \prod_{a \in \partial i \cap g} \frac{\cosh \hat{\eta}_{a \rightarrow i}}{\cosh \eta_{i \rightarrow a}}, \quad (162)$$

where

$$\begin{cases} u_i &= h_i + \sum_{a \in \partial i \setminus g} \hat{\eta}_{a \rightarrow i} \\ v_i &= \sum_{a \in \partial i \cap g} \eta_{i \rightarrow a} \\ w_i &= \sum_{a \in \partial i \cap g} \hat{\eta}_{a \rightarrow i}. \end{cases} \quad (163)$$

Again by a direct application of inequalities (154) and (157), we find

$$\begin{aligned} |K_i(g)| &\leq \frac{e^{(d_i(g)+1) \tanh^{-1} \theta} + (-1)^{d_i(g)} e^{-(d_i(g)+1) \tanh^{-1} \theta}}{e^{(d_i(g)+1) \tanh^{-1} \theta} + e^{-(d_i(g)+1) \tanh^{-1} \theta}} \\ &\quad \times \left(\frac{1 + \theta^2}{1 - \theta^2} \right)^{d_i(g)/2}. \end{aligned} \quad (164)$$

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

$$|K_i(g)| \leq \begin{cases} 1 + \frac{\beta}{2}(1 + 4d_i(g) + d_i(g)^2)\theta^2 & \text{if } d_i(g) \text{ even} \\ \beta(1 + d_i(g))\theta & \text{if } d_i(g) \text{ odd.} \end{cases} \quad (165)$$

Using the formulas (124) we derive the following estimate of subgraphs activities for $\theta < \theta_0$ small enough

$$|K(g)| \leq \bar{K}(\underline{n}(g), \underline{m}(g)), \quad (166)$$

where

$$\begin{aligned} \bar{K}(\underline{n}(g), \underline{m}(g)) &= (1 + \alpha\theta^r)^{m_r(g)} \prod_{t=2}^{r-1} (\alpha\theta^{r-t})^{m_t(g)} \\ &\quad \times \prod_{\substack{s=2, \\ \text{even}}}^{l-1} \left(1 + \frac{\beta}{2}(1 + 4s + s^2)\theta^2 \right)^{n_s(g)} \\ &\quad \times \prod_{\substack{s=3, \\ \text{odd}}}^l (\beta(1 + s)\theta)^{n_s(g)}. \end{aligned} \quad (167)$$

Estimate (167) is essentially optimal for small θ as can be checked by Taylor expanding $K(g)$ in powers of θ .

APPENDIX C POLYMER EXPANSION IDENTITY (41)

We recall that \mathcal{G}_M is the set of all connected graphs G with M labeled vertices $1, \dots, M$. In general, for $\mathcal{I} \subset \{1, \dots, M\}$, $\mathcal{G}_{\mathcal{I}}$ is the set of all connected graphs with labeled vertices in \mathcal{I} . Furthermore we denote the complete graph with M labeled vertices by \mathcal{K}_M . A partition of the set $\{1, \dots, M\}$ into q cells is an unordered list $\{\mathcal{I}_1, \dots, \mathcal{I}_q\}$ of disjoint nonempty subsets $\mathcal{I}_t \subset \{1, \dots, M\}$. The partitions of M elements into q cells form an ensemble denoted by \mathcal{P}_M^q .

The polymer partition function is

$$\begin{aligned} Z^{\text{polymer}} &= 1 + \sum_{M \geq 1} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M \subset \Gamma} \prod_{k=1}^M K(\gamma_k) \\ &\quad \times \prod_{k < k'} \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset). \end{aligned} \quad (168)$$

We recall that polymers γ are connected subgraphs of Γ that cannot intersect due to the presence of the hard core constraints $\mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset)$. The polymer expansion identity is based on the expansion of these hard core constraints using the binomial theorems on graphs (Eqn. (120))

$$\begin{aligned} \prod_{k < k'} \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset) &= \prod_{k < k'} (\mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset) - 1 + 1) \\ &= \sum_{G \subset \mathcal{K}_M} U(G), \end{aligned} \quad (169)$$

where to each $G \in \mathcal{K}_M$ we associate an Ursell function

$$U(G) = \prod_{(k, k') \in G} \left\{ \mathbb{I}(\gamma_k \cap \gamma_{k'} = \emptyset) - 1 \right\}, \quad (170)$$

and we take the convention that $U(G) = 1$ if G has no edges. The sum in (169) runs over subgraphs G that are not necessarily connected. But for every subgraphs G there exist a unique decomposition into disjoint connected subgraphs G_1, \dots, G_q . This allows us to re-sum (169) as

$$\sum_{G \subset \mathcal{K}_M} U(G) = \sum_{q=1}^M \sum_{\{\mathcal{I}_1, \dots, \mathcal{I}_q\} \in \mathcal{P}_M^q} \prod_{t=1}^q \sum_{G_t \subset \mathcal{G}_{\mathcal{I}_t}} U(G_t). \quad (171)$$

Together with (169) and (171), the polymer partition function (168) can be rewritten as

$$Z^{\text{polymer}} = 1 + \sum_{M \geq 1} \frac{1}{M!} \sum_{q=1}^M \sum_{\{\mathcal{I}_1, \dots, \mathcal{I}_q\} \in \mathcal{P}_M^q} \prod_{t=1}^q \phi(\mathcal{I}_t), \quad (172)$$

where

$$\phi(\mathcal{I}_t) := \sum_{\gamma_k \in \mathcal{I}_t} \prod_{k \in \mathcal{I}_t} K(\gamma_k) \sum_{G_t \subset \mathcal{G}_{\mathcal{I}_t}} U(G_t). \quad (173)$$

The function introduced in (173) depends only on the size of the ensemble

$$\phi(\mathcal{I}_t) = \phi(|\mathcal{I}_t|), \quad (174)$$

as $k \in \mathcal{I}_t$ in (173) are just dummy indices. The number of partitions of $\{1, \dots, M\}$ with prescribed size $|\mathcal{I}_1| = m_1, \dots, |\mathcal{I}_q| = m_q$ is

$$\sum_{\{|\mathcal{I}_1|=m_1, \dots, |\mathcal{I}_q|=m_q\} \in \mathcal{P}_M^q} 1 = \frac{M!}{q!} \prod_{t=1}^q \frac{1}{m_t!}, \quad (175)$$

where m_1, \dots, m_q are non-zero integers satisfying $m_1 + \dots + m_q = M$. These considerations allow us to rewrite (172) as

$$\begin{aligned} Z^{\text{polymer}} &= 1 + \sum_{M \geq 1} \sum_{q=1}^M \frac{1}{q!} \sum_{m_1 + \dots + m_q = M} \prod_{t=1}^q \frac{\phi(m_t)}{m_t!}. \\ &= 1 + \sum_{q=1}^M \frac{1}{q!} \sum_{M=q}^{\infty} \sum_{m_1 + \dots + m_q = M} \prod_{t=1}^q \frac{\phi(m_t)}{m_t!} \\ &= 1 + \sum_{q=1}^M \frac{1}{q!} \left(\sum_{M=1}^{\infty} \frac{\phi(M)}{M!} \right)^q \\ &= \exp \left(\sum_{M=1}^{\infty} \frac{\phi(M)}{M!} \right). \end{aligned} \quad (176)$$

The logarithm of the polymers partition function (168) can thus be expressed as

$$\ln Z^{\text{polymer}} = \sum_{M=1}^{\infty} \frac{1}{M!} \sum_{\gamma_1, \dots, \gamma_M \subset \Gamma} \prod_{k=1}^M K(\gamma_k) \sum_{G \subset \mathcal{G}_M} U(G). \quad (177)$$

APPENDIX D PROOF OF LEMMA 15

In this appendix we prove the Lemma 15 which is restated below for convenience

Lemma. Fix $\delta > 0$. Assume $l \geq 3$ odd and $l < r$. There exists a constant $C > 0$ depending only on l and r such that for θ small enough

$$\mathbb{P}_{\Gamma} \left[|R_{\text{large}}| \geq \delta \right] \leq \frac{1}{\delta} e^{-Cn}, \quad (178)$$

where

$$R_{\text{large}} = \sum_{g \subset \Gamma \text{ s.t. } \exists \gamma \subset g \text{ with } |\gamma| \geq \lambda n} K(g). \quad (179)$$

Proof: Let $\Omega_{\Gamma}(\underline{n}, \underline{m})$ be the set of all $g \subset \Gamma$ with prescribed type $(\underline{n}(g), \underline{m}(g))$. By (167) and the Markov bound

$$\mathbb{P} \left[|R_{\text{large}}| \geq \delta \right] \leq \frac{1}{\delta} \sum_{\underline{n}, \underline{m} \in \Delta} \bar{K}(\underline{n}, \underline{m}) \mathbb{E}_{\Gamma} \left[|\Omega_{\Gamma}(\underline{n}, \underline{m})| \right], \quad (180)$$

where

$$\begin{aligned} \Delta \equiv & \left\{ (\underline{n}, \underline{m}) \mid \lambda n \leq \sum_{s=2}^l n_s + \sum_{t=2}^r m_t, \sum_{s=2}^l s n_s = \sum_{t=2}^r t m_t, \right. \\ & \left. \sum_{s=2}^l n_s < n, \sum_{t=2}^r m_t < nl/r \right\}. \end{aligned} \quad (181)$$

The expectation of the number of $g \subset \Gamma$ with prescribed type can be estimated by combinatorial bounds provided by McKay [16]. It turns out that these subgraphs proliferate exponentially in n only for a subdomain of Δ where $\bar{K}(\underline{n}, \underline{m})$ is exponentially much smaller in n . In the subdomain where $\bar{K}(\underline{n}, \underline{m})$ is not small (but it is always bounded) the number of subgraphs is sub-exponential when l is odd and $l < r$. As a consequence for l odd and $l < r$, we are able to prove that the sum on the right hand side of (180) is smaller than e^{-Cn} .

Let us now give the details of this calculation. Let $\omega = 4r^2 - 2r + 2$, a number independent of n . The combinatorial bound is only valid for subgraphs g with number of edges at most equal to $nl - \omega$. Thus we have to separate the domain of summation (181) into

$$\Delta_{\omega} = \Delta \cap \left\{ (\underline{n}, \underline{m}) \mid \sum_{s=2}^l s n_s \leq nl - \omega \right\} \text{ and } \Delta_{\omega}^c = \Delta \setminus \Delta_{\omega}, \quad (182)$$

and handle each part separately.

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

$$\begin{aligned} \mathbb{E}_{\Gamma} \left[|\Omega_{\Gamma}(\underline{n}, \underline{m})| \right] &\leq \binom{nl}{\omega} \\ &= O(n^{\omega}). \end{aligned} \quad (183)$$

This is nothing but simply counting the possible subgraphs obtained by removing ω edges from Γ . For the same reason the activity (167) is upper-bounded by

$$\begin{aligned} |\bar{K}(\underline{n}, \underline{m})| &\leq (1 + \alpha \theta^r)^{n \frac{l}{r}} \left(1 + \frac{\beta}{2} (1 + 4l + l^2) \theta^2 \right)^{\omega} \\ &\times (\beta(1+l)\theta)^{n-\omega} \\ &= O \left(\left(\beta(1+l)(1+\alpha)^{\frac{l}{r}} \theta \right)^{n-\omega} \right). \end{aligned} \quad (184)$$

Indeed the activity of the total graph is upper-bounded by $\bar{K}(\Gamma) = (1 + \alpha \theta^r)^{n \frac{l}{r}} (\beta(1+l)\theta)^n$. The worst case scenario for the activity of subgraphs obtained by removing ω edges from Γ is then bounded by (184). Therefore for every $\theta < \theta_1 = \left(\beta(1+l)(1+\alpha)^{\frac{l}{r}} \right)^{-1}$ and n large enough, there exists a constant $C_1 > 0$ depending on l and r such that

$$\sum_{\underline{n}, \underline{m} \in \Delta_{\omega}^c} \bar{K}(\underline{n}, \underline{m}) \mathbb{E}_{\Gamma} \left[|\Omega_{\Gamma}(\underline{n}, \underline{m})| \right] \leq e^{C_1(n-\omega) \ln \left(\frac{\theta}{\theta_1} \right)}. \quad (185)$$

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

$$\mathbb{P}_\Gamma [g \in \Omega_\Gamma(\underline{n}, \underline{m})] \leq \frac{\prod_{s=2}^l \binom{nl}{(l-s)!}^{n_s} \prod_{t=2}^r \binom{r!}{(r-t)!}^{m_t}}{\frac{(nl-\omega)!}{(nl-\sum_{s=2}^l sn_s-\omega)!}}. \quad (186)$$

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

$$\begin{aligned} \mathbb{E}_\Gamma [|\Omega_\Gamma(\underline{n}, \underline{m})|] &\leq \binom{n}{n_2, \dots, n_l} \binom{n \frac{l}{r}}{m_2, \dots, m_r} \\ &\times \frac{\left(\sum_{s=2}^l sn_s\right)!}{\prod_{s=2}^l s!^{n_s} \prod_{t=2}^r t!^{m_t}} \mathbb{P}_\Gamma [g \in \Omega_\Gamma(\underline{n}, \underline{m})]. \end{aligned} \quad (187)$$

Setting $x_s = \frac{n_s}{n}$, $y_t = \frac{r m_t}{l n}$, we perform an asymptotic analysis for n large of the bound (187). Therefore we transform factorials using Stirling approximation valid for $k > 0$

$$e^{\frac{1}{12k+1}} \leq \frac{k!}{\sqrt{2\pi k} e^{-k} k^k} \leq e^{\frac{1}{12k}}. \quad (188)$$

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

$$\begin{aligned} (1-z) \ln(1-z) + h_2\left(\frac{\omega}{nl}\right) &\geq \left(1-z - \frac{\omega}{nl}\right) \ln\left(1-z - \frac{\omega}{nl}\right) \\ &\quad - \left(1 - \frac{\omega}{nl}\right) \ln\left(1 - \frac{\omega}{nl}\right). \end{aligned} \quad (189)$$

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

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

where H is the Shannon entropy in nat.

Using the relations (187), (188), (189) along with $h_2\left(\frac{\omega}{nl}\right) \leq \frac{\omega}{nl} \ln\left(e^{\frac{nl}{\omega}}\right)$ gives the following bound on the number of subgraphs of Γ

$$\mathbb{E}_\Gamma [|\Omega_\Gamma(\underline{n}, \underline{m})|] \leq C_{l,r} n^{\frac{\omega}{l}+2} \exp(nlf(\underline{x}, \underline{y})), \quad (190)$$

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

$$\begin{aligned} f(\underline{x}, \underline{y}) &= \left(1 - \sum_{s=2}^l \frac{s}{l} x_s\right) \ln\left(1 - \sum_{s=2}^l \frac{s}{l} x_s\right) \\ &\quad + \left(\sum_{s=2}^l \frac{s}{l} x_s\right) \ln\left(\sum_{s=2}^l \frac{s}{l} x_s\right) \\ &\quad + \frac{1}{l} \left(\sum_{s=2}^l x_s \ln\binom{l}{s}\right) + \frac{1}{r} \left(\sum_{t=2}^r y_t \ln\binom{r}{t}\right) \\ &\quad - \frac{1}{r} \left(\left(1 - \sum_{t=2}^r y_t\right) \ln\left(1 - \sum_{t=2}^r y_t\right) + \sum_{t=2}^r y_t \ln y_t\right) \\ &\quad - \frac{1}{l} \left(\left(1 - \sum_{s=2}^l x_s\right) \ln\left(1 - \sum_{s=2}^l x_s\right) + \sum_{s=2}^l x_s \ln x_s\right). \end{aligned} \quad (191)$$

The bound on the activity (167) can also be put in a form where the growth rate in n is explicit

$$\overline{K}(\underline{n}, \underline{m}) = \exp(nlk_\theta(\underline{x}, \underline{y})), \quad (192)$$

where

$$\begin{aligned} k_\theta(\underline{x}, \underline{y}) &= \frac{y_r}{r} \ln(1 + \alpha\theta^r) + \sum_{t=2}^{r-1} \frac{y_t}{r} \ln(\alpha\theta^{r-t}) \\ &\quad + \sum_{\substack{s=2, \\ \text{even}}}^{l-1} \frac{x_s}{l} \ln\left(1 + \frac{\beta}{2}(1 + 4s + s^2)\theta^2\right) \\ &\quad + \sum_{\substack{s=3, \\ \text{odd}}}^l \frac{x_s}{l} \ln(\beta(1+s)\theta). \end{aligned} \quad (193)$$

Define the ensemble

$$\begin{aligned} \Delta' \equiv &\left\{(\underline{x}, \underline{y}) \in \mathbb{R}_+^{l-1} \times \mathbb{R}_+^{r-1} \mid \lambda \leq \frac{1}{l} \sum_{s=2}^l x_s + \frac{1}{r} \sum_{t=2}^r y_t, \right. \\ &\left. \sum_{s=2}^l \frac{s}{l} x_s = \sum_{t=2}^r \frac{t}{r} y_t, \sum_{s=2}^l x_s < 1, \sum_{t=2}^r y_t < 1\right\}. \end{aligned} \quad (194)$$

It is easy to verify that if $(\underline{n}, \underline{m}) \in \Delta_\omega$ then $(\underline{x}, \underline{y}) \in \Delta'$. Combining (185), (190) and (192) gives finally

$$\sum_{\underline{n}, \underline{m} \in \Delta_\omega} \overline{K}(\underline{n}, \underline{m}) \mathbb{E}_\Gamma [|\Omega_\Gamma(\underline{n}, \underline{m})|] \leq C'_{l,r} n^{\frac{\omega}{l}+l+r} \exp(nl\Lambda), \quad (195)$$

where

$$\Lambda(\theta) = \max_{(\underline{x}, \underline{y}) \in \Delta'} f(\underline{x}, \underline{y}) + k_\theta(\underline{x}, \underline{y}). \quad (196)$$

In (185) we estimate the sum over $(\underline{n}, \underline{m}) \in \Delta_\omega$ by the crude bound $|\Delta_\omega| \leq n^{l-1} \left(\frac{nl}{r}\right)^{r-1}$.

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

$$\Delta'_0 \equiv \left\{z \in \mathbb{R}_+^{\frac{l-1}{2}} \mid \lambda \leq \sum_{s=1}^{\frac{l-1}{2}} z_s < 1\right\}. \quad (197)$$

If $z \in \Delta'_0$ then $(\underline{x}, \underline{y}) \in \Delta_0$, as we can express the variable $y_r = \sum_{s=1}^{\frac{l-1}{2}} \frac{2s}{l} z_s$ with the second constraint in (194). The restriction to Δ_0 of $f(\underline{x}, \underline{y})$ can be recast into the form

$$lf(\underline{x}, \underline{y}) = f_0(z) - \left(1 - \frac{l}{r}\right) h_2\left(\sum_{s=1}^{\frac{l-1}{2}} \frac{2s}{l} z_s\right), \quad (198)$$

where

$$\begin{aligned}
f_0(\underline{z}) = & -(l-1)h_2\left(\sum_{s=1}^{\frac{l-1}{2}}\frac{2s}{l}z_s\right) \\
& + \left(\sum_{s=1}^{\frac{l-1}{2}}z_s\ln\binom{l}{2s}\right) \\
& - \left(\left(1-\sum_{s=1}^{\frac{l-1}{2}}z_s\right)\ln\left(1-\sum_{s=1}^{\frac{l-1}{2}}z_s\right) + \sum_{s=1}^{\frac{l-1}{2}}z_s\ln z_s\right).
\end{aligned} \tag{199}$$

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

$$lf(\underline{x}, \underline{y}) < -\left(1 - \frac{l}{r}\right)\min\left\{h_2(2\lambda), h_2\left(\frac{1}{l}\right)\right\} < 0. \tag{200}$$

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

$$\sum_{\underline{n}, \underline{m} \in \Delta_\omega} \bar{K}(\underline{n}, \underline{m}) \mathbb{E}_\Gamma[|\Omega_\Gamma(\underline{n}, \underline{m})|] \leq e^{-nC_2}. \tag{201}$$

Combining Markov's inequality (180) and inequalities (185), (201) ends the proof.

Notice that the condition $\frac{l}{r} < 1$ appears naturally in (200). It is thus necessary that the graph Γ describes a code (i.e. with positive rate). ■

REFERENCES

- [1] P. O. Vontobel, "Counting in graph covers: a combinatorial characterization of the bethe entropy function," *arXiv preprint arXiv:1012.0065*, 2010.
- [2] E. B. Sudderth, M. J. Wainwright, and A. S. Willsky, "Loop series and bethe variational bounds in attractive graphical models," in *NIPS*, 2007.
- [3] N. Ruozi, "The bethe partition function of log-supermodular graphical models," *arXiv preprint arXiv:1202.6035*, 2012.
- [4] V. Chandrasekaran, M. Chertkov, D. Gamarnik, D. Shah, and J. Shin, "Counting independent sets using the bethe approximation," *SIAM Journal on Discrete Mathematics*, vol. 25, no. 2, pp. 1012–1034, 2011.
- [5] M. Chertkov and V. Y. Chernyak, "Loop series for discrete statistical models on graphs," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, no. 06, p. P06009, 2006.
- [6] D. C. Brydges, "A short course on cluster expansions," *Critical Phenomena, Random Systems, Gauge Theories*, vol. Les Houches, session XLIII, Part I, pp. 129–183, 1984.
- [7] A. Montanari, "Tight bounds for ldpc and ldgm codes under map decoding," *Information Theory, IEEE Transactions on*, vol. 51, no. 9, pp. 3221–3246, 2005.
- [8] N. Macris, "Griffith–kelly–sherman correlation inequalities: A useful tool in the theory of error correcting codes," *Information Theory, IEEE Transactions on*, vol. 53, no. 2, pp. 664–683, 2007.
- [9] S. Kudekar and N. Macris, "Sharp bounds for optimal decoding of low-density parity-check codes," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4635–4650, 2009.
- [10] T. Richardson and R. L. Urbanke, *Modern coding theory*. Cambridge University Press, 2008.
- [11] S. B. Korada, S. Kudekar, and N. Macris, "Exact solution for the conditional entropy of poissonian ldpc codes over the binary erasure channel," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*. IEEE, 2007, pp. 1016–1020.
- [12] S. Kudekar and N. Macris, "Decay of correlations for sparse graph error correcting codes," *SIAM Journal on Discrete Mathematics*, vol. 25, no. 2, pp. 956–988, 2011.
- [13] J. M. Mooij and H. J. Kappen, "Sufficient conditions for convergence of the sum-product algorithm," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4422–4437, 2007.
- [14] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.
- [15] V. Y. Chernyak and M. Chertkov, "Loop calculus and belief propagation for q-ary alphabet: loop tower," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*. IEEE, 2007, pp. 316–320.
- [16] B. D. McKay, "Subgraphs of random graphs with specified degrees," in *Proceedings of the International Congress of Mathematicians*, vol. 4, 2010, pp. 2489–2501.