

Tight bounds for LDPC and LDGM codes under MAP decoding

Andrea Montanari*

May 23, 2006

Abstract

A new method for analyzing low density parity check (LDPC) codes and low density generator matrix (LDGM) codes under bit maximum a posteriori probability (MAP) decoding is introduced. The method is based on a rigorous approach to spin glasses developed by Francesco Guerra. It allows to construct lower bounds on the entropy of the transmitted message conditional to the received one. Based on heuristic statistical mechanics calculations, we conjecture such bounds to be tight. The result holds for standard irregular *ensembles* when used over binary input output symmetric channels.

The method is first developed for Tanner graph ensembles with Poisson left degree distribution. It is then generalized to ‘multi-Poisson’ graphs, and, by a completion procedure, to arbitrary degree distribution.

Keywords : Conditional entropy, Low Density Parity Check Codes, Maximum A Posteriori probability decoding, Spin glasses, Statistical physics.

*Laboratoire de Physique Théorique de l’Ecole Normale Supérieure, 24, rue Lhomond, 75231 Paris CEDEX 05, FRANCE Internet: Andrea.Montanari@lpt.ens.fr (UMR 8549, Unité Mixte de Recherche du CNRS et de l’ENS). Financial support has been provided by the European Community under contracts STIPCO and EVERGROW (IP No. 1935 in the complex systems initiative of the Future and Emerging Technologies directorate of the IST Priority, EU Sixth Framework).

1 Introduction

Codes based on random graphs are of huge practical and theoretical relevance. The analysis of such communication schemes is currently in a mixed status. From a practical point of view, the most relevant issue is the analysis of linear-time decoding algorithms. As far as message-passing algorithms are concerned, our understanding is rather satisfactory. Density evolution [1–4] allows to compute exact thresholds for vanishing bit error probability, at least in the large blocklength limit. These results have been successfully employed for designing capacity-approaching code ensembles [5, 6].

A more classical problem is the evaluation of decoding schemes which are optimal with respect to some fidelity criterion, such as word MAP (minimizing the block error probability) or symbol MAP decoding (minimizing the bit error probability). Presently, this issue has smaller practical relevance than the previous one, and nonetheless its theoretical interest is great. Any progress in this direction would improve our understanding of the effectiveness of belief propagation (and similar message-passing algorithms) in general inference problems. Unhappily, the status of this research area [7, 10] is not as advanced as the analysis of iterative techniques. In most cases one is able to provide only upper and lower bounds on the thresholds for vanishing bit error probability in the large blocklength limit. Moreover, the existing techniques seems completely unrelated from the ones employed in the analysis of iterative decoding. This is puzzling. We know that, at least for some code constructions and some channel models, belief propagation has performances which are close to optimal. The same has been observed empirically in inference problems. Such a convergence in behavior hints at a desirable convergence in the analysis techniques.

This paper aims at bridging this gap. We introduce a new technique which allows to derive lower bounds on the entropy of the transmitted message conditional to the received one. We conjecture that the lower bound provided by this approach is indeed tight. Interestingly enough, the basic objects involved in the new bounds are probability densities over \mathbb{R} , as in the density evolution analysis of iterative decoding. These densities are required moreover to satisfy the same ‘symmetry’ condition (see Sec. 4 for a definition) as the messages distributions in density evolution. The bound can be optimized with respect to the densities. A necessary condition for the densities to be optimal is that they correspond to a fixed point of density evolution for belief propagation decoding.

The method presented in this paper is based on recent developments in the rigorous theory of mean field spin glasses. Mean field spin glasses are theoretical models for disordered magnetic alloys, displaying an extremely rich probabilistic structure [12, 13]. As shown by Nicolas Sourlas [14–16], there exists a precise mathematical correspondence between such models and error correcting codes. Exploiting this correspondence, a number of heuristic techniques from statistical mechanics have been applied to the analysis of coding systems, including LDPC codes [17–20] and turbo codes [21, 22]. Unhappily, the results obtained through this approach were non-rigorous, although, most of the times, they were expected to be exact.

Recently, Francesco Guerra and Fabio Lucio Toninelli [23, 24] succeeded in developing a general technique for constructing bounds on the free energy of mean field spin glasses. The technique, initially applied to the Sherrington-Kirkpatrick model, was later extended by Franz and Leone [25] to deal with Ising systems on random graph with Poisson distributed degrees. Finally, Franz, Leone and Toninelli [26] adapted it to systems on graphs with general degree distributions. This paper adds two improvements to this line of research. It generalizes it to

Ising systems with (some classes of) biased coupling distributions¹. Furthermore, it introduces a new way of dealing with general degree distributions which (in our view) is considerably simpler than the approach of Ref. [26]. Using the new technique, we are able to prove that the asymptotic expression for the conditional entropy of irregular LDPC ensembles derived in [20] is indeed a lower bound on the real conditional entropy. This gives further credit to the expectation that the results of [20] are exact, and we formalize this expectation as a conjecture in Sec. 10.

The new technique is based upon an interpolation procedure which progressively eliminates right (parity check) nodes from the Tanner graph. This procedure is considerably simpler for graph ensembles with Poisson left (variable) degree distribution. Such graph can be in fact constructed by adding a uniformly random right node at a time, independently from the others. We shall therefore adopt a three steps strategy. We first prove our bound for Poisson ensembles. This allows to explain the important ideas of the interpolation technique in the simplest possible context. Unhappily Poisson ensembles may have quite bad error-floor properties due to low degree node and are not very interesting for practical purposes². Next, we generalize the bound to ‘multi-Poisson’ ensembles. These can be constructed by a sequence of rounds such that, within each round, right nodes are added independently of each other. In other words multi-Poisson graphs are obtained as the superposition of several Poisson graphs. Finally, we show that a general degree distribution can be approximated arbitrarily well using a ‘multi-Poisson’ construction. Together with continuity of the bound, this implies our general result.

In Section 2 we introduce the code ensembles to be considered. Symbol-MAP decoding scheme is defined in Sec. 3, together with some basic probabilistic notations. Section 4 collects some remarks on symmetric random variables to be used in the proof of our main results. We then prove that the per-bit conditional entropy of the transmitted message concentrates in probability with respect to the code realization. This serves as a justification for considering its ensemble average. Our main result, i.e. a lower bound on the average conditional entropy is stated in Sec. 6. This Section also contains the proof for Poisson ensembles. The proof for multi-Poisson and standard ensembles is provided (respectively) in Sections 7 and 8. Section 9 presents several applications of the new bound together with a general strategy for optimizing it. Finally, we draw our conclusion and discuss extensions of our work in Section 10. Several technical calculations are deferred to the Appendices.

2 Code ensembles

In this Section we define the code ensembles to be analyzed in the rest of the paper. By ‘standard ensembles’ we refer to the irregular ensembles considered, e.g., in Refs. [2,3]. Poisson ensembles are characterized by Poisson left degree distribution. Finally multi-Poisson codes can be thought as ‘combinations’ of Poisson codes, and are mainly a theoretical device for approximating standard ensembles.

For each of the three families, we shall proceed by introducing a family of Tanner graph ensembles. In order to specify a Tanner graph, we need to exhibit a set of left (variable) nodes \mathcal{V} , of size $|\mathcal{V}| = n$, a set of right (check) nodes \mathcal{C} , with $|\mathcal{C}| = m$ and a set of edges \mathcal{E} ,

¹The reader unfamiliar with the statistical physics jargon may skip this statement.

²One exception to this statement is provided by Luby Transform codes [27]. These can be regarded as Poisson ensembles, and, due to the large average right degree, have an arbitrary small error floor

each edge joining a left and a right node. If $i \in \mathcal{V}$ denotes a generic left node and $a \in \mathcal{C}$ a generic right node, an edge joining them will be given as $(i, a) \in \mathcal{E}$. Multiple edges are allowed (although only their parity matters for the code definition). Furthermore, two graphs obtained through a permutation of the variable or of the check nodes are regarded as distinct (nodes are ‘labeled’). The neighborhood of the variable (check) node $i \in \mathcal{V}$ ($a \in \mathcal{C}$) is denoted by ∂i (∂a). In formulae $\partial i = \{a \in \mathcal{C} : (i, a) \in \mathcal{E}\}$ and $\partial a = \{i \in \mathcal{V} : (i, a) \in \mathcal{E}\}$.

A Tanner graph ensemble will be generically indicated as (\dots) , where ‘ \dots ’ is a set of relevant parameters. Expectation with respect to a Tanner graph ensemble will be denoted by $\mathbb{E}_{\mathcal{G}}$.

Next, we define LDPC(\dots) and LDGM(\dots) codes as the LDPC and LDGM codes associated to a random Tanner graph from the (\dots) ensemble. Since this construction does not depend upon the particular family of Tanner graphs to be considered, we formalize it here.

Definition 1 *Let $\mathbb{H} = \{H_{ai} : a \in \mathcal{C}; i \in \mathcal{V}\}$ be the adjacency matrix of a Tanner graph from the ensemble (\dots) : $H_{ai} = 1$ if (i, a) appears in \mathcal{E} an odd number of times, and $H_{ai} = 0$ otherwise. Then*

1. *A code from the LDGM(\dots) ensemble is the linear code on $GF[2]$ having \mathbb{H} as generator matrix. The design rate of this ensemble is defined as $r_{\text{des}} = n/\mathbb{E}_{\mathcal{G}}m$.*
2. *A code from the LDPC(\dots) ensemble is the linear code on $GF[2]$ having \mathbb{H} as parity check matrix. The design rate of this ensemble is $r_{\text{des}} = 1 - \mathbb{E}_{\mathcal{G}}m/n$.*

Before actual definitions of graph ensembles, it is convenient to introduce some notations for describing them. For a given graph we define the degree profile $(\hat{\Lambda}, \hat{P})$ as a couple of polynomials

$$\hat{\Lambda}(x) = \sum_{l=2}^{l_{\max}} \hat{\Lambda}_l x^l, \quad \hat{P}(x) = \sum_{k=2}^{k_{\max}} \hat{P}_k x^k, \quad (2.1)$$

such that $\hat{\Lambda}_i$ (\hat{P}_i) is the fraction of left (right) nodes of degree i . The degree profile $(\hat{\Lambda}, \hat{P})$, will be in general a random variable depending on the particular graph realization. On the other hand, each ensemble will be assigned a non-random ‘design degree sequence’ (Λ, P) . This is the degree profile that the ensemble is designed to achieve (and in some cases achieves with probability approaching one in the large blocklength limit). Both $\Lambda(x)$ and $P(x)$ will have non-negative coefficients and satisfy the normalization condition $\Lambda(1) = P(1) = 1$. Finally, it is useful to introduce the ‘edge perspective’ degree sequences: $\lambda(x) = \sum_l \lambda_l x^{l-1} \equiv \Lambda'(x)/\Lambda'(1)$, and $\rho(x) = \sum_k \rho_k x^{k-1} \equiv P'(x)/P'(1)$.

In the following Sections, ‘with high probability’ (w.h.p.) and similar expressions will refer to the large blocklength limit, with the other code parameters kept fixed.

2.1 Standard ensembles

Standard ensembles are discussed in several papers [2–5, 29, 30]. Their performances under iterative decoding have been thoroughly investigated allowing for ensemble optimization [2, 6]. A standard ensemble of factor graphs is defined by assigning the blocklength n and the design degree sequence (Λ, P) . We shall assume the maximum left and right degrees l_{\max} and k_{\max} to be finite.

Definition 2 *A graph from the standard ensemble (n, Λ, P) includes n left nodes and $m \equiv n\Lambda'(1)/P'(1)$ right nodes (i.e. $\mathcal{V} = [n]$ and $\mathcal{C} = [m]$), and is constructed as follows. Partition*

the set of left nodes uniformly at random into l_{\max} subsets $\{\mathcal{V}_l\}$ with $|\mathcal{V}_l| = n\Lambda_l$. For any $l = 2, \dots, l_{\max}$, associate l ‘sockets’ to each $i \in \mathcal{V}_l$. Analogously, partition the right nodes into sets $\{\mathcal{C}_k\}$ with $|\mathcal{C}_k| = nP_k$, and associate k sockets to the nodes in \mathcal{C}_k . Notice that the total number of sockets on each side is $n\Lambda'(1)$.

Choose a uniformly random permutation over $n\Lambda'(1)$ objects and connect the sockets accordingly (two connected sockets form an edge).

The ensemble (n, Λ, P) is non-empty only if the numbers $n\Lambda'(1)/P'(1)$ and $\{n\Lambda_l, mP_k\}$ are integers. The design rate of the LDPC (n, Λ, P) ensemble is $r_{\text{des}} = 1 - \Lambda'(1)/P'(1)$, while for the LDGM (n, Λ, P) ensemble $r_{\text{des}} = P'(1)/\Lambda'(1)$. It is clear that the degree profile $(\hat{\Lambda}, \hat{P})$ concentrates around the design degree sequences (Λ, P) .

An equivalent construction of a graph in the standard ensemble is the following. As before, we shall partition nodes and associate them sockets. Furthermore we shall keep track of the number of ‘free’ sockets at variable node i after t steps in the procedure, through an integer $\mathbf{d}_i(t)$. Therefore, at the beginning set $\mathbf{d}_i(0) = l$ for any $i \in \mathcal{V}_l$. Next, for any $a = 1, \dots, m$ consider the a -th check node, and assume that $a \in \mathcal{C}_k$. For $r = 1, \dots, k$ do the following operations: (i) choose i_r^a in \mathcal{V} with probability distribution $w_i(t) = \mathbf{d}_i(t)/(\sum_j \mathbf{d}_j(t))$; (ii) Set $\mathbf{d}_{i_r^a}(t+1) = \mathbf{d}_{i_r^a}(t) - 1$ and $\mathbf{d}_i(t+1) = \mathbf{d}_i(t)$ for any $i \neq i_r^a$; (iii) increment t by 1. Finally a is connected to i_1^a, \dots, i_k^a . The graph \mathcal{G} obtained after the last right node $a = m$ is connected to the left side, is distributed according to the standard ensemble as defined above.

2.2 Poisson ensembles

A Poisson ensemble is specified by the blocklength n , a real number $\gamma > 0$, and a right degree design sequence $P(x)$. Again, we require the maximum right degree k_{\max} to be finite.

Definition 3 A Tanner graph from the Poisson ensemble (n, γ, P) is constructed as follows. The graph has n variable nodes $i \in \mathcal{V} \equiv [n]$. For any $2 \leq k \leq k_{\max}$, choose m_k from a Poisson distribution with parameter $n\gamma P_k/P'(1)$. The graph has $m = \sum_k m_k$ check nodes $a \in \mathcal{C} \equiv \{(\hat{a}, k) : \hat{a} \in [m_k], 2 \leq k \leq k_{\max}\}$.

For each parity check node $a = (\hat{a}, k)$ choose k variable nodes i_1^a, \dots, i_k^a uniformly at random in \mathcal{V} , and connect a with i_1^a, \dots, i_k^a .

A few remarks are in order: they are understood to hold in the large blocklength limit $n \rightarrow \infty$ with γ and P fixed. (i) The number of check nodes is a Poisson random variable with mean $\mathbb{E}_{\mathcal{G}} m = n\gamma/P'(1)$. Moreover, m concentrates in probability around its expectation. (ii) The right degree profile \hat{P} concentrates around its expectation $\mathbb{E}_{\mathcal{G}} \hat{P}_k = P_k + O(1/n)$. (iii) The left degree profile $\hat{\Lambda}$ has expectation $\mathbb{E}_{\mathcal{G}} \hat{\Lambda}_l = \gamma^l e^{-\gamma}/l! + O(1/n)$, and concentrates around its average. In view of these remarks, we define the design left degree sequence Λ of a Poisson ensemble to be given by $\Lambda(x) = e^{\gamma(x-1)}$.

The design rate³ of the LDGM (n, γ, P) ensemble is $r_{\text{des}} = P'(1)/\gamma$, while, for a LDPC (n, γ, P) ensemble, $r_{\text{des}} = 1 - \gamma/P'(1) = 1 - \mathbb{E}_{\mathcal{G}} m/n$.

³In the first case, the actual rate r is equal to n/m , and therefore, because of the observation (i) above, concentrates around the design rate. In the LDPC (n, γ, P) ensemble, the actual rate r is always larger or equal to $1 - m/n$, because the rank of the parity check matrix \mathbb{H} is not larger than m . Notice that $1 - m/n$ concentrates around the design rate. It is not hard to show that the actual rate is, in fact, strictly larger than r_{des} with high probability. A lower bound on the number of codewords $N(\mathbb{H})$ can in fact be obtained by counting all the codewords such that $x_i = 0$ for all the variable nodes i with degree larger than 0. This implies $r \geq \hat{\Lambda}_0$. Take γ and $P(x)$ such that $e^{-\gamma} > 1 - \gamma/P'(1) + \delta$ for some $\delta > 0$ (this can be always done by choosing γ large enough). Since $\hat{\Lambda}_0$ is closely concentrated around $e^{-\gamma}$ in the $n \rightarrow \infty$ limit, we have $r > r_{\text{des}} + \delta$ with high probability.

The important simplification arising for Poisson ensembles is that their rate can be easily changed in a continuous way. Consider for instance the problem of sampling a graph from the $(n, \gamma + \Delta\gamma, P)$ ensemble. To the first order in $\Delta\gamma$ this can be done as follows. First generate a graph from the (n, γ, P) . Then, for each $k \in \{3, \dots, k_{\max}\}$, add a check node of degree k with probability $\Delta\gamma P_k/P'(1)$, and connect it to k uniformly random variable nodes i_1, \dots, i_k . Technically, this property will allow to compute derivatives with respect to the code rate, cf. App. C.

2.3 Multi-Poisson ensembles

We introduce multi-Poisson ensembles them in order to ‘approximate’ graphs in standard ensembles as the union of several Poisson sub-graphs. The construction proceeds by a finite number of rounds. During each round, we add a certain number of right nodes to the graph. The adjacent left nodes are drawn independently using a biased distribution. The bias drives the procedure towards the design left degree distribution, and is most effective in the limit of a large number of ‘small’ stages. A multi-Poisson ensemble is fully specified by the blocklength n , a design degree sequence (Λ, P) (with $\Lambda(x)$ and $P(x)$ having maximum degree, respectively, l_{\max} , and k_{\max}), and a real number $\gamma > 0$ describing the number of checks to be added at each round. The number of rounds is defined to be $t_{\max} \equiv \lfloor \Lambda'(1)/\gamma \rfloor - 1$. Below we adopt the notation $[x]_+ = x$ if $x \geq 0$ and $= 0$ otherwise.

Definition 4 *A Tanner graph \mathcal{G} from the multi-Poisson ensemble (n, Λ, P, γ) is defined by the following procedure. The graph has n variable nodes $i \in \mathcal{V} \equiv [n]$, which are partitioned uniformly at random into l_{\max} subsets $\{\mathcal{V}_l\}$, with $|\mathcal{V}_l| = n\Lambda_l$. For each $l = 2, \dots, l_{\max}$ and each $i \in \mathcal{V}_l$, let $\mathbf{d}_i(0) = l$. Let \mathcal{G}_0 be the graph with variable nodes \mathcal{V} and without any check node. We shall define a sequence $\mathcal{G}_0, \dots, \mathcal{G}_{t_{\max}}$, and set $\mathcal{G} = \mathcal{G}_{t_{\max}}$.*

For any $t = 0, \dots, t_{\max} - 1$, \mathcal{G}_{t+1} is constructed from \mathcal{G}_t as follows. For any $2 \leq k \leq k_{\max}$, choose $m_k^{(t)}$ from a Poisson distribution with parameter $n\gamma P_k/P'(1)$. Add $m^{(t)} = \sum_k m_k^{(t)}$ check nodes to \mathcal{G}_t and denote them by $\mathcal{C}_t \equiv \{(\hat{a}, k, t) : \hat{a} \in [m_k^{(t)}], 2 \leq k \leq k_{\max}\}$. For each node $a = (\hat{a}, k, t) \in \mathcal{C}_t$, choose k variable nodes i_1^a, \dots, i_k^a independently in \mathcal{V} , with distribution $w_i(t) = [\mathbf{d}_i(t)]_+ / (\sum_j [\mathbf{d}_j(t)]_+)$, and connect a with i_1^a, \dots, i_k^a . Finally, set $\mathbf{d}_i(t+1) = \mathbf{d}_i(t) - \Delta_i(t)$, where $\Delta_i(t)$ is the number of times node i has been chosen during round t .

Notice that the above procedure may fail if $\sum_j [\mathbf{d}_j(t)]_+$ vanishes for some $t < t_{\max} - 1$. However, it is easy to understand that, in the large blocklength limit, the procedure will succeed with high probability (see the proof of Lemma 1 below).

The motivation for the above definition is that, as $\gamma \rightarrow 0$ at n fixed, it reproduces the definition of standard ensembles (see the formulation at the end of Sec. 2.1). At non-zero γ , the multi-Poisson ensemble differ from the standard one in that the probabilities $w_i(t)$ are changed only every about $n\gamma$ edge connections. On the other hand, we shall be able to analyze multi-Poisson ensembles in the asymptotic limit $n \rightarrow \infty$ with the other parameters –the design distributions (Λ, P) and the stage ‘step-size’ γ – kept fixed. It is therefore crucial to estimate the ‘distance’ between multi-Poisson and standard ensembles for large n at fixed γ .

We formalize this idea using a coupling argument. Let us recall here that, given two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, a coupling among them is a random variable $(X', Y') \in \mathcal{X} \times \mathcal{Y}$ such that the marginal distributions of X' and Y' are the same as (respectively) those of X and Y . Furthermore, we define a ‘rewiring’ as the elementary operation of either adding or removing a function node from a Tanner graph. The following Lemma is proved in Appendix A.

Lemma 1 *Let $0 < \gamma < 1$ and (Λ, P) be a degree sequence pair. Then there exist two n -independent positive numbers $A(\Lambda, P), b(\Lambda, P) > 0$ and a coupling $(\mathcal{G}_s, \mathcal{G}_{\text{mP}})$, between the standard ensemble (n, Λ, P) and the multi-Poisson ensemble (n, Λ, P, γ) , such that w.h.p. \mathcal{G}_s is obtained from \mathcal{G}_{mP} with a number of rewirings smaller than $A(\Lambda, P)n\gamma^{b(\Lambda, P)}$.*

In other words, we can obtain a random Tanner graph from the standard ensemble by first generating a multi-Poisson graph with the same design degree sequences and small γ , and then changing a small fraction of edges.

Although it is convenient to define multi-Poisson ensembles in terms of the design degree distribution $\Lambda(x)$, such a distribution *does not* coincide with the actual degree profile achieved by the above construction, even in the large blocklength limit. In order to clarify this statement, let us define the expected degree distribution $\Lambda^{(n, \gamma)}(x) \equiv \mathbb{E}_C \hat{\Lambda}(x)$ for blocklength n and step parameter γ . Furthermore, let⁴ $\Lambda^{(\gamma)}(x) \equiv \lim_{n \rightarrow \infty} \Lambda^{(n, \gamma)}(x)$. We claim that, in general, $\Lambda^{(\gamma)}(x) \neq \Lambda(x)$. This can be verified explicitly using the characterization of $\Lambda^{(\gamma)}(x)$ provided in Appendix B. However, as a consequence of Lemma 1, for small γ , $\Lambda^{(\gamma)}(x)$ is ‘close’ to $\Lambda(x)$.

Corollary 1 *Let $0 < \gamma < 1$ and (Λ, P) be a degree sequence pair. Then there exist two n -independent positive numbers $A(\Lambda, P), b(\Lambda, P) > 0$ such that $|\Lambda_l^{(\gamma)} - \Lambda_l| \leq A(\Lambda, P)n\gamma^{b(\Lambda, P)}$ for each $l \in \{2, \dots, l_{\text{max}}\}$. Moreover*

$$\lim_{\gamma \rightarrow 0} \|\Lambda^{(\gamma)} - \Lambda\| = 0, \quad \text{where} \quad \|\Lambda^{(\gamma)} - \Lambda\| \equiv \frac{1}{2} \sum_l |\Lambda_l^{(\gamma)} - \Lambda_l|. \quad (2.2)$$

The distance $\|\mu - \nu\|$ defined above is often called the ‘total variation distance’: we refer to App. A for some properties.

Proof: Let $(\mathcal{G}_s, \mathcal{G}_{\text{mP}})$ be a pair of Tanner graphs distributed as in the coupling of Lemma 1. Their marginal distributions are, respectively, the standard (n, Λ, P) and the multi-Poisson (n, Λ, P, γ) ones. Denote by $\hat{\Lambda}_l(\mathcal{G}_\cdot)$ the fraction of degree- l variable nodes in graph \mathcal{G}_\cdot . Then, there exist A and b such that

$$\lim_{n \rightarrow \infty} \mathbb{P}[\exists l \text{ s.t. } |\hat{\Lambda}_l(\mathcal{G}_s) - \hat{\Lambda}_l(\mathcal{G}_{\text{mP}})| > A\gamma^b] = 0. \quad (2.3)$$

This follows from Lemma 1 together with the fact that each rewiring induces a change bounded by k_{max}/n in the degree profile. Therefore, using the notation $\Lambda_l^{(n)} = \mathbb{E} \hat{\Lambda}_l(\mathcal{G}_s)$ for the expected degree profile in the standard ensemble at finite blocklength, we get

$$|\Lambda_l^{(\gamma, n)} - \Lambda_l^{(n)}| = |\mathbb{E}[\hat{\Lambda}_l(\mathcal{G}_s) - \hat{\Lambda}_l(\mathcal{G}_{\text{mP}})]| \leq \mathbb{E}|\hat{\Lambda}_l(\mathcal{G}_s) - \hat{\Lambda}_l(\mathcal{G}_{\text{mP}})| \leq A\gamma^b + o_n(1). \quad (2.4)$$

The first thesis follows by taking the $n \rightarrow \infty$ limit.

Convergence in total variation distance follows immediately:

$$\begin{aligned} \|\Lambda^{(\gamma)} - \Lambda\| &= \frac{1}{2} \sum_l |\Lambda_l^{(\gamma)} - \Lambda_l| \leq \frac{1}{2} \sum_{l \leq l_*} |\Lambda_l^{(\gamma)} - \Lambda_l| + \frac{1}{2} \sum_{l > l_*} \Lambda_l + \frac{1}{2} \sum_{l > l_*} \Lambda_l^{(\gamma)} = \\ &= \frac{1}{2} \sum_{l \leq l_*} |\Lambda_l^{(\gamma)} - \Lambda_l| + \frac{1}{2} (1 - \sum_{l \leq l_*} \Lambda_l) + \frac{1}{2} (1 - \sum_{l > l_*} \Lambda_l^{(\gamma)}). \end{aligned} \quad (2.5)$$

⁴In App. B we shall prove the existence (in an appropriate sense) of this limit and provide an efficient way of computing $\Lambda^{(\gamma)}(x)$.

The last expression can be made arbitrarily small by choosing γ and l_* appropriately. For instance, one can choose l_* in such a way that $\sum_{l \leq l_*} \Lambda_l \geq 1 - \varepsilon$, and then γ such that $A\gamma^b \leq \varepsilon/l_*$. This implies $\|\Lambda^{(\gamma)} - \Lambda\| \leq 3\varepsilon$. \square

3 Decoding Schemes

In the LDGM case the codeword bits are naturally associated to the check nodes. A string $\hat{\underline{x}} = \{\hat{x}_a : a \in \mathcal{C}\} \in \{0, 1\}^m$ is a codeword if and only if there exists an information message $\underline{x} = \{x_i : i \in \mathcal{V}\} \in \{0, 1\}^n$ such that

$$\hat{x}_a = x_{i_1^a} \oplus \cdots \oplus x_{i_k^a}, \quad (3.1)$$

for each $a = (\hat{a}, k) \in \mathcal{C}$. Here \oplus denotes the sum modulo 2. Encoding consists in choosing an information message \underline{x} with uniform probability distribution and constructing the corresponding codeword $\hat{\underline{x}}$ using the equations (3.1). Notice that, because the code is linear, each codeword is the image of the same number of information messages. Therefore choosing an information message uniformly at random probability is equivalent to choosing a codeword uniformly at random.

In the LDPC case the codeword bits can be associated to variable nodes. A string $\underline{x} = \{x_i : i \in \mathcal{V}\} \in \{0, 1\}^n$ is a codeword if and only if it satisfies the parity check equations

$$x_{i_1^a} \oplus \cdots \oplus x_{i_k^a} = 0, \quad (3.2)$$

for each $a = (\hat{a}, k) \in \mathcal{C}$. In the encoding process we pick a codeword with uniform distribution.

The codeword, chosen according to the above encoding process, is transmitted on a binary-input output symmetric channel (BIOS) with output alphabet \mathcal{A} and transition probability density $Q(y|x)$. In the following we shall use a discrete notation for \mathcal{A} . It is straightforward to adapt the formulas below to the continuous case. If, for instance $\mathcal{A} = \mathbb{R}$, sums should be replaced by Lebesgue integrals: $\sum_{y \in \mathcal{A}} \cdot \rightarrow \int dy \cdot$.

The channel output has the form $\hat{\underline{y}} = \{y_a : a \in \mathcal{C}\} \in \mathcal{A}^m$ in the LDGM case or $\underline{y} = \{y_i : i \in \mathcal{V}\} \in \mathcal{A}^n$ LDPC case. In order to keep unified notation for the two cases, we shall introduce a simple convention which introduces a fictitious output \underline{y} , associated to the variable nodes, (in the LDGM case) or $\hat{\underline{y}}$, associated to the check nodes (in the LDPC case). If an LDGM is used, \underline{y} takes by definition a standard value $\underline{y}_* = (*, \dots, *)$, while of course $\hat{\underline{y}}$ is determined by the transmitted codeword and the channel realization. The character $*$ should be thought as an erasure. If we are considering a LDPC, \underline{y} is the channel output, while $\hat{\underline{y}}$ takes a standard value $\hat{\underline{y}}_0 = (0, \dots, 0)$.

We will focus on the probability distribution $P(\underline{x}|\underline{y}, \hat{\underline{y}})$ of the vector \underline{x} , conditional to the channel output $(\underline{y}, \hat{\underline{y}})$. Depending upon the family of codes employed (whether LDGM or LDPC), this distribution has different meanings. It is the distribution of the information message in the LDGM case, and the distribution of the transmitted codeword in the LDPC case. It can be always written in the form

$$P(\underline{x}|\underline{y}, \hat{\underline{y}}) = \frac{1}{\mathcal{Z}(\underline{y}, \hat{\underline{y}})} \prod_{a \in \mathcal{C}} Q_{\mathcal{C}}(\hat{y}_a | x_{i_1^a} \oplus \cdots \oplus x_{i_k^a}) \prod_{i \in \mathcal{V}} Q_{\mathcal{V}}(y_i | x_i). \quad (3.3)$$

The precise form of the functions $Q_{\mathcal{C}}(\cdot|\cdot)$ and $Q_{\mathcal{V}}(\cdot|\cdot)$ depends upon the family of codes:

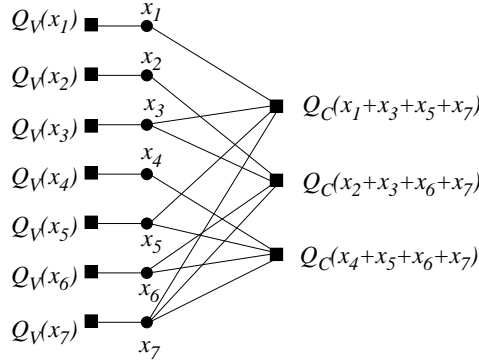


Figure 1: Factor graph representation of the probability distribution of the transmitted codeword (information message for LDGM codes) conditional to the channel output, cf. Eq. (3.3). Notice that, for the sake of compactness the y arguments have been omitted.

For LDGM's:

$$Q_C(\hat{y}|\hat{x}) = Q(\hat{y}|\hat{x}), \quad Q_V(y|x) = \begin{cases} 1 & \text{if } y = *, \\ 0 & \text{otherwise} \end{cases}. \quad (3.4)$$

For LDPC's

$$Q_C(\hat{y}|\hat{x}) = \begin{cases} 1 & \text{if } \hat{y} = \hat{x}, \\ 0 & \text{otherwise} \end{cases}, \quad Q_V(y|x) = Q(y|x). \quad (3.5)$$

The probability distribution (3.3) can be conveniently represented in the factor graphs language [31], see Fig. 1. There are two type of factor nodes in such a graph: nodes corresponding to $Q_V(\cdot)$ terms on the left, and nodes corresponding to $Q_C(\cdot)$ on the right. It is also useful to introduce a specific notation for the expectation with respect to the distribution (3.3). If $F : \{0,1\}^{\mathcal{V}} \rightarrow \mathbb{R}$ is a function of the codeword (information message for LDGM codes), we define:

$$\langle F(\underline{x}) \rangle \equiv \sum_{\underline{x}} F(\underline{x}) P(\underline{x}|\underline{y}, \hat{y}). \quad (3.6)$$

In the proof of our main result, see Sec. 6, it will be also useful to consider several i.i.d. copies of \underline{x} , each one having a distribution of the form (3.3). If $F : \{0,1\}^{\mathcal{V}} \times \dots \times \{0,1\}^{\mathcal{V}} \rightarrow \mathbb{R}$, $(\underline{x}^{(1)}, \dots, \underline{x}^{(q)}) \mapsto F(\underline{x}^{(1)}, \dots, \underline{x}^{(q)})$ is a real function of q such copies, we denote its expectation value by

$$\langle F(\underline{x}^{(1)}, \dots, \underline{x}^{(q)}) \rangle_* \equiv \sum_{\underline{x}^{(1)} \dots \underline{x}^{(q)}} F(\underline{x}^{(1)}, \dots, \underline{x}^{(q)}) P(\underline{x}^{(1)}|\underline{y}, \hat{y}) \cdots P(\underline{x}^{(q)}|\underline{y}, \hat{y}). \quad (3.7)$$

We are interested in two different decoding schemes: bit MAP (for short MAP hereafter) and iterative belief propagation (BP) decoding. In MAP decoding we follow the rule:

$$x_i^{\text{MAP}} = \arg \max_{x_i} P(x_i|\underline{y}, \hat{y}), \quad (3.8)$$

where $P(x_i|\underline{y}, \hat{y})$ is obtained by marginalizing the probability distribution (3.3) over $\{x_j; j \neq i\}$.

BP decoding [32] is a message passing algorithm. The received message is encoded in terms of log-likelihoods $\{h_i : i \in \mathcal{V}\}$ and $\{J_a : a \in \mathcal{C}\}$ as follows (notice the unusual normalization):

$$h_i = \frac{1}{2} \log \frac{Q_{\mathcal{V}}(y_i|0)}{Q_{\mathcal{V}}(y_i|1)}, \quad J_a = \frac{1}{2} \log \frac{Q_{\mathcal{C}}(\hat{y}_a|0)}{Q_{\mathcal{C}}(\hat{y}_a|1)}. \quad (3.9)$$

The messages $\{u_{a \rightarrow i}, v_{i \rightarrow a}\}$ are updated recursively following the rule

$$u_{a \rightarrow i} := \operatorname{arctanh}\left\{\tanh J_a \prod_{j \in \partial a \setminus i} \tanh v_{j \rightarrow a}\right\}, \quad (3.10)$$

$$v_{i \rightarrow a} := h_i + \sum_{b \in \partial i \setminus a} u_{b \rightarrow i}. \quad (3.11)$$

After iterating Eqs. (3.10), (3.11) a certain fixed number of times, all the incoming messages to a certain bit x_i are used to estimate its value.

In the following, we shall denote by $\mathbb{E}_{\mathcal{C}}$ the expectation with respect to one of the code ensembles defined in this Section. Which one of the ensembles defined above will be clear from the context. We will denote by $\mathbb{E}_{\mathbf{y}}$ the expectation with respect to the received message $(\mathbf{y}, \hat{\mathbf{y}})$, assuming the transmitted one to be the all-zero codeword (or, in some cases, with respect to one of the received symbols).

4 Random variables

It is convenient to introduce a few notations and simple results for handling some particular classes of random variables. Since these random variables will appear several times in the following, this introduction will help to keep the presentation more compact. Here, as in the other Sections, we shall sometimes use the standard convention of denoting random variables by upper case letters (e.g. X) and deterministic values by lower case letters (e.g. x). Moreover, we use the symbol $\stackrel{d}{=}$ to denote identity in distribution (i.e. $X \stackrel{d}{=} Y$ if X and Y have the same distribution).

The most important class of random variables in the present paper is the following.

Definition 5 *A random variable X is said to be symmetric (and we write $X \in \mathcal{S}$) if it takes values in $(-\infty, +\infty]$ and*

$$\mathbb{E}_X[g(x)] = \mathbb{E}_X[e^{-2x}g(-x)], \quad (4.1)$$

for any real function g such that at least one of the expectation values exists.

This definition was already given in [33]. Notice however that the two definitions differ by a factor 2 in the normalization of X . The introduction of symmetric random variables is motivated by the following observations [33]:

1. If $Q(y|x)$ is the transition probability of a BIOS and y is distributed according to $Q(y|0)$, then the log-likelihood

$$\ell(y) = \frac{1}{2} \log \frac{Q(y|0)}{Q(y|1)}, \quad (4.2)$$

is a symmetric random variable. In particular the log-likelihoods $\{h_i : i \in \mathcal{V}\}$ and $\{J_a : a \in \mathcal{C}\}$ defined in (3.9) are symmetric random variables.

2. Two important symmetric variables are: (i) $X = 0$ with probability 1; (ii) $X = +\infty$ with probability 1. Both can be considered as particular examples of the previous observation. Just take an erasure channel with erasure probability 1 (case i) or 0 (case ii).
3. If X and Y are symmetric then $Z = X + Y$ is symmetric. Here the sum is extended to the domain $(-\infty, +\infty]$ by imposing the rule $x + \infty = +\infty$.
4. If X and Y are symmetric then $Z = \operatorname{arctanh}(\tanh X \tanh Y)$ is symmetric. The functions $x \mapsto \tanh x$ and $x \mapsto \operatorname{arctanh} x$ are extended to the domains (respectively) $(-\infty, +\infty]$ and $(-1, +1]$ by imposing the rules $\tanh(+\infty) = +1$ and $\operatorname{arctanh}(+1) = +\infty$.

As a consequence of the above observations, if the messages $\{u_{a \rightarrow i}, v_{i \rightarrow a}\}$ in BP decoding are initialized to some symmetric random variable, remain symmetric at each step of the decoding procedure. This follows directly from the update equations (3.10) and (3.11).

Remarkably, the family of symmetric variables is ‘stable’ also under MAP decoding. This is stated more precisely in the result below [30].

Lemma 2 *Let $P(\underline{x}|\underline{y}, \hat{\underline{y}})$ be the probability distribution of the channel input (information message) $\underline{x} = (x_1 \dots x_n)$, conditional to the channel output $\underline{y} = (y_1 \dots y_n)$, $\hat{\underline{y}} = (\hat{y}_1 \dots \hat{y}_n)$ as given in Eq. (3.3). Assume the channel is BIOS and \underline{x} is the codeword of (is coded using a) linear code. Let $\underline{i} \equiv \{i_1 \dots i_k\} \subseteq [n]$ and define*

$$\ell_{\underline{i}}(\underline{y}) = \frac{1}{2} \log \frac{P(x_{i_1} \oplus \dots \oplus x_{i_k} = 0 | \underline{y})}{P(x_{i_1} \oplus \dots \oplus x_{i_k} = 1 | \underline{y})}. \quad (4.3)$$

If \underline{y} is distributed according to the channel, conditional to the all-zero codeword being transmitted then $\ell_{\underline{i}}(\underline{y})$ is a symmetric random variable.

To complete our brief review of properties of symmetric random variables, it is useful to collect a few identities to be used several times in the following (throughout the paper we use \log and \log_2 to denote, respectively, natural and base 2 logarithms).

Lemma 3 *Let X be a symmetric random variable. Then the following identities hold*

$$\mathbb{E}_X \tanh^{2k-1} X = \mathbb{E}_X \tanh^{2k} X = \mathbb{E}_X \frac{\tanh^{2k} X}{1 + \tanh X}, \quad \text{for } k \geq 1, \quad (4.4)$$

$$\mathbb{E}_X \log(1 + \tanh X) = \sum_{k=1}^{\infty} \left(\frac{1}{2k-1} - \frac{1}{2k} \right) \mathbb{E}_X \tanh^{2k} X. \quad (4.5)$$

Proof: The identities (4.4) follow from the observation that, because of Eq. (4.1), we have

$$\mathbb{E} g(X) = \frac{1}{2} \mathbb{E} [g(X) + e^{-2X} g(-X)]. \quad (4.6)$$

The desired result is obtained by substituting either $g(X) = \tanh^{2k} X$ or $g(X) = \tanh^{2k-1} X$.

In order to get (4.5), apply the identity (4.6) to $g(X) = \log(1 + \tanh X)$ and get

$$\mathbb{E} \log(1 + t) = \frac{1}{2} \mathbb{E} \frac{1}{1+t} [(1+t) \log(1+t) + (1-t) \log(1-t)] = \quad (4.7)$$

$$= \mathbb{E} \sum_{k=1}^{\infty} \left(\frac{1}{2k-1} - \frac{1}{2k} \right) \frac{t^{2k}}{1+t}, \quad (4.8)$$

where we introduced the shorthand $t = \tanh X$. At this point you can switch sum and expectation because of the monotone convergence theorem. \square

The space of symmetric random variables is useful because log-likelihoods (for our type of problems) naturally belong to this space. It is also useful to have a compact notation for the distribution of a binary variable x whose log-likelihood is $u \in (-\infty, +\infty]$. We therefore define

$$P_u(x) = \begin{cases} (1 + e^{-2u})^{-1}, & \text{if } x = 0, \\ (1 + e^{-2u})^{-1}e^{-2u}, & \text{if } x = 1. \end{cases} \quad (4.9)$$

5 Concentration

Our main object of interest will be the entropy per input symbol of the transmitted message conditional to the received one (we shall generically measure entropies in bits). We take the average of this quantity with respect to the code ensemble to define

$$h_n = \frac{1}{n} \mathbb{E}_C H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}}) = \quad (5.1)$$

$$= \frac{1}{n} \mathbb{E}_C \mathbb{E}_y \log_2 \mathcal{Z}(\underline{y}, \hat{\underline{y}}) - \frac{\gamma}{P'(1)} \sum_y Q_C(y|0) \log_2 Q_C(y|0) \\ - \sum_y Q_V(y|0) \log_2 Q_V(y|0). \quad (5.2)$$

In passing from Eq. (5.1) to (5.2), we exploited the symmetry of the channel, and fixed the transmitted message to be the all-zero codeword.

Intuitively speaking, the conditional entropy appearing in Eq. (5.1) allows to estimate the typical number of inputs with non-negligible probability for a given channel output. The most straightforward rigorous justification for looking at the conditional entropy is provided by Fano's inequality [34] which we recall here.

Lemma 4 *Let $P_B(C)$ ($P_b(C)$) be the block (bit) error probability for a code C having block-length n and rate r . Then*

1. $P_B(C) \geq [H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}}) - 1]/(nr)$.
2. $\mathbf{h}(P_b(C)) \geq H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}})/n$.

Here $\mathbf{h}(x) \equiv -x \log_2 x - (1-x) \log_2 (1-x)$ denotes the binary entropy function.

The rationale for taking the expectation with respect to the code in the definition (5.1) is the following concentration result

Theorem 1 *Let $H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}})$ be the relative entropy for a code drawn from any of the ensembles LDGM(\dots) or LDPC(\dots) defined in Section 2, when used to communicate over a binary memoryless channel. Then there exist two (n -independent) constants $A, B > 0$ such that, for any $\varepsilon > 0$*

$$\mathbb{P}_C\{|H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}}) - nh_n| > n\varepsilon\} \leq A e^{-nB\varepsilon^2}. \quad (5.3)$$

Here \mathbb{P}_C denotes the probability with respect to the code realization.

In particular this result implies that, if h_n is bounded away from zero in the $n \rightarrow \infty$ limit, then, with high probability with respect to the code realization, the bit error rate is bounded away from 0. The converse (namely that $\lim_{n \rightarrow \infty} h_n = 0$ implies $\lim_{n \rightarrow \infty} \mathbb{P}_C[P_b(C) > \delta] = 0$) is in general false. However, for many cases of interests (in particular for LDPC ensembles) we expect this to be the case. We refer to Sec. 10 for further discussion of this point.

Proof: We use an Azuma inequality [35] argument similar to the one adopted by Richardson and Urbanke to prove concentration under message passing decoding [3].

Notice that the code-dependent contribution to $H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}})$ is $\mathbb{E}_y \log_2 \mathcal{Z}(\underline{y}, \hat{\underline{y}})$, cf. Eq. (5.2). We are therefore led to construct a Doob martingale as follows. First of all fix $m_{\max} = (1+\delta)\mathbb{E}_{\mathcal{G}}m$, and condition to $m \leq m_{\max}$ (m being the number of right nodes). A code C in this ‘constrained’ ensemble can be thought as a sequence of m_{\max} random variables $c_1, \dots, c_{m_{\max}}$. The variable $c_t \equiv (k, \underline{i})$ consists in the degree k of the t -th check node, plus the list $\underline{i} = \{i_1^a \dots i_k^a\}$ of adjacent variable nodes on the Tanner graph. We adopt the convention $c_t = *$ if $t > m$. Let $C_t \equiv (c_1, \dots, c_t)$ and define the random variables

$$X_t = \mathbb{E}_C[\mathbb{E}_y \log_2 \mathcal{Z}(\underline{y}, \hat{\underline{y}}) | C_t, m \leq m_{\max}], \quad t = 0, 1, \dots, m_{\max}. \quad (5.4)$$

It is obvious that the sequence $X_0, \dots, X_{m_{\max}}$ form a martingale. In particular $\mathbb{E}[X_t | X_0 \dots X_{t-1}] = X_{t-1}$. In order to apply Azuma inequality we need an upper bound on the differences $|X_t - X_{t-1}|$. Consider two Tanner graphs which differ in a unique check node and let Δ be a uniform upper bound on the difference in $\mathbb{E}_y \log_2 \mathcal{Z}(\underline{y}, \hat{\underline{y}})$ among these two graphs. Since X_{t-1} is the expectation of X_t with respect to the choice of the t -th check node, $|X_t - X_{t-1}| \leq \Delta$ as well. In order to derive such an upper bound Δ , we shall compare two graphs differing in a unique check node, with the graph obtained by erasing this node.

More precisely, consider a Tanner graph in the ensemble having m check nodes, and channel output $\underline{y}, \hat{\underline{y}}$. Now add a single check node, to be denoted as 0. Let \hat{y}_0 be the corresponding observed value, and $i_1^0 \dots i_k^0$ the positions of the adjacent bits. In the LDGM case \hat{y}_0 will be drawn from the channel distribution, while it is fixed to 0 if the code is a LDPC. Evaluate the difference of the corresponding partition functions. We claim that

$$\mathbb{E}_{\hat{y}_0} \mathbb{E}_y \log_2 \mathcal{Z}_{m+1}(\underline{y}, \hat{\underline{y}}, \hat{y}_0) - \mathbb{E}_y \log \mathcal{Z}_m(\underline{y}, \hat{\underline{y}}) = \mathbb{E}_{\hat{y}_0} \mathbb{E}_y \log_2 \langle Q_C(\hat{y}_0 | x_{i_1} \oplus \dots \oplus x_{i_k}) \rangle, \quad (5.5)$$

where $\langle \cdot \rangle$ denotes the expectation with respect to the distribution (3.3) for the m -check nodes code. In order to prove such a formula, we write explicitly $\log_2 \langle Q_C(\hat{y}_0 | x_{i_1} \dots x_{i_k}) \rangle$ using Eq. (3.3). We get

$$\log_2 \left\{ \frac{1}{\mathcal{Z}_m(\underline{y}, \hat{\underline{y}})} \sum_{\underline{x}} \prod_{a \in [m]} Q_C(\hat{y}_a | x_{i_1^a} \dots x_{i_k^a}) \prod_{i \in [n]} Q_V(y_i | x_i) \cdot Q_C(\hat{y}_0 | x_{i_1} \dots x_{i_k}) \right\}. \quad (5.6)$$

Next we apply the definition of $\mathcal{Z}_{m+1}(\underline{y}, \hat{\underline{y}}, \hat{y}_0)$, and take expectation with respect to $\underline{y}, \hat{\underline{y}}$ and \hat{y}_0 .

Using the definitions (4.2) and (4.3) we obtain

$$\langle Q_C(\hat{y}_0 | x_{i_1} \oplus \dots \oplus x_{i_k}) \rangle = \frac{1}{2} [1 + \tanh l(y_0) \tanh \ell_{\underline{i}}(\underline{y})], \quad (5.7)$$

with $\underline{i} = (i_0 \dots i_k)$. It follows therefore from Lemma 3 that

$$-1 \leq \mathbb{E}_{\hat{y}_0} \mathbb{E}_y \log \langle Q_C(\hat{y}_0 | x_{i_1} \oplus \dots \oplus x_{i_k}) \rangle \leq 0. \quad (5.8)$$

Therefore the difference in conditional entropy among two Tanner graph which differ in a unique check node is at most 2 bits (one bit for removing the check node plus one bit for adding it in a different position). Arguing as above, this yields $|X_{t+1} - X_t| \leq 2$ and Azuma inequality implies

$$\mathbb{P}_C\{|H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}}) - nh_n| > n\varepsilon \mid m \leq m_{\max}\} \leq A_1 e^{-nB_1\varepsilon^2}. \quad (5.9)$$

with $A_1 = 2$ and $B_1 = P'(1)/[8\gamma(1 + \delta)]$.

In the case of standard ensembles, we are done, because the condition $m \leq m_{\max} = (1 + \delta)\mathbb{E}_G m$ holds surely (m is a deterministic quantity). For Poisson and multi-Poisson ensembles, we have still to show that the event $m > m_{\max}$ does not modify significantly the estimate (5.9). From elementary theory of Poisson random variables we know that, for any $\delta > 0$ there exist $A_2, B_2 > 0$ such that $\mathbb{P}_C[m > m_{\max}] \leq A_2 e^{-nB_2\delta^2}$. Notice that

$$\begin{aligned} \mathbb{P}_C\{|H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}}) - nh_n| > n\varepsilon\} &\leq \mathbb{P}_C\{|H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}}) - nh_n| > n\varepsilon \mid m \leq m_{\max}\} + \\ &\quad + \mathbb{P}_C[m > m_{\max}] \leq \end{aligned} \quad (5.10)$$

$$\leq A_1 e^{-nB_1\varepsilon^2} + A_2 e^{-nB_2\delta^2}. \quad (5.11)$$

The thesis is obtained by choosing $\delta = \varepsilon$, $B = \min(B_1, B_2)$, and $A = A_1 + A_2$. \square

6 Main result and proof for Poisson ensembles

As briefly mentioned in the Introduction, the main ideas in the proof are most clearly described in the context of simple Poisson ensembles. We shall therefore discuss them in detail here, will be more succinct when using similar arguments for multi-Poisson ensembles. Some of the calculations and technical details are deferred to Appendices C and D.

In order to state our main result in compact form, it is useful to introduce two infinite families $\{U_A\}$, $\{V_B\}$ of i.i.d. random variables. The indices A , and B run over whatever set including the cases occurring in the paper. We adopt the convention that any two variables of these families carrying distinct subscripts are independent. The distribution of the U and V variables shall moreover satisfy a couple of requirements specified in the definition below.

Definition 6 *Fix a degree sequence pair (Λ, P) , and let ρ_k be the edge perspective right degree distribution $\rho_k = P_k/P'(1)$. Let $V_1, V_2, \dots \stackrel{d}{=} V$ be a family of i.i.d. symmetric random variables, k an integer with distribution ρ_k , J a symmetric random variable distributed as the log-likelihoods $\{J_a\}$ of the parity check bits, cf. Eq. (3.9), and*

$$U^V \equiv \operatorname{arctanh} \left[\tanh J \prod_{i=1}^{k-1} \tanh V_i \right]. \quad (6.1)$$

The random variables U, V are said to be admissible if they are independent, symmetric and $U \stackrel{d}{=} U^V$.

For any couple of admissible random variables U, V , we define the associated trial entropy as follows

$$\begin{aligned} \phi_V(\Lambda, P) &\equiv -\Lambda'(1) \mathbb{E}_{u,v} \log_2 \left[\sum_x P_u(x) P_v(x) \right] + \mathbb{E}_l \mathbb{E}_y \mathbb{E}_{\{u_i\}} \log_2 \left[\sum_x \frac{Q_V(y|x)}{Q_V(y|0)} \prod_{i=1}^l P_{u_i}(x) \right] + \\ &\quad + \frac{\Lambda'(1)}{P'(1)} \mathbb{E}_k \mathbb{E}_{\hat{y}} \mathbb{E}_{\{v_i\}} \log_2 \left[\sum_{x_1 \dots x_k} \frac{Q_C(\hat{y}|x_1 \oplus \dots \oplus x_k)}{Q_C(\hat{y}|0)} \prod_{i=1}^k P_{v_i}(x_i) \right], \end{aligned} \quad (6.2)$$

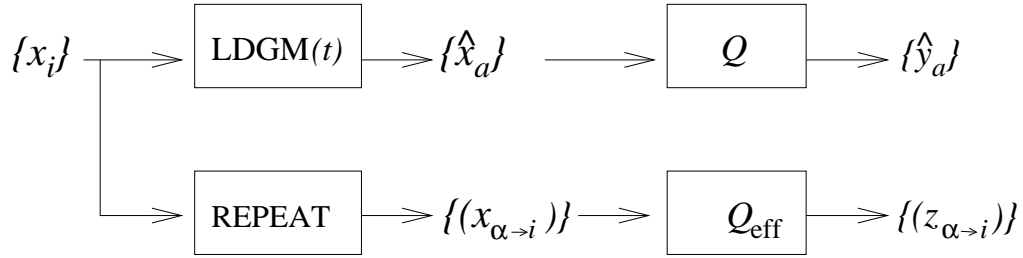


Figure 2: A communication scheme interpolating between an LDGM code and an irregular repetition code. Notice that the repetition part is transmitted through a different (effective) channel.

where l and k are two integer random variables with distribution (respectively) Λ_l and P_k . Hereafter we shall drop the reference to the degree distributions in $\phi_V(\Lambda, P)$ whenever this is clear from the context.

Notice that, in the notation for the trial entropy, we put in evidence its dependence just on the distribution of the V -variables. In fact we shall think of the distribution of the U -variables as being determined by the relation $U \stackrel{d}{=} U^V$, see Eq. (6.1).

The main result of this paper is stated below.

Theorem 2 *Let $P(x)$ be a polynomial with non-negative coefficients such that $P(0) = P'(0) = 0$, and assume that $P(x)$ is convex for $x \in [-x_0, x_0]$.*

1. *Let h_n be the expected conditional entropy per bit for either of the Poisson ensembles LDGM(n, γ, P) or LDPC(n, γ, P). If $x_0 \geq 1$, then*

$$h_n \geq \sup_V \phi_V(\Lambda, P). \quad (6.3)$$

2. *Let h_n be the expected conditional entropy per bit for either of the multi-Poisson ensembles LDGM(n, γ, Λ, P) or LDPC(n, γ, Λ, P). If $x_0 > e$, then*

$$\liminf_{n \rightarrow \infty} h_n \geq \sup_V \phi_V(\Lambda^{(\gamma)}, P). \quad (6.4)$$

3. *Let h_n be the expected conditional entropy per bit for either of the standard ensembles LDGM(n, Λ, P) or LDPC(n, Λ, P). If $x_0 > e$, then*

$$\liminf_{n \rightarrow \infty} h_n \geq \sup_V \phi_V(\Lambda, P). \quad (6.5)$$

Here the sup has to be taken over the space of admissible random variables.

Proof [Poisson ensembles]: Computing the conditional entropy (5.2) is difficult because the probability distribution (3.3) does not factorize over the bits $\{x_i\}$. Guerra's idea [23] consists in interpolating between the original distribution (3.3) and a much simpler one which factorizes over the bits. In the LDPC case, this corresponds to progressively removing parity check conditions (3.2), from the code definition. For LDGM's, it amounts to removing bits from the codewords (3.1). In both cases the design rate is augmented. In order to compensate the increase in conditional entropy implied by this transformation we imagine to re-transmit some

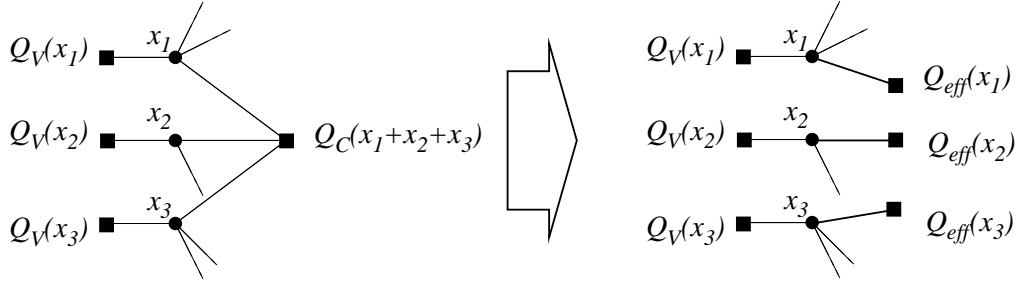


Figure 3: Action of the interpolation procedure on the factor graph representing the probability distribution of the channel input (or, in LDGM case, the information message) conditional to the channel output. For the sake of simplicity we dropped the arguments y_i (for Q_V), \hat{y}_a (for Q_C) and $z_{\alpha \rightarrow i}$ (for Q_{eff}).

of the bits $\{x_i\}$ through an ‘effective’ channel. It turns out that the transition probability of this effective channel can be chosen in such a way to ‘match’ the transformation of the original code.

In practice, for any $s \in [0, \gamma]$ we define the interpolating system as follows. We construct a code from the same ensemble family as the original one, but with modified parameters (n, s, P) . Notice that both the block length and the right degree distribution remained unchanged. The design rate, on the other hand has been increased to $r_{des}(s) = P'(1)/s$ (in the LDGM case) or $r_{des}(t) = 1 - s/P'(1)$ (in the LDPC case). A codeword from this code is transmitted through the original channel, with transition matrix $Q(\cdot|\cdot)$. It is useful to denote by \mathcal{C}_s the set of factor nodes for a given value of t . Of course \mathcal{C}_s is a random variable.

As anticipated, we must compensate for the rate loss in the above procedure. We therefore repeat each bit x_i l_i times and transmit it through an auxiliary channel with transition probability $Q_{eff}(\cdot|\cdot)$. The l_i ’s are taken to be independent Poisson random variables with parameter $(\gamma - s)$. We can therefore think this construction as a code formed of two parts: a code from the LDGM(n, s, P) or LDPC(n, s, P) ensemble, plus an irregular repetition code. Each part of the code is transmitted through a different channel. In Fig. 2 we present a schematic description of this two-parts coding system (the scheme refers to the LDGM case).

The received message will have the form $(\underline{y}, \underline{\hat{y}}, \underline{z})$, with $\underline{z} = \{z_{\alpha \rightarrow i}\}$, $\alpha_i \in [l_i]$. We can write the conditional probability for $\underline{x} = (x_1 \dots x_n)$ to be the transmitted codeword conditional to the output $(\underline{y}, \underline{\hat{y}}, \underline{z})$ of the channel as follows:

$$P(\underline{x}|\underline{y}, \underline{\hat{y}}, \underline{z}) = \frac{1}{\mathcal{Z}_s} \prod_{a \in \mathcal{C}_s} Q_C(\hat{y}_a | x_{i_1^a} \oplus \dots \oplus x_{i_k^a}) \prod_{i \in \mathcal{V}} Q_V(y_i | x_i) \prod_{i \in \mathcal{V}} \prod_{\alpha_i=1}^{l_i} Q_{eff}(z_{\alpha_i \rightarrow i} | x_i), \quad (6.6)$$

with $\mathcal{Z}_s = \mathcal{Z}(\underline{y}, \underline{\hat{y}}, \underline{z})$ fixed by the normalization condition of $P(\underline{x}|\underline{y}, \underline{\hat{y}}, \underline{z})$. Notice that, for $s = \gamma$ the original distribution (3.3) is recovered since $l_i = 0$ for any $i \in \mathcal{V}$. On the other hand, if $s = 0$, then $m = 0$ and the Tanner graph contains no check nodes: $\mathcal{C}_0 = \emptyset$. We are left with a simple irregular repetition code. The action of the interpolation procedure on the factor graph is depicted in Fig. 3.

The following steps in the proof will depend upon the effective channel transition probability $Q_{eff}(\cdot|\cdot)$ only through its log-likelihood distribution. We therefore define the random variable

U as

$$U \stackrel{d}{=} \frac{1}{2} \log \frac{Q_{\text{eff}}(Z|0)}{Q_{\text{eff}}(Z|1)}. \quad (6.7)$$

Where Z is distributed according to $Q_{\text{eff}}(z|0)$. Notice that U is symmetric and that, for any symmetric U we can construct at least one BIOS channel whose log-likelihood is distributed as U [3]. Hereafter we shall consider $Q_{\text{eff}}(\cdot|\cdot)$ to be determined by such a construction. When necessary, we shall adopt a discrete notation for the output alphabet of the effective channel $Q_{\text{eff}}(\cdot|\cdot)$. However, our derivation is equally valid for a continuous output alphabet.

It is natural to consider the conditional entropy-per-bit of the interpolating model. With an abuse of notation we denote by \mathbb{E}_C the expectation both with respect to the code ensemble and with respect to the l_i 's. We define

$$h_n(s) \equiv \frac{1}{n} \mathbb{E}_C H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}}, \underline{Z}) = \quad (6.8)$$

$$\begin{aligned} &= \frac{1}{n} \mathbb{E}_C \mathbb{E}_{y,z} \log_2 \mathcal{Z}(\underline{y}, \hat{\underline{y}}, \underline{z}) - \frac{s}{P'(1)} \sum_y Q_C(y|0) \log_2 Q_C(y|0) - \\ &\quad - \sum_y Q_V(y|0) \log_2 Q_V(y|0) - (\gamma - s) \sum_z Q_{\text{eff}}(z|0) \log_2 Q_{\text{eff}}(z|0). \end{aligned} \quad (6.9)$$

Notice that $h_n(s)$ depends implicitly upon the distribution of U . In passing from Eq. (6.8) to Eq. (6.9) we used the symmetry condition for the two channels and assumed the transmitted message to be the all-zero codeword.

It is easy to compute the conditional entropy (6.8) at the extremes of the interpolation interval. For $s = \gamma$ we recover the original probability distribution (3.3) and the conditional entropy h_n . When $s = 0$, on the other hand, the factor graph no longer contains function nodes of degree larger than one and the partition function can be computed explicitly. Therefore we have

$$h_n(\gamma) = h_n \quad (6.10)$$

$$h_n(0) = \mathbb{E}_{y,z} \mathbb{E}_l \log_2 \left[\sum_x \frac{Q_V(y|x)}{Q_V(y|0)} \prod_{\alpha=1}^l Q_{\text{eff}}(z_\alpha|x) \right] - \gamma \sum_z Q_{\text{eff}}(z|0) \log_2 Q_{\text{eff}}(z|0). \quad (6.11)$$

where l is a poissonian variable with parameter γ . As anticipated, $h_n(0)$ can be expressed uniquely in terms of the distribution of the U variables. Using Eq. (6.7), we get

$$h_n(0) = \mathbb{E}_y \mathbb{E}_l \mathbb{E}_{\{u_i\}} \log_2 \left[\sum_x \frac{Q_V(y|x)}{Q_V(y|0)} \prod_{\alpha=1}^l P_{u_i}(x) \right] + \gamma \mathbb{E}_u \log_2(1 + e^{-2u}). \quad (6.12)$$

The next step is also very natural. Since we are interested in estimating h_n , we write

$$h_n = h_n(0) + \int_0^\gamma \frac{dh_n}{ds}(s) ds. \quad (6.13)$$

A straightforward calculation yields:

$$\begin{aligned} \frac{dh_n}{ds}(s) &= \sum_k \frac{P_k}{P'(1)} \frac{1}{n^k} \sum_{i_1 \dots i_k} \mathbb{E}_{\hat{y}} \mathbb{E}_s \log_2 \left\langle \frac{Q_C(\hat{y}|x_{i_1} \oplus \dots \oplus x_{i_k})}{Q_C(y|0)} \right\rangle_s - \\ &\quad - \frac{1}{n} \sum_{i \in \mathcal{V}} \mathbb{E}_z \mathbb{E}_s \log_2 \left\langle \frac{Q_{\text{eff}}(z|x_i)}{Q_{\text{eff}}(z|0)} \right\rangle_s, \end{aligned} \quad (6.14)$$

where the average $\langle \cdot \rangle_s$ is taken with respect to the interpolating probability distribution (6.6). The details of this computation are reported in App. C. Of course the right-hand side of Eq. (6.14) is still quite hard to evaluate because the averages $\langle \cdot \rangle_s$ are complicated objects. We shall therefore approximate them by much simpler averages under which the x_i 's are independent and have log-likelihoods which are distributed as V -variables. More precisely we define

$$\begin{aligned} \frac{d\hat{h}_n}{dt}(s) &= \sum_k \frac{P_k}{P'(1)} \frac{1}{n^k} \sum_{i_1 \dots i_k} \mathbb{E}_{\hat{y}} \mathbb{E}_v \log_2 \left\{ \sum_{x_{i_1} \dots x_{i_k}} \frac{Q_{\mathcal{C}}(\hat{y} | x_{i_1} \oplus \dots \oplus x_{i_k})}{Q_{\mathcal{C}}(y|0)} \prod_{j=1}^k P_{v_j}(x_{i_j}) \right\} - \\ &\quad - \frac{1}{n} \sum_{i \in \mathcal{V}} \mathbb{E}_z \mathbb{E}_v \log_2 \left\{ \sum_{x_i} \frac{Q_{\text{eff}}(z | x_i)}{Q_{\text{eff}}(z|0)} P_v(x_i) \right\}. \end{aligned} \quad (6.15)$$

Notice that in fact this expression does not depend upon s . Summing and subtracting it to Eq. (6.13), and after a few algebraic manipulations, we obtain

$$h_n = \phi_V + \int_0^\gamma \left[\frac{dh_n}{dt}(s) - \frac{d\hat{h}_n}{dt}(s) \right] ds \equiv \phi_V + \int_0^\gamma R_n(s) ds. \quad (6.16)$$

The proof is completed by showing that $R_n(s) \geq 0$ for any $s \in [0, \gamma]$. This calculation is reported in App. D. \square

7 Proof for multi-Poisson ensembles

The proof for multi-Poisson ensembles follows the same strategy as for Poisson ensembles. The only difference is that the interpolating system is obviously more complex.

Given the ensemble parameters (n, γ, Λ, P) , and defining as in Sec. 2.3, $t_{\max} \equiv \lfloor \Lambda'(1)/\gamma \rfloor - 1$, we introduce an interpolating ensemble for each pair (t_*, s) , with $t_* \in \{0, \dots, t_{\max} - 1\}$ and $s \in [0, \gamma]$ as follows. The first t_* rounds (i.e. $t = 0, \dots, t_* - 1$ in Definition 4) in the graph construction are the same as for the original multi-Poisson graph.

Next, during round t_* , $m_k^{(t_*)}$ is drawn from a Poisson distribution with mean $nsP_k/P'(1)$ (instead of $n\gamma P_k/P'(1)$), and $m_k^{(t_*)}$ right (check) nodes of degree k are added for each $k = 2, \dots, k_{\max}$. The neighbors of each of this check node are i.i.d. random variables in \mathcal{V} with distribution $w_i(t) = [\mathbf{d}_i(t_*)]_+ / (\sum_j [\mathbf{d}_j(t_*)]_+)$. In order to compensate for the smaller number of right nodes, an integer $l_i(t_*)$ is drawn for each $i \in \mathcal{V}$ from a Poisson distribution with mean $n(\gamma - s)w_i(t)$. As in the previous Section, this means that $l_i(t_*)$ repetitions of the bit x_i will be transmitted through an effective channel with transition probability $Q_{\text{eff}}(\cdot | \cdot)$. Finally, the number of free sockets is updated by setting $\mathbf{d}_i(t_* + 1) = \mathbf{d}_i(t_*) - \Delta_i(t_*) - l_i(t_*)$, where $\Delta_i(t_*)$ is the number of times the node i has been chosen as neighbor of a right node in this round.

During rounds $t = t_* + 1, \dots, t_{\max} - 1$, no right node is added to the factor graph. On the other hand, for each $i \in \mathcal{V}$ a random integer $l_i(t)$ is drawn from a Poisson distribution with mean $n\gamma w_i(t_*)$. This means that the bit x_i will be further retransmitted l_i times through the effective channel. Furthermore the number of free sockets is updated at the end of each round by setting $\mathbf{d}_i(t + 1) = \mathbf{d}_i(t) - l_i(t)$.

This completes the definition of the Tanner graph ensemble. We denote by $\mathcal{C}_{(t_*, s)}$ the set of function nodes and by $l_i = \sum_{t=t_*}^{t_{\max}-1} l_i(t)$ the total number of times the bit x_i is transmitted

through the effective channel. The overall conditional distribution of the channel input given the output has the same form (6.6) as for the simple Poisson ensemble.

The overall number of function nodes in the (t_*, s) ensemble is easily seen to be a Poisson variable with mean $n(\gamma t_* + s)/P'(1)$. In fact we add $\text{Poisson}(n\gamma/P'(1))$ function nodes during each of the first t_* rounds, and $\text{Poisson}(ns/P'(1))$ during the $(t_* + 1)$ -th round. Analogously, the overall number of repetitions $\sum_i l_i$ is a Poisson variable with mean $n[\gamma(t_{\max} - t_* - 1) + \gamma - s]$. Using these remarks, we get the expression

$$h_n(t_*, s) = \frac{1}{n} \mathbb{E}_C H_n(\underline{X}|\underline{Y}, \hat{\underline{Y}}, \underline{Z}) = \quad (7.1)$$

$$\begin{aligned} &= \frac{1}{n} \mathbb{E}_C \mathbb{E}_{y,z} \log_2 \mathcal{Z}(\underline{y}, \hat{\underline{y}}, \underline{z}) - \frac{1}{P'(1)} (\gamma t_* + s) \sum_y Q_C(y|0) \log_2 Q_C(y|0) - \\ &\quad - \sum_y Q_V(y|0) \log_2 Q_V(y|0) - (\gamma(t_{\max} - t_*) - s) \sum_z Q_{\text{eff}}(z|0) \log_2 Q_{\text{eff}}(z|0). \end{aligned} \quad (7.2)$$

Let us now look at a few particular cases of the interpolating system. For $(t_* = t_{\max} - 1, s = \gamma)$ we recover the original multi-Poisson ensemble. Moreover, for any $t_* \in \{0, \dots, t_{\max} - 2\}$ the ensembles $(t_*, s = \gamma)$ and $(t_* + 1, s = 0)$ are identically distributed. Finally, when $(t_* = 0, s = 0)$ the set of factor nodes is empty with probability one, and the resulting coding scheme is just an irregular repetition code (bit x_i being repeated l_i times) used over the channel $Q_{\text{eff}}(\cdot|\cdot)$. If we denote by $h_n(t_*, s)$ the expected entropy per bit in the interpolating ensemble, we get

$$h_n(t_{\max} - 1, \gamma) = h_n \quad (7.3)$$

$$h_n(0, 0) = \mathbb{E}_{y,z} \mathbb{E}_l \log_2 \left[\sum_x \frac{Q_V(y|x)}{Q_V(y|0)} \prod_{\alpha=1}^l Q_{\text{eff}}(z_\alpha|x) \right] - \gamma t_{\max} \sum_z Q_{\text{eff}}(z|0) \log_2 Q_{\text{eff}}(z|0). \quad (7.4)$$

Here the expectations on $y, \{z_\alpha\}$ are taken with respect to the distributions $Q_V(y|0), Q_{\text{eff}}(z|0)$, while l is distributed according to the expected degree profile $\Lambda_l^{(n,\gamma)}$. Moreover, we used the fact that $\sum_l \Lambda_l^{(n,\gamma)} l = n\gamma t_{\max}$. Finally, as in the previous Section (cf. Eq. (6.12)) the conditional entropy $h_n(0, 0)$ depends upon the effective channel transition probability only through the distribution of the log-likelihood U .

The next step consist in writing

$$h_n(t_*, \gamma) = h_n(t_*, 0) + \int_0^\gamma \frac{dh_n}{ds}(t_*, s) ds, \quad (7.5)$$

for $t_* \in \{0, \dots, t_{\max} - 1\}$. Using the fact that $h_n(t_*, \gamma) = h_n(t_* + 1, 0)$, this implies

$$h_n = h_n(0, 0) + \sum_{t_*=0}^{t_{\max}-1} \int_0^\gamma \frac{dh_n}{ds}(t_*, s) ds. \quad (7.6)$$

The derivative with respect to the interpolation parameter is similar the simple Poisson case:

$$\begin{aligned} \frac{dh_n}{ds}(t_*, s) &= \sum_k \frac{P_k}{P'(1)} \sum_{i_1 \dots i_k} \mathbb{E}_{\hat{y}} \mathbb{E}^{\{i_1 \dots i_k\}} w_{i_1}(t_*) \dots w_{i_k}(t_*) \log_2 \left\langle \frac{Q_C(\hat{y}|x_{i_1} \oplus \dots \oplus x_{i_k})}{Q_C(y|0)} \right\rangle_{(t_*, s)} - \\ &\quad - \sum_{i \in \mathcal{V}} \mathbb{E}_z \mathbb{E}^{\{i\}} w_i(t_*) \log_2 \left\langle \frac{Q_{\text{eff}}(z|x_i)}{Q_{\text{eff}}(z|0)} \right\rangle_{(t_*, s)} + \varphi(n; t_*, s), \end{aligned} \quad (7.7)$$

where $\langle \cdot \rangle_{(t_*, s)}$ denotes expectation with respect to the conditional distribution (6.6) appropriate for the (t_*, s) ensemble and the expectations $\mathbb{E}^{\{i_1 \dots i_k\}}$ are defined in App. E. Moreover, the term $\varphi(n; t_*, s)$ has the following properties.

- A. $|\varphi(n; t_*, s)| \leq C_1$ for some constant C_1 which depends uniquely upon the ensemble parameters Λ , P and γ .
- B. $\varphi(n; t_*, s) \leq C_2(t_*, s) \sqrt{(\log n)^3/n}$ for some function $C_2(t_*, s)$ which does not depend upon n .

We refer to App. E for the details of this computation. Notice that the equivalent expression for Poisson codes, cf. Eq. (6.14), is recovered by setting $w_i(t_*) = 1/n$ and dropping the correction $\varphi(\cdot)$.

Finally, we introduce an ‘approximation’ $\frac{d\hat{h}_n}{ds}(t_*, s)$ to Eq. (7.7) analogous to Eq. (6.15). More precisely, we replace the expectations $\langle \cdot \rangle_{(t_*, s)}$ with expectations over product measures of the form $\prod_i P_{v_i}(x_i)$, the ensemble averages $\mathbb{E}^{\{i_1 \dots i_k\}}$ with averages over i.i.d. V_i ’s, and we drop the remainder $\varphi(\cdot)$. Using Eq. (7.6) and after rearranging various terms (the relation $\Lambda^{(\gamma, n)'}(1) = \gamma t_{\max}$ is useful here) we end up with

$$\begin{aligned} h_n &= \phi_V(\Lambda^{(\gamma, n)}, P) + \sum_{t_*=0}^{t_{\max}-1} \int_0^\gamma \left[\frac{dh_n}{dt}(t_*; s) - \frac{d\hat{h}_n}{dt}(t_*; s) \right] ds = \\ &= \phi_V(\Lambda^{(\gamma, n)}, P) + \sum_{t_*=0}^{t_{\max}-1} \int_0^\gamma R_n(t_*, s) ds + o_n(1), \end{aligned} \quad (7.8)$$

where the second inequality follows by applying the dominated convergence theorem to $\varphi(\cdot)$. The proof is finished by showing that $R_n(t_*, s)$ is non-negative for any $t_* \in \{0, \dots, t_{\max} - 1\}$ and $s \in [0, \gamma]$. this calculation is very similar to the simple Poisson case, and is discussed in App. F. \square

8 Proof for standard ensembles

We proved points 1 and 2 of Theorem 2 directly. We will now show that 2 implies the lower bound for standard ensembles (point 3) which is the practically more relevant case.

The idea is that the standard ensemble (n, Λ, P) is indeed ‘very close’ to the multi-Poisson ensemble (n, γ, Λ, P) for small γ . In order to implement this idea, we state a preliminary result here.

Lemma 5 *Let C_1 and C_2 be two codes with the same blocklength n from any of the ensembles defined in Section 2 (the ensemble does not need to be the same). Assume they are used to communicate through the same noisy channel and let $H_1(X|Y_1)$, $H_2(X|Y_2)$ be the corresponding conditional entropies. If C_1 can be obtained from C_2 with n_r rewirings, then $|H_1(X|Y_1) - H_2(X|Y_2)| \leq n_r$.*

Proof: Recall that a rewiring was defined as either the removal or the addition of a function node to the Tanner graph. We already proved (cf. Eqs. (5.5) to (5.8) and relative discussion) that the introduction of a single function node induces a change in the conditional entropy which is smaller than 1 bit. \square

Let now $0 < \gamma < 1$, and consider a pair of Tanner graphs $(\mathcal{G}_s, \mathcal{G}_{\text{mP}})$ distributed according to the coupling in Lemma 1. In particular, the marginal distribution of $\mathcal{G}_s \stackrel{\text{d}}{=} (n, \Lambda, P)$ and $\mathcal{G}_{\text{mP}} \stackrel{\text{d}}{=} (n, \Lambda, P, \gamma)$. With an abuse of notation, denote by $h_n(\mathcal{G}_s)$ and $h_n(\mathcal{G}_{\text{mP}})$ the corresponding conditional entropy *per bit* and by $h_n, h_n^{(\gamma)}$ their ensemble averages. From Lemmas 1 and 5 it follows that

$$\lim_{n \rightarrow \infty} \mathbb{P}[|h_n(\mathcal{G}_{\text{mP}}) - h_n(\mathcal{G}_s)| \geq A\gamma^b] = 0, \quad (8.1)$$

where the constants A and b are as in Lemma 1. Therefore

$$|h_n^{(\gamma)} - h_n| = |\mathbb{E}[h_n(\mathcal{G}_{\text{mP}}) - h_n(\mathcal{G}_s)]| \leq \mathbb{E}|h_n(\mathcal{G}_{\text{mP}}) - h_n(\mathcal{G}_s)| \leq A\gamma^b + o_n(1), \quad (8.2)$$

where the last inequality follows from Eq. (8.1) together with the remark that $h_n(\mathcal{G}) \leq 1$ for any code. By taking the large blocklength limit, we get

$$\liminf_{n \rightarrow \infty} h_n \geq \liminf_{n \rightarrow \infty} h_n^{(\gamma)} - A\gamma^b \geq \phi_V(\Lambda^{(\gamma)}, P) - A\gamma^b, \quad (8.3)$$

where the last inequality follows from our lower bound on multi-Poisson ensembles, Eq. (6.4). Next we notice that $\phi_V(\Lambda, P)$ is a continuous function of Λ (with respect to the total variation distance, see App. A for a definition), once the degree distribution P , and the V -variables distribution have been fixed. Moreover by Corollary 1 $\lim_{\gamma \rightarrow 0} \Lambda^{(\gamma)} = \Lambda$ in total variation distance sense. Therefore we can obtain the thesis (6.5) by taking the $\gamma \rightarrow 0$ limit in the last expression. \square

9 Examples and applications

The optimization problem in Eq. (6.3) is, in general, rather difficult. Nevertheless, one can easily obtain sub-optimal bounds on the entropy h_n , by cleverly choosing the distribution of the V -variables to be used in Eq. (6.2). Moreover, bounds can be optimized through density evolution. Although a complete discussion of the optimization problem is beyond the scope of this paper, a rather simple approach, cf. Sec. 9.5 and Tab. 2, already gives very good results (indeed we believe them to be optimal).

Our main focus will be here on standard (n, Λ, P) ensembles. As in our original definition, cf. Eq. (2.1), we shall generically consider the case in which $\Lambda_0 = \Lambda_1 = 0$ (no degree 0 or degree 1 variable nodes). However, most of the arguments can be adapted to Poisson ensembles too. On the other hand, we shall *always* assume $P_0 = P_1 = 0$ (no degree 0 or degree 1 check nodes).

Throughout this Section, we shall use the notation $h = \liminf_{n \rightarrow \infty} h_n$ for the asymptotic conditional entropy per symbol.

9.1 Shannon threshold

Assume $V = 0$ with probability 1, and therefore, from Eq. (6.1), $U = 0$ with probability 1. Plugging these distributions into Eq. (6.2) we get

$$\phi_V = -\frac{\Lambda'(1)}{P'(1)} + \mathbb{E}_y \log_2 \left[\sum_x \frac{Q_V(y|x)}{Q_V(y|0)} \right] + \frac{\Lambda'(1)}{P'(1)} \mathbb{E}_{\hat{y}} \log_2 \left[\sum_x \frac{Q_C(\hat{y}|x)}{Q_C(\hat{y}|0)} \right]. \quad (9.1)$$

Using Theorem 2, and after a few manipulations, we get

$$h \geq 1 - \frac{\mathfrak{C}(Q)}{r_{\text{des}}}, \quad \text{for the LDGM}(n, \Lambda, P) \text{ ensemble}, \quad (9.2)$$

$$h \geq r_{\text{des}} - \mathfrak{C}(Q), \quad \text{for the LDPC}(n, \Lambda, P) \text{ ensemble}. \quad (9.3)$$

Where we denoted $\mathfrak{C}(Q)$ the capacity of the BIOS channel with transition probability $Q(y|x)$:

$$\mathfrak{C}(Q) = 1 - \sum_{y \in \mathcal{A}} Q(y|0) \log_2 \left[\sum_x \frac{Q(y|x)}{Q(y|0)} \right]. \quad (9.4)$$

In other words reliable communication (which requires $h_n \rightarrow 0$ as $n \rightarrow \infty$) can be achieved only if the design rate is smaller than channel capacity. For the LDGM ensemble this statement is equivalent to the converse of channel coding theorem, because r_{des} is concentrated around the actual rate. This is the case also for standard LDPC ensembles with no degree 0 or degree 1 variable nodes [36].

For general LDPC ensembles, Eq. (9.3) is slightly weaker than the channel coding theorem because the actual rate can be larger than the design rate. However, as shown in the next Sections, the bound can be easily improved changing the distribution of V .

Of course this results could have been derived from information-theoretic arguments. However it is nice to see that it is indeed contained in Theorem 2.

9.2 Non-negativity of the entropy

Let us consider the opposite limit: $V = +\infty$ with probability 1, and distinguish two cases:

LDPC(n, Λ, P) ensemble. As a consequence of Eq. (6.1), also $U = +\infty$ with probability 1. Using Theorem 2 and Eq. (6.2), it is easy to obtain

$$h \geq \Lambda_0 H_Q(X|Y), \quad (9.5)$$

where $H_Q(X|Y) = 1 - \mathfrak{C}(Q)$ is the relative entropy for a single bit transmitted across the channel $Q(y|x)$. The interpretation of Eq. (9.5) is straightforward. Typically, a fraction Λ_0 of the variable nodes have degree zero. The relative entropy (5.1) is lower-bounded by the entropy of these variables.

LDGM(n, Λ, P) ensemble. Equation (6.1) implies that U is distributed as the log-likelihoods $J_a = (1/2) \log Q(y|0)/Q(y|1)$, see Eq. (3.9). It is easy then to evaluate the bound:

$$h \geq \mathbb{E}_l \mathbb{E}_y \log_2 \left[\sum_x \prod_{i=1}^l \frac{Q(y_i|x)}{\sum_{x'} Q(y_i|x')} \right]. \quad (9.6)$$

The meaning of this inequality is, once again, quite clear:

$$H_n(\underline{X}|\hat{Y}) \geq \sum_{i=1}^n H_n(X_i|\underline{X}^{(i)}, \hat{Y}) \geq \sum_{i=1}^n H_n(X_i|\{x_j=0 \forall j \neq i\}; \hat{Y}), \quad (9.7)$$

where we introduced $\underline{X}^{(i)} \equiv \{X_j\}_{j \neq i}$. The above inequalities are consequences of the entropy chain rule and of the fact that conditioning reduces entropy. Taking the expectation with respect to the code ensemble and letting $n \rightarrow \infty$ yields (9.6).

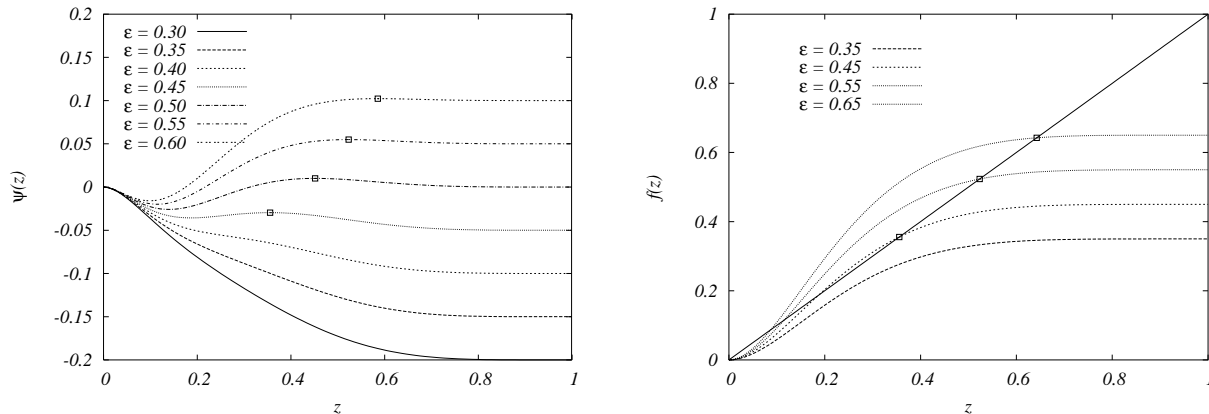


Figure 4: Left frame: trial entropy ϕ_V on the erasure channel $\text{BEC}(\epsilon)$ as a function of the parameter z characterizing the V -distribution, cf. Eq. (9.10). Here we consider the $\text{LDPC}(n, \Lambda, P)$ ensemble with $\Lambda(x) = x^3$ and $P(x) = x^6$ (i.e. a regular (3, 6) Gallager code). A square (\square) marks the high bit-error-rate local maximum $z_{\text{bad}}(\epsilon)$. Right frame: graphical representation of the equation $z = f(z)$ yielding the local extrema of the trial entropy, cf. Eq. (9.11).

9.3 Binary Erasure Channel

Let us define the binary erasure channel $\text{BEC}(\epsilon)$. We have $\mathcal{A} = \{0, 1, *\}$ and $Q(0|0) = Q(1|1) = 1 - \epsilon$, $Q(*|0) = Q(*|1) = \epsilon$. Since the log-likelihoods (3.9) take values in $\{0, +\infty\}$ it is natural to assume the same property to hold for the variables U and V . We denote by z (\hat{z}) the probability for V (U) to be 0. As in the previous example we distinguish two cases

LDPC(n, Λ, P) ensemble. Equation (6.1) yields

$$\hat{z} = 1 - \rho(1 - z). \quad (9.8)$$

It is easy to show that Eq. (6.5) implies the bound

$$h \geq \sup_{z \in [0,1]} \phi(z, 1 - \rho(1 - z)), \quad (9.9)$$

where

$$\phi(z, \hat{z}) = \Lambda'(1)z(1 - \hat{z}) - \frac{\Lambda'(1)}{P'(1)}[1 - P(1 - z)] + \epsilon \Lambda(\hat{z}). \quad (9.10)$$

Notice that $\phi(z, 1 - \rho(1 - z)) \equiv \psi(z)$ is a smooth function of $z \in [0, 1]$. Therefore the sup in Eq. (9.9) is achieved either in $z = 0, 1$ or for a $z \in (0, 1)$ such that $\psi(z)$ stationary. It is easy to realize that the stationarity condition is equivalent to the equations

$$z = \epsilon \lambda(\hat{z}), \quad \hat{z} = 1 - \rho(1 - z). \quad (9.11)$$

The reader will recognize the fixed point conditions for BP decoding [29].

Let us consider a specific example: the (3, 6) regular Gallager code. We have $\Lambda(x) = x^3$ and $P(x) = x^6$: $P(x)$ is convex for any $x \in \mathbb{R}$ and therefore Theorem 2 applies. The design rate is $r_{\text{des}} = 1/2$. In Fig. 4 we show the function $\psi(z)$ for several values of the erasure probability.

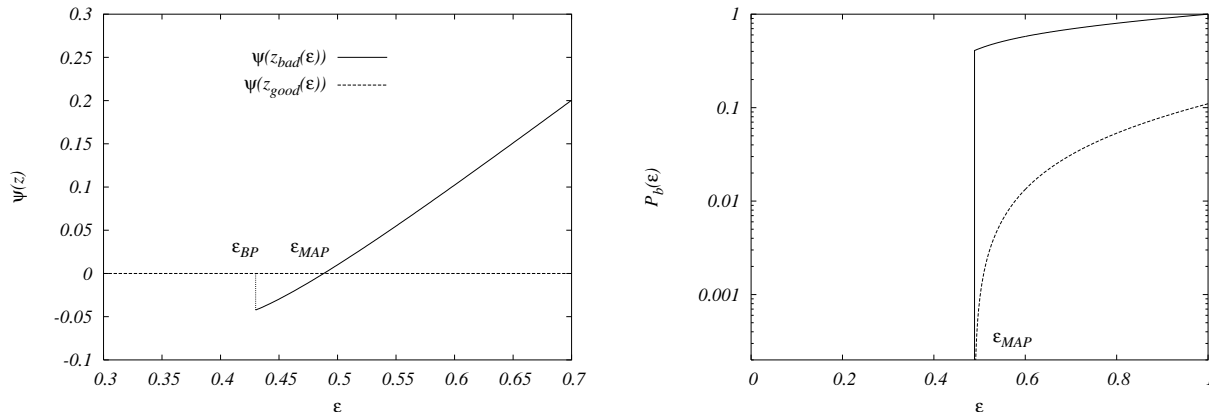


Figure 5: Left frame: trial entropy evaluated at its local maxima $z_{\text{good}}(\epsilon)$ and $z_{\text{bad}}(\epsilon)$, as a function of the erasure probability ϵ (notice that $\psi(z_{\text{good}}(\epsilon)) = 0$ identically). Right frame: bit error rate under maximum likelihood decoding. The dashed line is the lower bound obtained from Fano inequality. The continuous curve is the conjectured exact result.

In the right frame we present the function $f(z) = \epsilon\lambda(1 - \rho(1 - z))$ for some of these values. At small ϵ , the conditions (9.11) have a unique solution at $z_{\text{good}}(\epsilon) = 0$, and $\psi(z)$ has its unique local maximum there. The corresponding lower bound on the conditional entropy is $\psi(z_{\text{good}}(\epsilon)) = 0$. For $\epsilon > \epsilon_{\text{BP}} \approx 0.4294398$ a secondary maximum $z_{\text{bad}}(\epsilon)$ appears. Density evolution converges to $z_{\text{bad}}(\epsilon)$ and therefore this fixed point controls the bit error rate under BP decoding. For $\epsilon_{\text{BP}} < \epsilon < \epsilon_{\text{MAP}} \approx 0.4881508$, $\psi(z_{\text{bad}}) < \psi(z_{\text{good}})$ and therefore this local maximum is irrelevant for MAP decoding. Above ϵ_{MAP} , $\psi(z_{\text{bad}}) > \psi(z_{\text{good}})$ and therefore $z_{\text{bad}}(\epsilon)$ controls the properties of MAP decoding too.

In Fig. 5, left frame, we reproduce the function $\psi(z_{\text{bad}}(\epsilon))$ as a function of ϵ . Fano inequality, cf. Lemma 4, can be used for obtaining lower bounds on block and bit error rates in terms of the quantity $\max\{\psi(z_{\text{good}}(\epsilon)), \psi(z_{\text{bad}}(\epsilon))\}$. The result for our running example is presented in Fig. 5, right frame. It is evident that the result is not tight because of the sub-optimality of Fano inequality. For instance, in the $\epsilon \rightarrow 1$ limit, $\psi(z_{\text{bad}}(\epsilon))$ yields the lower bound $h_n \geq 1/2$. This result is easily understood: since no bit has been received, all the $2^{n/2}$ codewords are equiprobable. On the other hand Fano inequality yields a poor $P_b(\epsilon = 1) \geq 0.11003$.

A better (albeit non-rigorous) estimate is provided by the following recipe. Notice that BP decoding yields (in the large blocklength limit)

$$P_b^{\text{BP}}(\epsilon) = \begin{cases} \epsilon\Lambda(z_{\text{good}}(\epsilon)) & \text{for } \epsilon < \epsilon_{\text{BP}}, \\ \epsilon\Lambda(z_{\text{bad}}(\epsilon)) & \text{for } \epsilon \geq \epsilon_{\text{BP}}. \end{cases} \quad (9.12)$$

Our prescription consists in taking

$$P_b^{\text{MAP}}(\epsilon) = \begin{cases} \epsilon\Lambda(z_{\text{good}}(\epsilon)) & \text{for } \epsilon < \epsilon_{\text{MAP}}, \\ \epsilon\Lambda(z_{\text{bad}}(\epsilon)) & \text{for } \epsilon \geq \epsilon_{\text{MAP}}. \end{cases} \quad (9.13)$$

In other words, BP is asymptotically optimal except in the interval $[\epsilon_{\text{BP}}, \epsilon_{\text{MAP}}]$. Generalizations and heuristic justification of this recipe will be provided in the next Sections. The resulting curve for our running example is reported in Fig. 5, right frame.

k	γ	ϵ_{BP}	$\epsilon_{\text{MAP: New UB}}$	Gallager UB	Gallager LB
2	4	1/3	1/3	*	*
3	6	0.4294398	0.4881508	0.4999118	0.4833615
4	8	0.3834465	0.4977409	0.4999118	0.4967571
5	10	0.3415500	0.4994859	0.4999997	0.4992593
6	12	0.3074623	0.4998757	0.5000000	0.4998207

Table 1: Maximum a posteriori probability and belief propagation thresholds for several ensembles of the form LDPC(n, γ, x^k) with $\gamma = (1 - r_{\text{des}})k$ and $r_{\text{des}} = 1/4$. For the MAP threshold we compare several different thresholds: ‘New UB’ is the upper bound derived in this paper; ‘Gallager UB’ is Gallager lower bound as generalized in Ref. [9]; ‘Gallager LB’ is the upper bound derived using Gallager’s technique, as applied in Ref. [11].

The analysis of this simple example uncovers the existence of three distinct regimes: (i) A low noise regime, $\epsilon < \epsilon_{\text{BP}}$: both BP and MAP decoding are effective in this case: the bit error rate vanishes in the large blocklength limit; (ii) An intermediate noise regime, $\epsilon_{\text{BP}} < \epsilon < \epsilon_{\text{MAP}}$. Only MAP decoding can produce vanishing error rates here. (iii) An high noise regime, $\epsilon_{\text{MAP}} < \epsilon$. The bit error rate under MAP decoding is bounded from below. In Table 1 we report the values of ϵ_{BP} and ϵ_{MAP} for a few ensembles LDPC(n, Λ, P) with $\Lambda(x) = x^l$, $P(x) = x^k$ and $r_{\text{des}} = 1/2$. As we shall discuss below, this pattern is quite general.

LDGM(n, Λ, P) ensemble. It is interesting to look at the differences between LDGM and LDPC ensembles within the BEC context. The requirement (6.1) implies

$$\hat{z} = 1 - (1 - \epsilon)\rho(1 - z). \quad (9.14)$$

Applying Theorem 2 we get the bound

$$h \geq \sup_{z \in [0,1]} \phi(z, 1 - (1 - \epsilon)\rho(1 - z)), \quad (9.15)$$

where, with a slight abuse of notation, we defined

$$\phi(z, \hat{z}) = \Lambda'(1)z(1 - \hat{z}) - \frac{\Lambda'(1)(1 - \epsilon)}{P'(1)}[1 - P(1 - z)] + \Lambda(\hat{z}). \quad (9.16)$$

As in the LDPC case we look at the stationarity condition of the function $\psi(z) \equiv \phi(z, 1 - (1 - \epsilon)\rho(1 - z))$. Elementary calculus yields the couple of equations

$$z = \lambda(\hat{z}), \quad \hat{z} = 1 - (1 - \epsilon)\rho(1 - z), \quad (9.17)$$

that, once again, coincide with the fixed point conditions for BP decoding. These equations have a noise-independent solution $z_{\text{bad}}(\epsilon) = 1$ (implying $\hat{z} = 1$ because of Eq. (9.14)). Theorem 2 yields $h \geq 1 - \mathcal{C}(\epsilon)/r_{\text{des}}$, with $\mathcal{C}(\epsilon) = 1 - \epsilon$ the channel capacity: we recover in this context the simple lower bound of Sec. 9.1.

A better understanding of the peculiarities LDGM ensembles is obtained by looking at a particular example. Consider, for instance, $\Lambda(x) = x^8$ and $P(x) = x^4$ which corresponds to a design rate $r_{\text{des}} = 1/2$. Theorem 2 applies because $P(x)$ is convex on \mathbb{R} . In Fig. 6, left frame, we plot the function $\psi(z)$ for several values of the erasure probability ϵ . It is clear

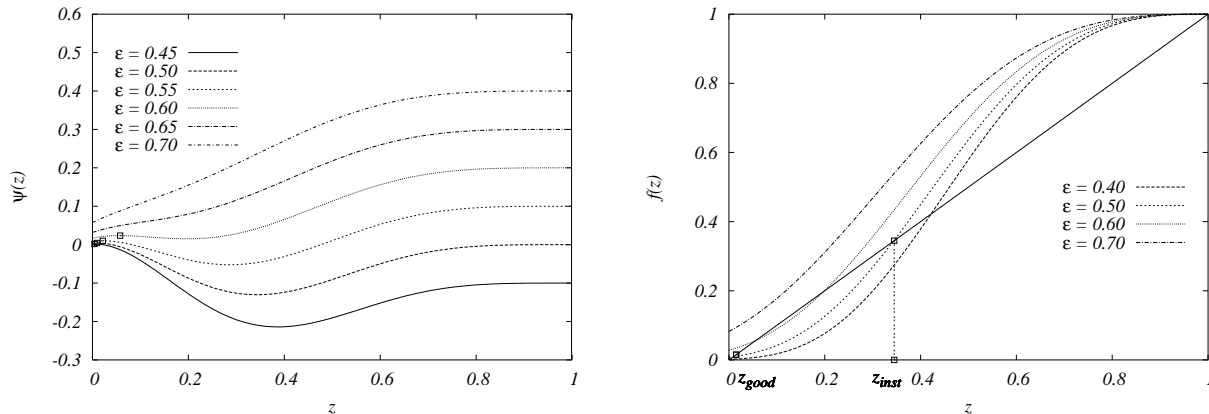


Figure 6: Left frame: trial entropy for the LDGM(n, Λ, P) ensemble on the BEC(ϵ) channel. Here $\Lambda(x) = x^8$ and $P(x) = x^4$ (the design rate is $r_{\text{des}} = 1/2$). A square (\square) marks the small- P_b stationary point $z_{\text{good}}(\epsilon)$. Right frame: graphical representation of the stationarity condition $z = f(z)$.

that $z_{\text{bad}} = 1$ is always a local maximum. A second local maximum $z_{\text{good}}(\epsilon)$ appears when the erasure probability becomes smaller than $\epsilon_{\text{BP}} \approx 0.6165534$. The extremum at $z_{\text{good}}(\epsilon)$ becomes a global maximum for $\epsilon < \epsilon_{\text{MAP}}$, with $\epsilon_{\text{MAP}} \approx 0.5022591$. In Fig. 6, right frame, we reproduce the function $f(z) = \lambda(1 - (1 - \epsilon)\rho(1 - z))$ in terms of which the stationarity condition (9.17) reads $z = f(z)$. We also mark the solutions $z_{\text{good}}(\epsilon)$ (corresponding to a local maximum of $\psi(z)$) and $z_{\text{inst}}(\epsilon)$ (corresponding to a local minimum of $\psi(z)$).

The interpretation of these results is straightforward. Maximum likelihood decoding is controlled by the stationary point $z_{\text{bad}} = 1$ for $\epsilon > \epsilon_{\text{MAP}}$. In this regime the lower bound (9.16) yields the same conditional entropy as for the random code ensemble. We expect the bit error rate in this regime to be $P_b(\epsilon) = 1/2$. At low noise ($\epsilon < \epsilon_{\text{MAP}}$) the fixed point $z_{\text{good}}(\epsilon)$ controls the MAP performances. Analogously to what argued in the previous Subsection, we expect this to imply to a bit error rate $P_b(\epsilon) = \Lambda(1 - (1 - \epsilon)\rho(1 - z_{\text{good}}(\epsilon)))$.

As for BP decoding, it has a unique fixed point $z_{\text{bad}} = 1$ for $\epsilon > \epsilon_{\text{BP}}$. This corresponds to a high bit error rate $P_b(\epsilon) = 1/2$. A second, locally stable, fixed point appears at ϵ_{BP} . If BP is initialized using only erased messages (as is usually done), all the messages remain erased (BP does not know where to start from). The same remains true is a small number of non-erased (correct) messages is introduced: density evolution is still controlled by the z_{bad} fixed point. If however the initial conditions contains a large enough fraction (namely, larger than $1 - z_{\text{inst}}$) of correct messages, the small- P_b fixed point z_{good} is eventually reached.

Let us finally notice that the present results can be shown to be consistent with the ones of Refs. [37, 38].

9.4 General channel: a simple minded bound

The previous Section suggests a simple bound for the LDPC(n, Λ, P) ensemble on general BIOS channel. Take, as for the BEC case, $V = 0$ with probability z and $= \infty$ with probability $1 - z$, while $U = 0$ with probability \hat{z} and $= \infty$ with probability $1 - \hat{z}$. These conditions are consistent with the admissibility requirement (6.1) if

$$\hat{z} = 1 - \rho(1 - z). \quad (9.18)$$

Plugging into Eq. (6.2) we get a bound of the same form as for the BEC, cf. Eq. (9.9) with

$$\phi(z, \hat{z}) = \Lambda'(1)z(1 - \hat{z}) - \frac{\Lambda'(1)}{P'(1)}[1 - P(1 - z)] + [1 - \mathfrak{C}(Q)]\Lambda(\hat{z}). \quad (9.19)$$

Passing from the BEC to a general BIOS, amounts, under this simple ansatz, to substituting $1 - \mathfrak{C}(Q)$ to ϵ .

9.5 General channel: optimizing the bound

We saw in Sec. 9.3 that, for the BEC, stationary points of the trial entropy function correspond to fixed points of the density-evolution equations. This fact is indeed quite general and holds for a general BIOS channel.

In order to discuss this point, it is useful to have a concrete representation for the random variables U, V entering in the definition of the trial entropy (6.2). A first possibility is to identify them with the distributions $U(x) \equiv \mathbb{P}[U \leq x]$ and $V(x) \equiv \mathbb{P}[V \leq x]$ as explained in [30]. The distributions are right continuous, non decreasing functions such that $\lim_{x \rightarrow -\infty} A(x) = 0$, and $\lim_{x \rightarrow +\infty} A(x) \leq 1$. Viceversa, to any such function we may associate a well defined random variable. It is convenient to introduce the densities $u(x)$ and $v(x)$ which are formal Radon-Nikodyn derivatives of $U(x)$ and $V(x)$. We also introduce the log-likelihood distributions associated to channel output, cf Eq. (3.9): $J(x) \equiv \mathbb{P}[J \leq x]$ and $H(x) \equiv \mathbb{P}[h \leq x]$. The corresponding densities will be denoted by $j(x)$ and $h(x)$.

The admissibility condition (6.1) translate of course into a condition on the distributions $U(x)$ and $V(x)$. Following once again [30], we express this condition through ‘ G -distributions’. More precisely for any number $x \in (-\infty, +\infty]$, we define $\gamma(x) = (\gamma_1(x), \gamma_2(x)) \in \{\pm 1\} \times [0, +\infty)$ by taking $\gamma_1(x) = \text{sign } x$, and $\gamma_2(x) = -\log |\tanh x|$. We define Γ to be the change-of-measure operator associated to the mapping γ . If X is a random variable with distribution A and (formal) density \mathfrak{a} , $\Gamma(\mathfrak{a})$ is defined to be the density associated to $\gamma(X)$. Despite the notation Γ is defined on distributions, and only formally on densities. The action of Γ is described in detail in [30]. Among the other properties, it has a well defined inverse Γ^{-1} . We can now write the condition implied by (6.1) in a compact form:

$$\mathfrak{u} = \sum_{k=2}^{k_{\max}} \rho_k \Gamma^{-1} \left[\Gamma(\mathfrak{j}) \otimes \Gamma(\mathfrak{v})^{\otimes(k-1)} \right] \equiv \rho_{\mathfrak{j}}(\mathfrak{v}). \quad (9.20)$$

A second concrete representation is obtained by inverting the distributions $U(x), V(x)$. Of course this is not possible unless $U(x), V(x)$ are continuous and increasing. However, we can always define the ‘inverse distributions’:

$$\bar{U} : (0, 1) \rightarrow (-\infty, +\infty] \quad (9.21)$$

$$\xi \mapsto \bar{U}(\xi) \equiv \min\{x \text{ such that } U(x) \geq \xi\}, \quad (9.22)$$

with an analogous definition for $\bar{V}(\xi)$. We introduce analogously the inverse distributions $\bar{J}(\xi)$ and $\bar{H}(\xi)$ for the log-likelihoods J_a and h_i , cf. Eq. (3.9). Notice that, given a real valued random variable X , its inverse distribution $\bar{A}(\xi)$ is non-decreasing and left-continuous. We shall denote by $\bar{\mathcal{A}}$ the space of inverse distributions. Moreover, it has a simple practical meaning: if one is able to sample ξ uniformly in $(0, 1)$, then $\bar{A}(\xi)$ is distributed like X . From this observation it follows that any inverse distribution $\bar{A}(\xi)$ uniquely determines its associated random variable.

We can now re-express the trial entropy (6.2) as a functional over $\overline{\mathcal{A}}$, $\phi = \phi(\overline{\mathbf{U}}, \overline{\mathbf{V}})$ using the above correspondence⁵. After some straightforward computations we get

$$\begin{aligned} \phi &= -\Lambda'(1) \int \log_2[1 + \tanh \overline{\mathbf{U}}(\xi_1) \tanh \overline{\mathbf{V}}(\xi_2)] d\xi_0 d\xi_1 + \\ &+ \sum_l \Lambda_l e^{-\gamma} \int \log_2 \left[\sum_{\sigma=\pm 1} (1 + \sigma \tanh \overline{\mathbf{H}}(\xi_0)) \prod_{i=1}^l (1 + \sigma \tanh \overline{\mathbf{U}}(\xi_i)) \right] \prod_{i=0}^l d\xi_i + \\ &+ \frac{\gamma}{P'(1)} \sum_k P_k \int \log_2 \left[1 + \tanh \overline{\mathbf{J}}(\xi_0) \prod_{i=1}^k \tanh \overline{\mathbf{V}}(\xi_i) \right] \prod_{i=0}^k d\xi_i - \mathfrak{C}(Q_V) - \frac{\gamma}{P'(1)} \mathfrak{C}(Q_C), \end{aligned} \quad (9.23)$$

with all the integrals on the ξ_i 's being on the interval $(0, 1)$. This representation allows to easily derive the following result.

Lemma 6 *Assume the supremum of the trial free energy ϕ_V over the space of admissible random variables is achieved for some couple (U, V) . Then*

$$V \stackrel{\text{d}}{=} h + \sum_{i=1}^l U_i, \quad (9.24)$$

l being a Poisson random variable of parameter γ and h distributed according to the definition (3.9).

Proof: Look at ϕ as a functional $(\overline{\mathbf{U}}, \overline{\mathbf{V}}) \rightarrow \phi(\overline{\mathbf{U}}, \overline{\mathbf{V}})$ of the inverse distributions $\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$. The idea is to differentiate this functional at its (assumed) maximum. Let $\overline{\mathbf{D}} : (0, 1) \rightarrow (-\infty, +\infty]$ be left continuous and non decreasing. It is an easy calculus exercise to show that

$$\left. \frac{d}{d\varepsilon} \phi(\overline{\mathbf{U}} + \varepsilon \overline{\mathbf{D}}, \overline{\mathbf{V}}) \right|_{\varepsilon=0} = -\Lambda'(1) \int [1 - \tanh^2 \overline{\mathbf{U}}(\xi)] \overline{\mathbf{D}}(\xi) \mathcal{F}_\xi(\overline{\mathbf{U}}, \overline{\mathbf{V}}), \quad (9.25)$$

$$\left. \frac{d}{d\varepsilon} \phi(\overline{\mathbf{U}}, \overline{\mathbf{V}} + \varepsilon \overline{\mathbf{D}}) \right|_{\varepsilon=0} = -\Lambda'(1) \int [1 - \tanh^2 \overline{\mathbf{V}}(\xi)] \overline{\mathbf{D}}(\xi) \mathcal{G}_\xi(\overline{\mathbf{U}}, \overline{\mathbf{V}}), \quad (9.26)$$

where

$$\mathcal{F}_\xi(\overline{\mathbf{U}}, \overline{\mathbf{V}}) = \int \frac{\tanh \overline{\mathbf{V}}(\xi_1)}{1 + \tanh \overline{\mathbf{U}}(\xi) \tanh \overline{\mathbf{V}}(\xi_1)} d\xi_1 - \quad (9.27)$$

$$- \sum_l \lambda_l \int \frac{\tanh[\overline{\mathbf{H}}(\xi_0) + \sum_{i=1}^l \overline{\mathbf{U}}(\xi_i)]}{1 + \tanh \overline{\mathbf{U}}(\xi) \tanh[\overline{\mathbf{H}}(\xi_0) + \sum_{i=1}^l \overline{\mathbf{U}}(\xi_i)]} \prod_{i=0}^l d\xi_i,$$

$$\mathcal{G}_\xi(\overline{\mathbf{U}}, \overline{\mathbf{V}}) = \int \frac{\tanh \overline{\mathbf{U}}(\xi_1)}{1 + \tanh \overline{\mathbf{U}}(\xi_1) \tanh \overline{\mathbf{V}}(\xi)} d\xi_1 - \quad (9.28)$$

$$- \sum_k \rho_k \int \frac{\tanh \overline{\mathbf{J}}(\xi_0) \prod_{i=1}^{k-1} \tanh \overline{\mathbf{V}}(\xi_i)}{1 + \tanh \overline{\mathbf{J}}(\xi_0) \prod_{i=1}^{k-1} \tanh \overline{\mathbf{V}}(\xi_i) \tanh \overline{\mathbf{V}}(\xi)} \prod_{i=0}^{k-1} d\xi_i.$$

Notice that $\mathcal{G}_\xi(\overline{\mathbf{U}}, \overline{\mathbf{V}})$ vanishes because of the admissibility condition (6.1). It is then straightforward to show that $\mathcal{F}_\xi(\overline{\mathbf{U}}, \overline{\mathbf{V}})$ must vanish for any ξ such that $\overline{\mathbf{U}}(\xi) < \infty$, in order for $(\overline{\mathbf{U}}, \overline{\mathbf{V}})$ to be a maximum. This in turns implies the thesis. \square

⁵Since throughout this Section the degree sequences are kept fixed, we shall drop the dependence of ϕ on (Λ, P) , and (with an abuse of notation) replace it with its dependence upon $\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$.

k	l	r_{des}	New UB	Gallager UB	Gallager LB	Shannon limit
4	3	1/4	0.2101(1)	0.2109164	0.2050273	0.2145018
5	3	2/5	0.1384(1)	0.1397479	0.1298318	0.1461024
6	3	1/2	0.1010(2)	0.1024544	0.0914755	0.1100279
6	4	1/3	0.1726(1)	0.1726268	0.1709876	0.1739524

Table 2: Thresholds for regular LDPC codes over the binary symmetric channel $\text{BSC}(p)$ (k and l are, respectively, the check and variable node degrees, and r_{des} the design rate). The new upper bound proved in this paper is evaluated numerically following the approach described in the text. The quoted error comes from Monte Carlo sampling of the random variables U and V . ‘Gallager UB’ and ‘Gallager LB’ refer respectively to the upper and lower bounds obtained by Gallager [1].

9.6 Numerical estimates and comparison with previous bounds

The discussion in the previous Section suggests a natural possibility for evaluating numerically the lower bound in Theorem 2. Run density evolution [3] for T iterations and then evaluate the trial entropy (6.2) taking U and V to be random variables with the density of (respectively) right-to-left or left-to-right messages. Notice that, in order for Eq. (6.1) to be satisfied, the right-to-left density must be updated one last time before evaluating the trial entropy.

This still leaves a lot of freedom. The first question is: how large T (the number of iterations) should be? While it is difficult to provide a quantitative answer, in order to approach the supremum in Eqs. (6.3) to (6.5), one should get a good approximation of fixed point densities. Generically, this happens only as T is let to infinity.

The next question is: how the densities should be initialized? This question has a very simple answer in usual applications of density evolution: just initialize to the message density seen at the zeroth step of message passing. This generally means U, V identically equal to 0. Hereafter, we shall refer to this as the ‘0-initialization’. This answer is no longer complete in the present context. In fact any initial condition, such that U and V are symmetric random variables, corresponds eventually to a valid lower bound of the form in Eqs. (6.3) to (6.5). At least one other simple initial condition consist in taking $U = V = +\infty$ identically. In the case of standard ensembles with minimum left degree at least 2 this is in fact a fixed point and the corresponding trial entropy vanishes. We shall refer to this as the ‘ ∞ -initialization’.

Despite this freedom, Eqs. (6.3) to (6.5) always provide a lower bound, no matter how we implement the general strategy. In Tab. 2 we report the numerically-evaluated upper bound for a few regular ensembles over the binary symmetric channel. We implemented a sampled (Monte Carlo) version of density evolution (with 10^4 to 10^5 sampling points) and adopted the 0-initialization. The trial entropy (6.2) was averaged over 10^4 iterations after 10^4 equilibration iterations. The threshold was estimated as the smallest noise parameter such that the trial entropy is positive.

In the same Table (and analogously in Table 1 for the erasure channel), we compare our upper bounds with previous upper and lower bounds. In his thesis [1], Gallager used an estimate of the distance spectrum, together with a clever modification of the union bound in order to obtain lower bounds. This technique was further generalized and refined over the years, see for instance [8, 10, 11] but it’s fair to say that there were no major modification over the initial result for the simplest regular ensembles and channel models. We should stress that a different technique, based on typical pairs decoding was proposed in [7]. Our evaluation of

Gallager bound wasn't numerically distinguishable from the results of [7].

As for upper bounds, Gallager approach is based on an information theoretic argument. Also in this case, despite some improvements [9], the main idea remained essentially unchanged. Moreover, the quantitative estimates by Gallager remained essentially the state-of-the art for the simplest regular ensembles and channel models considered here.

Despite the various estimates in Tables 1 and 2 are numerically close (which is partially due to the proximity of capacity), the bound of Theorem 2 is clearly superior to previous upper bounds.

9.7 Relation with the Bethe free energy

Until now we studied the average properties of the code *ensembles* defined in Sec. 2. Although the concentration result of Sec. 5 justify this approach, it may be interesting to take a step backward and consider the decoding problem for a single realization of the code and of the channel. It is convenient to introduce the 'Bethe free-energy' [39] $F_B(\underline{b})$ associated to the probability distribution (3.3). We have

$$F_B(\underline{b}) = U_B(\underline{b}) - H_B(\underline{b}), \quad (9.29)$$

where

$$U_B(\underline{b}) = - \sum_{a \in \mathcal{C}} \sum_{\underline{x}_a} b_a(\underline{x}_a) \log_2 Q_{\mathcal{C}}(\hat{y}_a | \underline{x}_a) - \sum_{i \in \mathcal{V}} \sum_{x_i} b_i(x_i) \log_2 Q_{\mathcal{V}}(y_i | x_i), \quad (9.30)$$

$$H_B(\underline{b}) = - \sum_{a \in \mathcal{C}} \sum_{\underline{x}_a} b_a(\underline{x}_a) \log_2 b_a(\underline{x}_a) + \sum_{i \in \mathcal{V}} (|\partial i| - 1) \sum_{x_i} b_i(x_i) \log_2 b_i(x_i), \quad (9.31)$$

and we used the shorthands $\underline{x}_a = (x_{i_1^a}, \dots, x_{i_k^a})$ and $Q_{\mathcal{C}}(\hat{y}_a | \underline{x}_a) = Q_{\mathcal{C}}(\hat{y}_a | x_{i_1^a} \oplus \dots \oplus x_{i_k^a})$. The parameters $\{b_a(\underline{x}_a) : a \in \mathcal{C}\}$ and $\{b_i(x_i) : i \in \mathcal{V}\}$ are probability distributions subject to the marginalization conditions

$$\sum_{x_j, j \in \partial a \setminus i} b_a(\underline{x}_a) = b_i(x_i) \quad \forall i \in \partial a, \quad (9.32)$$

$$\sum_{x_i} b_i(x_i) = 1 \quad \forall i. \quad (9.33)$$

For LDGM codes $F_B(\underline{b})$ is always finite. For LDPC codes, it takes values in $(-\infty, +\infty]$ and is finite if $b_a(\underline{x}_a)$ vanishes whenever $x_{i_1^a} \oplus \dots \oplus x_{i_k^a} = 0$ (as always we use the convention $0 \log 0 = 0$). Moreover, as explained in [39], its stationary points are fixed points of BP decoding.

Following [39], we consider the stationary points of the Bethe free energy (9.29) under the constraints (9.32), (9.33). This can be done by introducing a set of Lagrange multipliers $\{\lambda_{ia}(x_i)\}$ for Eq. (9.32), the constraint (9.33) being easily satisfied by normalizing the beliefs. One then consider the Lagrange function

$$L_B(b, \lambda) = F_B(b) - \sum_{(ia) \in \mathcal{E}} \sum_{x_i} \lambda_{ia}(x_i) \left[\sum_{x_j: j \in \partial a \setminus i} b_a(\underline{x}_a) - b_i(x_i) \right]. \quad (9.34)$$

We refer to [39] for further details of this computation. Stationary points are eventually shown to have the form

$$b_a(\underline{x}_a) = \frac{1}{z_a} Q_{\mathcal{C}}(\hat{y}_a | \underline{x}_a) \prod_{j \in \partial a} P_{v_{j \rightarrow a}}(x_j), \quad (9.35)$$

$$b_i(x_i) = \frac{1}{z_i} Q_{\mathcal{V}}(y_i | x_i) \prod_{a \in \partial i} P_{u_{a \rightarrow i}}(x_i), \quad (9.36)$$

with $P(x)$ being defined as in Eq. (4.9). The z_a 's and z_i 's are fixed by the normalization conditions $\sum_{\underline{x}} b_a(\underline{x}) = 1$ and $\sum_x b_i(x_i) = 1$. The messages $\{v_{i \rightarrow a}\}$ are related to the Lagrange multipliers $\{\lambda_{ia}(x_i)\}$ by the relation

$$P_{v_{i \rightarrow a}}(x_i) \propto \exp\{\lambda_{ia}(x_i)\}, \quad (9.37)$$

while the $\{u_{a \rightarrow i}\}$ must satisfy the equation

$$v_{i \rightarrow a} = h_i + \sum_{b \in \partial i \setminus a} u_{b \rightarrow i}, \quad (9.38)$$

for any $i \in \mathcal{V}$. The marginalization constraint (9.32) is satisfied if the equation

$$u_{a \rightarrow i} = \operatorname{arctanh}\left\{\tanh J_a \prod_{j \in \partial a \setminus i} \tanh v_{j \rightarrow a}\right\} \quad (9.39)$$

holds for any $a \in \mathcal{C}$.

If we substitute the beliefs (9.35), (9.36) into Eq. (9.29) we can express the Bethe free energy as a function of the messages $\underline{u} \equiv \{u_{a \rightarrow i}\}$ and $\underline{v} \equiv \{v_{i \rightarrow a}\}$. Using Eqs. (9.38), (9.39), we get the following expression (with a slight abuse of notation we do not change the symbol denoting the free energy)

$$\begin{aligned} F_{\text{B}}(\underline{u}, \underline{v}) &= \sum_{(ia) \in \mathcal{E}} \log_2 \left[\sum_{x_i} P_{u_{a \rightarrow i}}(x_i) P_{v_{i \rightarrow a}}(x_i) \right] - \sum_{i \in \mathcal{V}} \log_2 \left[\sum_{x_i} Q_{\mathcal{V}}(y_i | x_i) \prod_{a \in \partial i} P_{u_{a \rightarrow i}}(x_i) \right] - \\ &\quad - \sum_{a \in \mathcal{C}} \log_2 \left[\sum_{\underline{x}_a} Q_{\mathcal{C}}(\hat{y}_a | \underline{x}_a) \prod_{i \in \partial a} P_{v_{i \rightarrow a}}(x_i) \right]. \end{aligned} \quad (9.40)$$

A simple comparison of this expression with Eq. (6.2) yields the following interesting result.

Proposition 1 *Let $F_{\text{B}}(\underline{u}, \underline{v})$ be the Bethe free energy for any of the code ensembles LDGM(n, γ, P), LDPC(n, γ, P), LDGM(n, Λ, P), LDPC(n, Λ, P), with the beliefs parameterized as in Eqs. (9.35) and (9.36), and assume that the messages are i.i.d. random variables $u_{i \rightarrow a} \stackrel{\text{d}}{=} U$ and $v_{a \rightarrow i} \stackrel{\text{d}}{=} V$. Then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} F_{\text{B}}(\underline{u}, \underline{v}) = -\phi_V(\Lambda, P) + \kappa \sum_y Q(y|0) \log_2 Q(y|0), \quad (9.41)$$

where the expectation $\mathbb{E}(\cdot)$ is taken both with respect to the messages distribution and the code ensemble, and $\kappa = \gamma/P'(1)$ (for LDGM ensembles) or $\kappa = 1$ (for LDPC ensembles). For multi-Poisson ensembles LDGM(n, γ, Λ, P) and LDPC(n, γ, Λ, P), the same formula holds with Λ being replaced by $\Lambda^{(\gamma)}$ on the right hand side.

Proof: In order to compute the expectation on the left hand side of Eq. (9.41), let us proceed in two steps. In a first step, we shall take the expectation with respect to the messages $\{u_{i \rightarrow a}, v_{a \rightarrow i}\}$, which in the Theorem statement are assumed to be i.i.d.'s, as well as with respect to the channel output $\{y_i, \hat{y}_a\}$. Let us denote by $\mathcal{V}_l(\mathcal{C}_k)$ the set of variable nodes (check nodes) of degree l (degree k). By linearity of expectation, we get

$$\begin{aligned} \mathbb{E}_{u,v} F_B(\underline{u}, \underline{v}) &= |\mathcal{E}| \mathbb{E}_{u,v} \log_2 \left[\sum_x P_u(x) P_v(x) \right] - \sum_l |\mathcal{V}_l| \mathbb{E}_y \mathbb{E}_u \log_2 \left[\sum_x Q_{\mathcal{V}}(y|x) \prod_{a=1}^l P_{u_a}(x) \right] - \\ &\quad - \sum_k |\mathcal{C}_k| \mathbb{E}_v \mathbb{E}_u \log_2 \left[\sum_{\underline{x}} Q_{\mathcal{C}}(\hat{y}|x_1 \oplus \dots \oplus x_k) \prod_{i=1}^k P_{u_i}(x_i) \right]. \end{aligned} \quad (9.42)$$

Now notice that the number of edges is equal to the number of variable nodes times the average left degree: $|\mathcal{E}| = n \hat{\Lambda}'(1)$. The number of variable nodes of degree l is, by definition $|\mathcal{V}_l| = n \hat{\Lambda}_l$. Furthermore the total number of check nodes is $n \hat{\Lambda}'(1) / \hat{P}'(1)$, and therefore $|\mathcal{C}_k| = (n \hat{\Lambda}'(1) / \hat{P}'(1)) \hat{P}_k$. Finally both for Poisson and standard ensembles, the expected degree profile converges to the design profile, see Sec. 2.2. In other words

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{G}} \hat{\Lambda}_l = \Lambda_l, \quad \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{G}} \hat{P}_k = P_k, \quad \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{G}} \hat{\Lambda}'(1) = \Lambda'(1). \quad (9.43)$$

Therefore (9.41) is proved by taking the expectation with respect to the graph ensemble in Eq. (9.42) and then taking the large blocklength limit⁶.

Finally, for the multi-Poisson ensemble we gave just to notice that the expected left degree profile converges to $\Lambda^{(\gamma)}$ rather than to Λ , see App. B \square

This result provides an appealing interpretation for the trial entropy entering in Theorem 2. Apart from a simple rescaling, it is asymptotically equal to the expected value of the Bethe free energy when the messages $\{u_{i \rightarrow a}\}$ and $\{v_{a \rightarrow i}\}$ are i.i.d. random variables. Viceversa, Theorem 2 can be interpreted as yielding a connection between the Bethe free energy, and the conditional entropy of the transmitted message.

10 Generalizations and conclusion

We expect that the results derived in this paper can be extended in several directions.

A first direction consists in proving the analogue of Theorem 2 for more general code ensembles. It is important to notice that the technique used in this paper (as well as in [26]) makes a crucial use of the convexity of $P(x)$. Although the non-rigorous calculations of [17–20] suggest that the result will have the same form for a non-convex $P(x)$, the proof is probably more difficult in this case.

A second direction consists in proving that the bound of Theorem 2 is indeed tight. We precise this claim as follows.

Conjecture 1 *Under the hypotheses of Theorem 2 we have*

$$\lim_{n \rightarrow \infty} h_n = \sup_V \phi_V, \quad (10.1)$$

⁶Notice that in the case of Poisson ensembles Λ_l as no bounded support (l can be arbitrarily large). However the thesis follows from convergence of $\mathbb{E}_{\mathcal{G}} \hat{\Lambda}_l$ in total variation distance.

where the sup has to be taken over the space of admissible random variables. The degree sequences to be used as argument of ϕ_V are the same as in Theorem 2.

Once again, this claim is supported by [17–20].

Finally, in this paper we limited ourselves to estimating the conditional entropy per channel use. As discussed in Section 5, this implies only sub-optimal bounds on the bit error rate. It would be therefore important to estimate directly this quantity without passing through Fano inequality. The results of [17–20] suggest the following recipe for computing the bit error rate under symbol MAP decoding. Determine the message densities maximizing the trial entropy, cf. Eq. (6.3). Compute the density of a posteriori likelihoods as in density evolution (this implies a convolution of all the densities incoming in a variable node). The bit error rate is simply given by the weight of negative log-likelihoods under this distribution.

Finally, one may hope that the strong connection between message passing techniques (density evolution) and MAP decoding (conditional entropy) highlighted in the present approach may lead to a better understanding of the former.

Acknowledgments

I am grateful to Tom Richardson and RÜdi Urbanke for their encouragement, and to Silvio Franz and Fabio Toninelli for many exchanges on the interpolation method.

Note: while this paper was being revised, an alternative technique has been put forward [41] which allows to rederive part of the present results.

A Coupling graph ensembles

In this Appendix we prove Lemma 1. Instead of exhibiting directly a coupling between a standard graph $\mathcal{G}_s \stackrel{d}{=} (n, \Lambda, P)$ and a multi-Poisson graph $\mathcal{G}_{\text{mP}} \stackrel{d}{=} (n, \Lambda, P, \gamma)$, we shall proceed in two steps. More precisely, we shall exhibit two couplings $(\mathcal{G}_s, \mathcal{G}_*)$ and $(\mathcal{G}_*, \mathcal{G}_{\text{mP}})$ where the distribution of $\mathcal{G}_* \stackrel{d}{=} (n, \Lambda, P, \gamma)_*$ is defined below (as in Sec. 2.3, we let $t_{\text{max}} \equiv \lfloor \Lambda'(1)/\gamma \rfloor - 1$ be the number of rounds).

In order to generate a random element in $(n, \Lambda, P, \gamma)_*$, proceed as for the multi-Poisson ensemble (see Definition 4) but the following modification. During stage t , for each check node $a = (\hat{a}, k, t) \in \mathcal{C}_t$, and for each $r = 1, \dots, k$, i_r^a is chosen randomly in \mathcal{V} with distribution $w_i(t, a, r) = (d_i(t) - \Delta_i(t, a, r)) / [\sum_j (d_i(t) - \Delta_i(t, a, r))]$, where $\Delta_i(t, a, r)$ is the number of times i has been already chosen during stage t . In other words, unlike in the multi-Poisson ensemble, we keep track faithfully of the number of free sockets.

Let us now describe how the two-step coupling works.

From \mathcal{G}_{mP} to \mathcal{G}_* : Consider round t . Let $d_i^{\text{mP}}(t)$ and $d_i^*(t)$ be the number of free sockets respectively for \mathcal{G}_{mP} and \mathcal{G}_* . Choose the variable nodes $(i_r^a)_{\text{mP}}$ and $(i_r^a)_*$ in the two graphs by coupling optimally (see discussion below) the distributions $w_i^{\text{mP}}(t) = [d_i^{\text{mP}}(t)]_+ / (\sum_j [d_j^{\text{mP}}(t)]_+)$ and $w_i^*(t; a, r) = (d_i^*(t) - \Delta_i(t; a, r)) / [\sum_j (d_j^*(t) - \Delta_j(t; a, r))]$. If $(i_r^a)_* \neq (i_r^a)_{\text{mP}}$, we say that a ‘discrepancy’ has occurred. We claim that, if $\gamma < 1$, then the total number of discrepancies is smaller than $An\gamma^b$ w.h.p. (C and b being n - and γ -independent constants). The proof of this claim is provided below. Of course the number of rewirings necessary to pass from \mathcal{G}_{mP} to \mathcal{G}_* is bounded by twice the number of discrepancies.

From \mathcal{G}_* to \mathcal{G}_s : Notice that \mathcal{G}_* is generated in the same way as \mathcal{G}_s (see discussion at the end of Sec. 2.1) but for the fact that it contains a random number of check nodes. In fact, the total number of check nodes of degree k (call it m_k) in a $(n, \Lambda, P, \gamma)_*$ graph is a Poisson random variable with mean $\bar{m}_k = t_{\max} n \gamma P_k / P'(1)$. Denoting by $m_k^{(s)} = n \Lambda'(1) P_k / P'(1)$ the number of check nodes in a standard (n, Λ, P) graph, it is easy to see that $m_k^{(s)} - 2Bn\gamma < \bar{m}_k \leq m_k^{(s)} - Bn\gamma$ for some positive (n and γ independent) constant B . By elementary properties of Poisson random variables, one obtains $m_k^{(s)} - 3Bn\gamma < m_k \leq m_k^{(s)} - (1/2)Bn\gamma$ for each $k \in \{2, \dots, k_{\max}\}$ with probability greater than $1 - 2e^{-C\gamma^2 n}$ for some constant C .

We therefore obtain the desired coupling as follows: first generate \mathcal{G}_* . If $m_k > m_k^{(s)}$ for any k , then generate an independent graph \mathcal{G}_s . In the opposite case, generate \mathcal{G}_s by adding $m_k^{(s)} - m_k$ check nodes for each k and connect them to variable nodes as described at the end of Sec. 2.1. Because of the above argument, the number of rewirings (check nodes added) is smaller than $A'n\gamma$ with high probability.

We are now left with the task of proving the claim in the first step. Before accomplishing this task, it is worth recalling an elementary fact which is useful in this proof [40]. Given two distributions $\{w_i^{(1)}\}$ and $\{w_i^{(2)}\}$ over $i \in [n]$, their total variation distance is defined as

$$\|w^{(1)} - w^{(2)}\| = \frac{1}{2} \sum_{i=1}^n |w_i^{(1)} - w_i^{(2)}|. \quad (\text{A.1})$$

Furthermore, if i_1 and i_2 are distributed according to $w^{(1)}$ and $w^{(2)}$, there exist coupling between them (i.e. a joint distribution which has $w^{(1)}, w^{(2)}$ as marginals) such that $\|w^{(1)} - w^{(2)}\| = \mathbb{P}(i_1 \neq i_2)$. Such coupling is ‘optimal’ in the sense that, for any coupling we have $\|w^{(1)} - w^{(2)}\| \leq \mathbb{P}(i_1 \neq i_2)$.

The proof of the claim is obtained by recursively estimating the number of discrepancies between \mathcal{G}_{mP} and \mathcal{G}_* . Suppose that we have terminated the first t rounds (denoted as $0, \dots, t-1$ in Definition 4) in the generation of the couple $(\mathcal{G}_{\text{mP}}, \mathcal{G}_*)$ and no more than $C_t n \gamma$ discrepancies occurred so far (with C_t n -independent). This hypothesis trivially holds for $t = 0$. We will determine an n -independent constant C_{t+1} , such that, at the end of the t -th step there will be, w.h.p., less than $C_{t+1} n \gamma$ discrepancies. By iterating this argument, we deduce that \mathcal{G}_{mP} and \mathcal{G}_* have less than $C_{t_{\max}} n \gamma$ discrepancies with high probability, and will be able to obtain the estimate $C_{t_{\max}} \leq A + B\gamma^{-\rho}$ with $0 < \rho < 1$, $A, B > 0$ three γ -independent constants. This implies Lemma 1 with $b = 1 - \rho$.

During the round t , $(i_r^a)_*$ and $(i_r^a)_{\text{mP}}$ are taken from the distributions $w_i^{\text{mP}}(t)$ and $w_i^*(t; a, r)$ described above. Let us start by noticing that, with high probability

$$\begin{aligned} \sum_{i=1}^n |[d_i^{\text{mP}}(t)]_+ - d_i^*(t) + \Delta_i(t; a, r)| &\leq \sum_{i=1}^n |d_i^{\text{mP}}(t) - d_i^*(t)| + \sum_{i=1}^n \Delta_i(t; a, r) \leq \\ &\leq C_t n \gamma^2 t + 2n\gamma. \end{aligned} \quad (\text{A.2})$$

The first inequality follows from the fact that $d_i^*(t)$ and $\Delta_i(t; a, r)$ are non-negative. The second one, from the induction hypothesis together with the observation that $\sum_{i=1}^n \Delta_i(t; a, r)$ is smaller than the total number of variable node choices during round t , which in turn is a Poisson random variable with mean $n\gamma$.

Next, we observe that, w.h.p.

$$\sum_{i=1}^n [d_i^{\text{mP}}(t)]_+ \geq n(\Lambda'(1) - \gamma t - \gamma). \quad (\text{A.3})$$

In fact, at $t = 0$ the sum on the left-hand side has value $n\Lambda'(1)$, and after each round it decreases at most by the number of left sockets which are occupied during that round. This is a Poisson random variable of mean $n\gamma$.

Now we can estimate the variation distance between the distributions of $(i_r^a)_{\text{mP}}$ and $(i_r^a)_*$:

$$\begin{aligned} \|w^{\text{mP}}(t) - w^*(t; a, r)\| &\leq \frac{\sum_{i=1}^n |[d_i^{\text{mP}}(t)]_+ - d_i^*(t) + \Delta_i(t; a, r)|}{\sum_{i=1}^n [d_i^{\text{mP}}(t)]_+} \leq \\ &\leq \frac{C_t \gamma^2 t + 2\gamma}{\Lambda'(1) - \gamma(t+1)} \leq \end{aligned} \quad (\text{A.4})$$

$$\leq \frac{C_t \Lambda'(1) + 2}{t_{\max} - t}, \quad (\text{A.5})$$

where the second inequality follows from $t < t_{\max}$ and $t_{\max} \leq \Lambda'(1)/\gamma - 1$. During round t , about $n\gamma$ couples $(i_r^a)_{\text{mP}}$ and $(i_r^a)_*$ are chosen and they differ with probability $\|w^{\text{mP}}(t) - w^*(t; a, r)\|$ (because we coupled $w^{\text{mP}}(t)$ and $w^*(t; a, r)$ optimally). The total number of discrepancies is therefore smaller than $2n\gamma\|w^{\text{mP}}(t) - w^*(t; a, r)\|$ with high probability. Unhappily this estimate worsen as t approaches t_{\max} because of the denominator in (A.5). This problem is overcome as follows. Fix $t_* \equiv t_{\max} - \lfloor t_{\max}^\rho \rfloor$, where $0 < \rho < 1$ is the solution of $\rho = 2\Lambda'(1)(1-\rho)$, and use the estimate (A.5) only for $t < t_*$. For $t_* \leq t < t_{\max}$ we just use the fact that during each round no more than $n\gamma$ discrepancies can be introduced. In other words

$$C_{t+1} = \begin{cases} C_t + 2\Lambda'(1)(C_t + \lambda)/(t_{\max} - t) & \text{if } t < t_*, \\ C_t + 1 & \text{if } t_* \leq t < t_{\max}, \end{cases} \quad (\text{A.6})$$

where we introduced the constant $\lambda \equiv 2/\Lambda'(1) > 0$. This recursion is easily summed up, yielding

$$\log(C_{t_*} + \lambda) - \log(\lambda) = \sum_{t=0}^{t_*-1} \log\left(1 + \frac{2\Lambda'(1)}{t_{\max} - t}\right) \leq \sum_{t=0}^{t_*-1} \frac{2\Lambda'(1)}{t_{\max} - t} \leq \quad (\text{A.7})$$

$$\leq 2\Lambda'(1) \sum_{t=\lfloor t_{\max}^\rho \rfloor + 1}^{t_{\max}} \frac{1}{t} \leq 2\Lambda'(1) \log\left(\frac{t_{\max} e}{\lfloor t_{\max}^\rho \rfloor}\right). \quad (\text{A.8})$$

Using the definition of t_{\max} and the relation $2\Lambda'(1)(1-\rho) = \rho$, we get $C_{t_*} \leq A + B\gamma^{-\rho}$ with A and B two γ independent constants. Finally $C_{t_{\max}} \leq C_{t_*} + \lfloor t_{\max}^\rho \rfloor \leq A' + B'\gamma^{-\rho}$.

B Degree distribution for multi-Poisson ensembles

In this Appendix we provide an asymptotic characterization of the variable-node degree profile for the multi-Poisson ensemble (n, Λ, P, γ) . We shall start by defining a construction which yields the degree profile in the large blocklength limit and then prove convergence to this construction.

Let $\Omega_{l,d}^{(t)}$ be a sequence of distributions over $l \in \{2, \dots, l_{\max}\}$ and $d \in \mathbb{Z}$ indexed by $t \in \{0, \dots, t_{\max}\}$ and defined as follows. Introduce the kernel

$$W_t(\Delta|d) = e^{-\lambda(d)} \frac{\lambda(d)^\Delta}{\Delta!}, \quad \lambda(d) \equiv \frac{\gamma[d]_+}{\sum_{d',l} \Omega_{l,soc}^{(t)}[d]_+}. \quad (\text{B.1})$$

Next define recursively $\Omega_{l,d}^{(t)}$ as

$$\Omega_{l,d}^{(t+1)} = \sum_{d' \geq d} \Omega_{l,d'}^{(t)} W(d' - d|d'), \quad \Omega_{l,d}^{(0)} = \Lambda_l \mathbb{I}_{d=0}, \quad (\text{B.2})$$

where \mathbb{I}_A is the indicator function of the event A . Notice that the sum in the denominator of Eq. (B.1) is always well-defined. In fact from the definition follows that $\Omega_{l,d}^{(t)} = 0$ if $d > l_{\max}$. Finally, we define the asymptotic degree profile to be

$$\Lambda_l^{(\gamma)} = \sum_{l'} \Omega_{l',l-l}^{(t_{\max})}. \quad (\text{B.3})$$

The following result implies that $\{\Lambda_l^{(\gamma)}\}$ is in fact the correct asymptotic degree profile.

Lemma 7 *Let $\{\hat{\Lambda}_l : l = 2, \dots, l_{\max}\}$ be the variable nodes degree profile for a random Tanner graph from the (n, Λ, P, γ) multi-Poisson ensemble. Denote by $\|\mu - \nu\|$ the total variation distance between distributions μ and ν (see previous Section). If $\{\Lambda_l^{(\gamma)}\}$ is defined as above, then there exist $A, B > 0$, such that*

$$(I). \quad \lim_{n \rightarrow \infty} \mathbb{E} \hat{\Lambda}_l = \Lambda_l^{(\gamma)}, \quad (\text{B.4})$$

$$(II). \quad \lim_{n \rightarrow \infty} \|\mathbb{E} \hat{\Lambda} - \Lambda^{(\gamma)}\| = 0, \quad (\text{B.5})$$

$$(III). \quad \mathbb{P} \left\{ |\hat{\Lambda}_l - \Lambda_l^{(\gamma)}| \geq \varepsilon \right\} \leq A e^{-Bn\varepsilon^2}, \quad (\text{B.6})$$

for any positive ε .

Proof: Notice that (III) obviously implies (I). Moreover (see proof of Corollary 1), if a sequence of distributions over the integers $\mu^{(n)}$ converges pointwise to a (normalized) distribution $\mu^{(\infty)}$, then $\|\mu^{(n)} - \mu^{(\infty)}\| \rightarrow 0$. Therefore (I) implies (III).

We are left with the task of proving (III). We shall in fact prove the following stronger statement. Consider a multi-Poisson graph generated as in Definition (4). Let $\hat{\Omega}_{l,d}^{(t)}$ be the fraction of variable nodes $i \in \mathcal{V}$ such that $i \in \mathcal{V}_l$ (or, equivalently, $\mathbf{d}_i(0) = l$) and $\mathbf{d}_i(t) = \mathbf{d}$. We claim that $\hat{\Omega}_{l,d}^{(t)}$ is well approximated by the sequence $\Omega_{l,d}^{(t)}$ defined above. More precisely, there exist constants A, B (which may depend on the ensemble parameters as well as on t) such that

$$\mathbb{P} \left\{ |\hat{\Omega}_{l,d}^{(t)} - \Omega_{l,d}^{(t)}| \geq \varepsilon \right\} \leq A e^{-Bn\varepsilon^2}. \quad (\text{B.7})$$

Recall that the degree of a variable node $i \in \mathcal{V}_l$ is $l - \mathbf{d}_i(t_{\max})$, we have

$$\hat{\Lambda}_l = \sum_{l'} \hat{\Omega}_{l',l-l}^{(t_{\max})}. \quad (\text{B.8})$$

The thesis therefore follows from Eq. (B.7) together with the observation that the sum (B.8) contains a finite number of terms.

The claim is proved by induction on t . It is obviously true for $t = 0$. Assume now that the claim is true up to stage t , and consider the distribution of $\widehat{\Omega}_{l,d}^{(t+1)}$ conditional to $\widehat{\Omega}_{l,d}^{(t)}$. By the induction hypothesis we can assume that $|\widehat{\Omega}_{l,d}^{(t)} - \Omega_{l,d}^{(t)}| \leq \varepsilon$. Furthermore, since $\widehat{\Omega}_{l,d}^{(t)} = 0$ whenever $d > l_{\max}$, we can also assume

$$\left| \sum_d \widehat{\Omega}_{l,d}^{(t)}[d]_+ - \sum_d \Omega_{l,d}^{(t)}[d]_+ \right| \leq \varepsilon. \quad (\text{B.9})$$

We shall neglect the exponentially rare cases in which these conditions do not hold.

The total number of variable nodes sampled during stage $t + 1$ is $\Delta_{\text{tot}} = \sum_k k m_k^{(t)}$ where $m_k^{(t)}$ is a Poisson random variable with mean $n\gamma P_k/P'(1)$. We can therefore assume that $|\Delta_{\text{tot}} - n\gamma| \leq n\varepsilon$ and neglect the rare cases in which this is false. Next, consider a variable i , such that $\mathbf{d}_i(t) = \mathbf{d}$. The probability that this is chosen when selecting one of the neighbors of a function node a is

$$w_i(t) = \frac{[\mathbf{d}_i(t)]_+}{\sum_j [\mathbf{d}_j(t)]_+} = \frac{[\mathbf{d}]_+}{\sum_{l,d} \widehat{\Omega}_{l,d}^{(t)}[d]_+} \equiv w(\mathbf{d}). \quad (\text{B.10})$$

The probability that, during stage t , this variable node is selected $\Delta_i(t) = \Delta$ times (conditional to the total number Δ_{tot}) is therefore

$$\begin{aligned} \mathbb{P}[\Delta_i(t) = \Delta | \mathbf{d}_i(t) = \mathbf{d}] &= \binom{\Delta_{\text{tot}}}{\Delta} w(\mathbf{d})^\Delta (1 - w(\mathbf{d}))^{\Delta_{\text{tot}} - \Delta} = \\ &= W_t(\Delta | \mathbf{d}) + O(1/n) + O(\varepsilon). \end{aligned} \quad (\text{B.11})$$

Therefore, the fraction of variable nodes such that $\mathbf{d}_i(0) = l$, $\mathbf{d}_i(t) = \mathbf{d}$ and $\Delta_i(t) = \Delta$ (to be denoted by $\widehat{\Omega}_{\Delta,l,d}^{(t)}$) is concentrated around $[W_t(\Delta | \mathbf{d}) + O(\varepsilon)] \widehat{\Omega}_{l,d}^{(t)}$. Using the induction hypothesis, this implies that

$$\mathbb{P} \left\{ \left| \widehat{\Omega}_{\Delta,l,d}^{(t)} - W_t(\Delta | \mathbf{d}) \Omega_{l,d}^{(t)} \right| \geq \varepsilon \right\} \leq A e^{-nB\varepsilon^2}. \quad (\text{B.12})$$

Next we notice that $\mathbf{d}_i(t) = \mathbf{d}$ and $\Delta_i(t) = \Delta$ implies $\mathbf{d}_i(t+1) = \mathbf{d} - \Delta$. Therefore $\widehat{\Omega}_{l,d}^{(t+1)} = \sum_{d' \geq d} \widehat{\Omega}_{d'-d,l,d'}^{(t)}$. Since a finite number of terms enter in this sum, Eq. (B.12) implies

$$\mathbb{P} \left\{ \left| \widehat{\Omega}_{l,d}^{(t+1)} - \sum_{d' \geq d} W_t(d' - d | d') \Omega_{l,d'}^{(t)} \right| \geq \varepsilon \right\} \leq A e^{-nB\varepsilon^2}, \quad (\text{B.13})$$

for some, eventually different, constants A and B . Notice that the sum in the above expression is exactly $\Omega_{l,d}^{(t+1)}$. Therefore, thesis (III) is proved. \square

C Derivative of the conditional entropy: Poisson ensemble

In this Appendix we compute derivative of the conditional entropy with respect to the interpolation parameter, cf. Eq. (6.14). The crucial observation is the following. Let n a poissonian

random variable with parameter $\lambda > 0$, and $f : \mathbb{N} \rightarrow \mathbb{R}$ any function on the positive integers, then:

$$\frac{d}{d\lambda} \mathbb{E}f(n) = \mathbb{E}[f(n+1) - f(n)]. \quad (\text{C.1})$$

Consider now the expression (6.8) for the interpolating conditional entropy. This depends upon t through the distributions of the m_k 's (i.e. the number of right nodes of degree k which is a poissonian variable with parameter $ntP_k/P'(1)$), and the distribution of the l_i 's (i.e. the number of repetitions for the variable x_i , which is a random variable with parameter $\gamma - t$). When differentiating with respect to t we get therefore a sum of several contributions. For the sake of clarity, let us compute in detail one of such terms. In order to single out a term, assume that the parameter entering in the distribution of m_k is t_k and is distinct from t (i.e. the number of right nodes of degree k is a poissonian variable with parameter $nt_kP_k/P'(1)$). Let $\mathcal{Z} = \mathcal{Z}(m_k)$ be the normalization defined in Eq. (6.6) for a graph with m_k check nodes of degree k . Applying the above formula we get

$$\frac{dh_n}{dt_k}(t, t_k) = \frac{P_k}{P'(1)} \mathbb{E}_s \log_2 \{ \mathcal{Z}(m_k + 1) / \mathcal{Z}(m_k) \}. \quad (\text{C.2})$$

The symbol \mathbb{E}_s includes expectation with respect to m_k , the choice of $m_k + 1$ check nodes of degree k , as well as with respect to the corresponding received message. We can however single out expectation with respect to the last of these nodes and use the fact that $\mathcal{Z}(m_k)$ does not depend on it. Denote $\mathcal{Z}_{\mathcal{C}}(i_1 \dots i_k; \hat{y})$ the normalization constant in Eq. (6.6) when a factor $Q_{\mathcal{C}}(\hat{y}|x_{i_1} \oplus \dots \oplus x_{i_k})$ is multiplied to the probability distribution. Then we have

$$\frac{dh_n}{dt_k}(t, t_k) = \frac{P_k}{P'(1)} \frac{1}{n^k} \sum_{i_1 \dots i_k} \mathbb{E}_{\hat{y}} \mathbb{E}_s \log_2 \{ \mathcal{Z}_{\mathcal{C}}(i_1 \dots i_k; \hat{y}) / \mathcal{Z} \}. \quad (\text{C.3})$$

The same calculation can be repeated for check each degree k as well as for the dependency upon t of the distribution of the l_i 's. We introduce the notation $\mathcal{Z}_{\text{eff}}(i; z)$ for the normalization constant in Eq. (6.6) when a factor $Q_{\text{eff}}(z|x_i)$ is multiplied. With these definitions we have (here we set again $t_k = t$)

$$\begin{aligned} \frac{dh_n}{dt}(t) &= \sum_k \frac{P_k}{P'(1)} \frac{1}{n^k} \sum_{i_1 \dots i_k} \mathbb{E}_{\hat{y}} \mathbb{E}_s \log_2 \{ \mathcal{Z}_{\mathcal{C}}(i_1 \dots i_k; \hat{y}) / \mathcal{Z} \} - \frac{1}{n} \sum_{i \in \mathcal{V}} \mathbb{E}_z \mathbb{E}_s \log_2 \{ \mathcal{Z}_{\text{eff}}(i; z) / \mathcal{Z} \} - \\ &\quad - \frac{1}{P'(1)} \sum_y Q_{\mathcal{C}}(y|0) \log_2 Q_{\mathcal{C}}(y|0) + \sum_z Q_{\text{eff}}(z|0) \log_2 Q_{\text{eff}}(z|0). \end{aligned} \quad (\text{C.4})$$

The expression (6.14) is recovered by noticing that $\mathcal{Z}_{\mathcal{C}}(i_1 \dots i_k; \hat{y}) / \mathcal{Z} = \langle Q_{\mathcal{C}}(y|x_{i_1} \oplus \dots \oplus x_{i_k}) \rangle_t$ and $\mathcal{Z}_{\text{eff}}(i; z) / \mathcal{Z} = \langle Q_{\text{eff}}(z|x_i) \rangle$.

D Positivity of $R_n(t)$: Poisson ensemble

In this Appendix we show that, under the hypotheses of Theorem 2, the remainder $R_n(t)$ in Eq. (6.16) is positive. This completes the proof of the Theorem. We start by writing the remainder in the form

$$R_n(t) = R_{a,n}(t) - R_{b,n}(t) - \widehat{R}_{a,n}(t) + \widehat{R}_{b,n}(t), \quad (\text{D.1})$$

where (throughout this Appendix entropies will be measured in nats: this clearly does not affect the sign of $R_n(t)$)

$$R_{a,n}(t) = \frac{1}{P'(1)} \sum_k P_k \frac{1}{n^k} \sum_{i_1 \dots i_k} \mathbb{E}_{\hat{y}} \mathbb{E}_s \log \left\langle \frac{2 Q_C(\hat{y} | x_{i_1} \oplus \dots \oplus x_{i_k})}{Q_C(\hat{y} | 0) + Q_C(\hat{y} | 1)} \right\rangle_t, \quad (\text{D.2})$$

$$R_{b,n}(t) = \frac{1}{n} \sum_i \mathbb{E}_z \mathbb{E}_s \log \left\langle \frac{2 Q_{\text{eff}}(z | x_i)}{Q_{\text{eff}}(z | 0) + Q_{\text{eff}}(z | 1)} \right\rangle_t. \quad (\text{D.3})$$

Analogous definitions are understood for $\widehat{R}_{a,n}(t)$ and $\widehat{R}_{b,n}(t)$ with the $\langle \cdot \rangle_t$ average being substituted by an average over $P_{v_1}(x_1) \dots P_{v_n}(x_n)$ as in the passage from Eq. (6.14) to Eq. (6.15). The code average \mathbb{E}_s has to be of course substituted by an average over V variables \mathbb{E}_v .

We shall treat each of the four terms $R_{a,n}(t) \dots \widehat{R}_{b,n}(t)$ separately and put everything together at the end. Let us start from the first term. Using the definitions (3.9) and (4.3) we get

$$R_{a,n}(t) = \frac{1}{P'(1)} \sum_k P_k \frac{1}{n^k} \sum_{i_1 \dots i_k} \mathbb{E}_J \mathbb{E}_s \log[1 + \tanh J \tanh \ell_{i_1 \dots i_k}]. \quad (\text{D.4})$$

Here we did not write explicitly the dependence of the log-likelihood $\ell_{i_1 \dots i_k}$ for the sum $x_{i_1} \oplus \dots \oplus x_{i_k}$ upon the received message $(\underline{y}, \hat{y}, \underline{z})$ and the code realization. We notice now that J and $\ell_{i_1 \dots i_k}$ are two independent symmetric random variables. We can therefore apply the observations of Sec. 4 (and in particular Lemma 3 to get

$$R_{a,n}(t) = \sum_{m=1}^{\infty} c_m R_{a,n}(t; m), \quad (\text{D.5})$$

where

$$c_m \equiv \left(\frac{1}{2m-1} - \frac{1}{2m} \right) > 0, \quad (\text{D.6})$$

and

$$R_{a,n}(t; m) \equiv \frac{1}{P'(1)} \mathbb{E}_J [(\tanh J)^{2m}] \sum_k P_k \frac{1}{n^k} \sum_{i_1 \dots i_k} \mathbb{E}_s [(\tanh \ell_{i_1 \dots i_k})^{2m}]. \quad (\text{D.7})$$

It is now convenient to introduce the ‘spin’ variables⁷ σ_i , $i \in \mathcal{V}$ as follows

$$\sigma_i = \begin{cases} +1 & \text{if } x_i = 0, \\ -1 & \text{if } x_i = 1. \end{cases} \quad (\text{D.8})$$

Notice that $\tanh \ell_{i_1 \dots i_k} = \langle \sigma_{i_1} \dots \sigma_{i_k} \rangle_t$. We can also write the $2m$ -th power of $\tanh \ell_{i_1 \dots i_k}$ introducing $2m$ i.i.d. copies $\underline{\sigma}^{(1)}, \dots, \underline{\sigma}^{(2m)}$. Using the notation introduced in Eq. (3.7) we get

$$(\tanh \ell_{i_1 \dots i_k})^{2m} = \langle (\sigma_{i_1}^{(1)} \dots \sigma_{i_k}^{(1)}) \dots (\sigma_{i_1}^{(2m)} \dots \sigma_{i_k}^{(2m)}) \rangle_{t,*}. \quad (\text{D.9})$$

We replace this formula in Eq. (D.7), and we are finally able to carry on the sums over $i_1 \dots i_k$ and k . The final result is remarkably compact

$$R_{a,n}(t; m) = \frac{1}{P'(1)} \mathbb{E}_J [(\tanh J)^{2m}] \mathbb{E}_s \langle P(Q_{2m}) \rangle_{t,*}, \quad (\text{D.10})$$

⁷This name comes from the statistical mechanics analogy [14].

where we defined the ‘multi-overlaps’

$$Q_{2m}(\underline{\sigma}^{(1)}, \dots, \underline{\sigma}^{(2m)}) = \frac{1}{n} \sum_{i=1}^n \sigma_i^{(1)} \dots \sigma_i^{(2m)}. \quad (\text{D.11})$$

Notice that $Q_{2m} \in [-1, 1]$.

The same procedure can be repeated for $R_{b,n}(t)$. We get $R_{b,n}(t) = \sum_m c_m R_{b,n}(t; m)$, with

$$R_{b,n}(t; m) = \mathbb{E}_u[(\tanh u)^{2m}] \frac{1}{n} \sum_i \mathbb{E}_s[(\tanh \ell_i)^{2m}] = \quad (\text{D.12})$$

$$= \mathbb{E}_u[(\tanh u)^{2m}] \frac{1}{n} \sum_i \mathbb{E}_s[\langle \sigma_i \rangle_t^{2m}] = \quad (\text{D.13})$$

$$= \mathbb{E}_u[(\tanh u)^{2m}] \frac{1}{n} \sum_i \mathbb{E}_s[\langle \sigma_i^{(1)} \dots \sigma_i^{(2m)} \rangle_{t,*}] = \quad (\text{D.14})$$

$$= \mathbb{E}_u[(\tanh u)^{2m}] \mathbb{E}_s \langle Q_{2m} \rangle. \quad (\text{D.15})$$

Let us now consider $\widehat{R}_{a,n}(t)$. Since the probability distribution for the bits x_i ’s is factorized, the averages can now be easily computed. We get

$$\widehat{R}_{a,n}(t) = \frac{1}{P'(1)} \sum_k P_k \mathbb{E}_J \mathbb{E}_v \log[1 + \tanh J \tanh v_1 \dots \tanh v_k]. \quad (\text{D.16})$$

Notice that in fact the right-hand side is independent both of n and t . Once again we observe that J and the v_i ’s are independent symmetric random variables. Using the properties exposed in Sec. 4 we obtain $\widehat{R}_{a,n}(t) = \sum_m c_m \widehat{R}_{a,n}(t; m)$, where

$$\widehat{R}_{a,n}(t; m) = \frac{1}{P'(1)} \mathbb{E}_J [(\tanh J)^{2m}] \sum_k P_k \{ \mathbb{E}_v [(\tanh v)^{2m}] \}^k = \quad (\text{D.17})$$

$$= \frac{1}{P'(1)} \mathbb{E}_J [(\tanh J)^{2m}] P(q_{2m}), \quad (\text{D.18})$$

where we defined $q_{2m} \equiv \mathbb{E}_v [(\tanh v)^{2m}] \in [-1, 1]$.

Finally, the same procedure is applied to $\widehat{R}_{b,n}(t)$. We obtain $\widehat{R}_{b,n}(t) = \sum_m c_m \widehat{R}_{b,n}(t; m)$ with

$$\widehat{R}_{b,n}(t; m) = \mathbb{E}_u [(\tanh u)^{2m}] q_{2m}. \quad (\text{D.19})$$

The next step consists in noticing that, because U and V are admissible, we can apply Eq. (6.1) to get

$$\mathbb{E}_u [(\tanh u)^{2m}] = \mathbb{E}_J [(\tanh J)^{2m}] \sum_k \rho_h \{ \mathbb{E}_v [(\tanh v)^{2m}] \}^{k-1} = \frac{1}{P'(1)} \mathbb{E}_J [(\tanh J)^{2m}] P'(q_{2m}) \quad (\text{D.20})$$

This identity allows us to rewrite Eqs. (D.15) and (D.19) in the form

$$R_{b,n}(t; m) = \frac{1}{P'(1)} \mathbb{E}_J [(\tanh J)^{2m}] P'(q_{2m}) \mathbb{E}_s \langle Q_{2m} \rangle, \quad (\text{D.21})$$

$$\widehat{R}_{b,n}(t; m) = \frac{1}{P'(1)} \mathbb{E}_J [(\tanh J)^{2m}] P'(q_{2m}) q_{2m}. \quad (\text{D.22})$$

All the series obtained are absolutely convergent because $c_m \sim m^{-2}$ as $m \rightarrow \infty$ and $|R_{.,n}(t; m)| \leq 1$. We can therefore obtain $R_n(t)$ by performing the sum in Eq. (D.1) term by term. We get

$$R_n(t) = \frac{1}{P'(1)} \sum_{m=1}^{\infty} c_m \mathbb{E}_J[(\tanh J)^{2m}] \mathbb{E}_t[\langle f(Q_{2m}, q_{2m}) \rangle_{t,*}], \quad (\text{D.23})$$

where

$$f(Q, q) \equiv P(Q) - P'(q)Q - P(q) + P'(q)q. \quad (\text{D.24})$$

Since we assumed $P(x)$ to be convex for $x \in [-1, 1]$, $f(Q, q) \geq 0$ for any $Q, q \in [-1, 1]$. This completes the proof.

E Derivative of the conditional entropy: multi-Poisson ensemble

Throughout this Section $t_* \in \{0, \dots, t_{\max} - 1\}$ and $s \in [0, \gamma]$ are fixed. Let us start by noticing that the expected conditional entropy with respect to the multi-Poisson has the structure (here we use the shorthand \underline{Y} for the received message, which in our formalism is in fact (\underline{Y}, \hat{Y})):

$$h_n = \frac{1}{n} \mathbb{E}_C H_n(\underline{X}|\underline{Y}) = \frac{1}{n} \mathbb{E}_y \mathbb{E}_{t_*-1} \mathbb{E}_{t_*|t_*-1} \mathbb{E}_{t_{\max}-1|t_*} H_n(\underline{X}|\underline{Y}). \quad (\text{E.1})$$

Here we denoted by $\mathbb{E}_{t_2|t_1}$, with $t_2 \geq t_1$ the expectation with respect to the rounds $t_1 + 1, \dots, t_2$ in the code construction, and by \mathbb{E}_{t_1} the unconditional expectation over the first t_1 rounds. Notice that parameter s enters uniquely in the state $\mathbb{E}_{t_*|t_*-1}$, and more precisely in the mean of the Poisson variables $\{m_k^{(t_*)}\}$ and $\{l_i(t_*)\}$. We can therefore apply Eq. (C.1) to the expression (7.1). We get

$$\begin{aligned} \frac{dh_n}{ds} &= \sum_k \frac{P_k}{P'(1)} \sum_{i_1 \dots i_k} \mathbb{E}_y \mathbb{E}_{t_*} w_{i_1} \dots w_{i_k} \underbrace{\left\{ \mathbb{E}_{t_{\max}-1|t_*}^{\{i_1 \dots i_k\}} \log_2 \mathcal{Z}_C(i_1 \dots i_k; \hat{y}) - \mathbb{E}_{t_{\max}-1|t_*} \log_2 \mathcal{Z} \right\}}_{(i)} \\ &\quad - \sum_{i \in \mathcal{V}} \mathbb{E}_z \mathbb{E}_{t_*} w_i \underbrace{\left\{ \mathbb{E}_{t_{\max}-1|t_*}^{\{i\}} \log_2 \mathcal{Z}_{\text{eff}}(i; z) - \mathbb{E}_{t_{\max}-1|t_*} \log_2 \mathcal{Z} \right\}}_{(ii)} - \\ &\quad - \frac{1}{P'(1)} \sum_y Q_C(y|0) \log_2 Q_C(y|0) + \sum_z Q_{\text{eff}}(z|0) \log_2 Q_{\text{eff}}(z|0), \end{aligned} \quad (\text{E.2})$$

where we used the shorthand $w_i = w_i(t_*)$. The definition of the modified partition functions $\mathcal{Z}_C(i_1 \dots i_k; \hat{y})$ and $\mathcal{Z}_{\text{eff}}(i; z)$ is the same as in App. C. The resulting expression is here more complicated because the expectation over the stages $t_* + 1, \dots, t_{\max} - 1$ is not independent of the graph realization after stage t_* . For instance, if an extra check node is added during round t_* (as a result of Eq. (C.1)), the following check nodes are going to be added with a slightly different distribution in rounds $t_* + 1, \dots, t_{\max} - 1$. This fact is taken into account by defining the state $\mathbb{E}_{t_{\max}-1|t_*}^{\{i_1 \dots i_k\}}$ as follows. At the end of round t_* , set $\mathbf{d}_i(t_* + 1) = \mathbf{d}_i(t_*) - \Delta_i(t_*) - l_i(t_*) - \nu_i$ with ν_i equal to the number of times i appears in $\{i_1, \dots, i_k\}$. Then proceed as for the interpolating ensemble introduced in Sec. 7 for rounds $t_* + 1, \dots, t_{\max} - 1$.

We now decompose the underbraced terms in (E.2) as follows:

$$\begin{aligned} \left\{ (i) \right\} &= \left\{ \mathbb{E}_{t_{\max}-1|t_*}^{\{i_1 \dots i_k\}} \log_2 \mathcal{Z}_{\mathcal{C}}(i_1 \dots i_k; \hat{y}) - \mathbb{E}_{t_{\max}-1|t_*}^{\{i_1 \dots i_k\}} \log_2 \mathcal{Z} \right\} + \\ &+ \underbrace{\left\{ \mathbb{E}_{t_{\max}-1|t_*}^{\{i_1 \dots i_k\}} \log_2 \mathcal{Z} - \mathbb{E}_{t_{\max}-1|t_*} \log_2 \mathcal{Z} \right\}}_{F(i_1 \dots i_k)}, \end{aligned} \quad (\text{E.3})$$

and analogously for terms of type (ii) . It is now a matter of simple algebra to obtain Eq. (7.7), where (dropping the dependence of φ upon t_* and s)

$$\varphi(n) \equiv \mathbb{E}_{y,z} \mathbb{E}_{t_*} \left\{ \sum_k \frac{P_k}{P'(1)} \sum_{i_1 \dots i_k} w_{i_1} \dots w_{i_k} F(i_1 \dots i_k) - \sum_{i \in \mathcal{V}} w_i F(i) \right\}. \quad (\text{E.4})$$

We are now faced with the task of proving that $\varphi(n)$ is bounded as claimed in Sec. 7. Denote the quantity in parenthesis as $\hat{\varphi}(n)$. Notice that $|\hat{\varphi}(n)| \leq 2n$: in fact $\hat{\varphi}(n)$ is the difference among the entropies of two n -bits distributions.

First, we shall show that $\hat{\varphi}(n) \leq C\sqrt{(\log n)^3/n}$, under the hypothesis that $\sum_i \mathbf{d}_i(t_* + 1) \geq An$ for some positive constant A . Notice that the condition holds with high probability at least $1 - 2e^{-Bn}$, for some $B > 0$. The thesis follows from the inequality (hereafter we set $\mathbf{d}_i \equiv \mathbf{d}_i(t_* + 1)$)

$$|\varphi(n)| \leq \mathbb{P} \left[\sum_i \mathbf{d}_i \geq An \right] C\sqrt{\frac{(\log n)^3}{n}} + \mathbb{P} \left[\sum_i \mathbf{d}_i < An \right] 2n \leq C'\sqrt{\frac{(\log n)^3}{n}} \quad (\text{E.5})$$

We start by two simple observations which hold under the above condition.

1. There exist a constant $F_0 > 0$ such that $|F(i_1 \dots i_k)| \leq F_0$. F_0 is understood to depend on the ensemble parameters as well as on k , but not on n , t_* or s . This fact is proved by noticing that $F(i_1 \dots i_k)$ is the difference between the expectation values of $\log \mathcal{Z}$ in two different ensembles. These ensembles can be coupled as \mathcal{G}_* and \mathcal{G}_{mP} in App. A. Each time a new variable node must be chosen in the two graphs, choose it by coupling optimally the corresponding w_i distributions. The number of discrepancies obtained in this way is bounded: there is probability $O(1/n)$ of discrepancy (here the condition on $\sum_i \mathbf{d}_i$ is used) at each step and less than $n\Lambda'(1)$ steps. Finally the variation in $\log \mathcal{Z}$ produced by a single rewiring is smaller than 2 in absolute value.
2. Let i_1, \dots, i_k be i.i.d. with distribution $\{w_i\}$. There exist a constant w_0 such that the probability that any two of them coincide is smaller than w_0/n . This is proved by noticing that $w_i = [\mathbf{d}_i(t_*)]_+ / \sum_i [\mathbf{d}_i(t_*)]_+ \leq l_{\max}/An$ because of the above condition. Therefore, the probability of having coinciding indices is smaller than $k(k-1)l_{\max}/An$.

In a nutshell, these two remarks imply that terms with coincident indices give a contribution bounded by C/n in Eq. (E.4). Moreover, the first of these observation implies $|\phi(n)| \leq C_1$ uniformly in t_* and s , as claimed in Sec. 7.

We next rewrite the function $F(i_1 \dots i_k)$ by singling out the stage $t_* + 1$ in the code construction

$$\begin{aligned} F(i_1 \dots i_k) &= \mathbb{E}_{t_*+1|t_*}^{\{i_1 \dots i_k\}} [\mathbb{E}_{t_{\max}-1|t_*+1} \log_2 \mathcal{Z}] - \mathbb{E}_{t_*+1|t_*} [\mathbb{E}_{t_{\max}-1|t_*+1} \log_2 \mathcal{Z}] \equiv \\ &\equiv \mathbb{E}_{t_*+1|t_*}^{\{i_1 \dots i_k\}} \Psi(j_1 \dots j_m) - \mathbb{E}_{t_*+1|t_*} \Psi(j_1 \dots j_m). \end{aligned} \quad (\text{E.6})$$

Here $\Psi(\dots)$ denotes the quantity in square brackets in the previous line, and we made explicit its dependency upon the variable nodes chosen during stage $t_* + 1$. Notice that $j_1 \dots j_m$ are i.i.d.'s with distributions

$$\hat{w}_j^{\{i_1 \dots i_k\}} = \frac{[\mathbf{d}_j - \nu_j]_+}{\sum_i [\mathbf{d}_i - \nu_i]_+}, \quad \hat{w}_j = \frac{[\mathbf{d}_j]_+}{\sum_i [\mathbf{d}_i]_+}, \quad (\text{E.7})$$

where ν_j is (as above) the number of times j appears in $\{i_1 \dots i_k\}$. Notice that \hat{w}_j is not the same as w_j , the former being computed in terms of the $\{\mathbf{d}_i(t_* + 1)\}$ while the latter depends upon $\{\mathbf{d}_i(t_*)\}$. The only property of $\Psi(j_1 \dots j_m)$ we shall need hereafter is that it concentrates exponentially when $j_1 \dots j_m$ are distributed according to \hat{w}_j . More precisely

$$\mathbb{P}[|\Psi - \mathbb{E}\Psi| \geq n\varepsilon] \leq Ae^{-nB\varepsilon^2}, \quad (\text{E.8})$$

for some positive constants A and B . This result is obtained by repeating the proof of Theorem 1 for the quantity $\Psi(j_1 \dots j_m)$.

Now, we use the general fact that, given two distributions $p(s)$ and $q(s)$, such that $p(s) = 0$ only if $q(s) = 0$, we can write

$$\mathbb{E}_q f(s) = \mathbb{E}_p [X(s) f(s)], \quad X(s) \equiv \frac{1}{z} \frac{q(s)}{p(s)}, \quad z \equiv \mathbb{E}_p \left[\frac{q(s)}{p(s)} \right]. \quad (\text{E.9})$$

Applying this general relation to Eq. (E.6), we get

$$F(i_1 \dots i_k) = \mathbb{E} [X_{\{i_1 \dots i_k\}}(j_1 \dots j_m); \Psi(j_1 \dots j_m)], \quad (\text{E.10})$$

$$X_{\{i_1 \dots i_k\}} \equiv \left(1 - \frac{[k]}{L}\right)^{-m} \prod_{a=1}^m \frac{[\mathbf{d}_{j_a} - \nu_{j_a}]_+}{[\mathbf{d}_{j_a}]_+}, \quad (\text{E.11})$$

where we denoted $L \equiv \sum_i [\mathbf{d}_i]_+$ and $[k] \equiv \sum_i ([\mathbf{d}_i]_+ - [\mathbf{d}_i - \nu_i]_+)$. Furthermore, we assumed $d_i > 0$. We denote by \mathcal{V}_+ the set of variable nodes satisfying this condition. Notice that $0 \leq [k] \leq k$. Moreover $0 \leq X_{\{i_1 \dots i_k\}} \leq C$ for some constant C (recall that $m = O(n)$) and $\mathbb{E}X_{\{i_1 \dots i_k\}} = 1$. In view of the remarks 1. and 2. above, we focus here that i_1, \dots, i_k are distinct. Under this assumption

$$\begin{aligned} X_{\{i_1 \dots i_k\}} &= \left(1 - \frac{k}{L}\right)^{-m} \left(1 - \frac{1}{l_{i_1}}\right)^{\sum_a \mathbb{1}_{j_a=i_1}} \dots \left(1 - \frac{1}{l_{i_r}}\right)^{\sum_a \mathbb{1}_{j_a=i_r}} = \\ &= (1 + \delta(n)) X_{i_1} \dots X_{i_k}, \end{aligned} \quad (\text{E.12})$$

$$X_i \equiv \left(1 - \frac{1}{L}\right)^{-m} \left(1 - \frac{1}{l_i}\right)^{\sum_a \mathbb{1}_{j_a=i}}, \quad (\text{E.13})$$

where $\mathbb{1}_A$ is the indicator function for the event A and $\delta(n)$ is a non random function, which can be bounded $|\delta(n)| \leq \delta_0/n$. Inserting into Eq. (E.10), and using observation 1 above, we get

$$F(i_1 \dots i_k) = (1 + \delta(n)) \mathbb{E} [X_{i_1} \dots X_{i_k}; \Psi] = \mathbb{E} [X_{i_1} \dots X_{i_k}; \Psi] + O(1/n). \quad (\text{E.14})$$

We can now plug this result in Eq. (E.4), to get

$$\begin{aligned} \hat{\varphi}(n) &= \sum_k \frac{P_k}{P'(1)} \sum_{i_1 \dots i_k} w_{i_1} \dots w_{i_k} \mathbb{E} [X_{i_1} \dots X_{i_k}; \Psi] - \sum_{i \in \mathcal{V}} w_i \mathbb{E} [X_i; \Psi] + O(1/n) = \\ &= \frac{1}{P'(1)} \mathbb{E} \{ [P(X) - P'(1)X]; \Psi \} + O(1/n) = \end{aligned} \quad (\text{E.15})$$

$$= \frac{1}{P'(1)} \mathbb{E} \{ f(X, \mathbb{E}X) (\Psi - \mathbb{E}\Psi) \} + O(1/n). \quad (\text{E.16})$$

Here $f(X, x)$ is defined as in Eq. (D.24) and we introduced the site average $X \equiv \sum_{i=1}^n w_i X_i$. Furthermore we used the fact that terms with coincident indices induce an error of order $O(1/n)$, and that $\mathbb{E}X = 1$. Since $f(X, x)$ is convex positive with $f(x, x) = 0$, we have

$$|\hat{\varphi}(n)| \leq C \mathbb{E} \{ (X - \mathbb{E}X)^2 (\Psi - \mathbb{E}\Psi) \} + O(1/n). \quad (\text{E.17})$$

Finally, we notice that X satisfies a concentration law of the form

$$\mathbb{P}[|X - 1| \geq \varepsilon] \leq A e^{-nB\varepsilon^2}. \quad (\text{E.18})$$

The proof is, once again, the same as for Theorem 1.

Using the expression (E.16) together with Eqs. (E.8), (E.18), and the fact that $\Psi \leq n$ we finally get

$$|\hat{\varphi}(n)| \leq C_2 n \varepsilon^3 + C_2 n n e^{-nB\varepsilon^2} + O(1/n) \leq C \sqrt{\frac{(\log n)^3}{n}}. \quad (\text{E.19})$$

where the last inequality follows by choosing $\varepsilon = \alpha \sqrt{\log n/n}$.

F Positivity of $R_n(t_*; s)$: multi-Poisson ensemble

We start as for the Poisson ensemble, by writing the remainder in the form

$$R_n(t_*, s) = R_{a,n}(t_*, s) - R_{b,n}(t_*, s) - \hat{R}_{a,n}(t_*, s) + \hat{R}_{b,n}(t_*, s), \quad (\text{F.1})$$

where (to lighten formulae, entropies will be measured in nats in this Appendix)

$$R_{a,n}(t_*, s) = \sum_k \frac{P_k}{P'(1)} \sum_{i_1 \dots i_k} \mathbb{E}_{\hat{y}} \mathbb{E}^{\{i_1 \dots i_k\}} w_{i_1} \dots w_{i_k} \log \left\langle \frac{2 Q_{\mathcal{C}}(\hat{y} | x_{i_1} \oplus \dots \oplus x_{i_k})}{Q_{\mathcal{C}}(\hat{y} | 0) + Q_{\mathcal{C}}(\hat{y} | 1)} \right\rangle, \quad (\text{F.2})$$

$$R_{b,n}(t_*, s) = \sum_i \mathbb{E}_z \mathbb{E}^{\{i\}} w_i \log \left\langle \frac{2 Q_{\text{eff}}(z | x_i)}{Q_{\text{eff}}(z | 0) + Q_{\text{eff}}(z | 1)} \right\rangle. \quad (\text{F.3})$$

Analogous expressions hold for $\hat{R}_{a,n}(t_*, s)$, $\hat{R}_{b,n}(t_*, s)$ with the conditional measure $\langle \cdot \rangle$ substituted by the product measure $P_{v_1}(x_1) \dots P_{v_n}(x_n)$. Here we set $w_i = w_i(t_*)$ as in the previous Section, and we use the notation $\mathbb{E}^{\{i_1 \dots i_k\}}$ introduced there.

The treatment of the four terms $R_{a,n}(\cdot)$, $R_{b,n}(\cdot)$, $\hat{R}_{a,n}(\cdot)$, $\hat{R}_{b,n}(\cdot)$ parallels closely the calculations in App. D. Here we limit ourself to discussing $R_{a,n}(\cdot)$: this should be more than sufficient for understanding the necessary changes with respect to the Poisson case. As in that case, we use (3.9) and (4.3) to write

$$R_{a,n}(t_*, s) = \sum_k \frac{P_k}{P'(1)} \sum_{i_1 \dots i_k} \mathbb{E}_J \mathbb{E}^{\{i_1 \dots i_k\}} w_{i_1} \dots w_{i_k} \log [1 + \tanh J \tanh \ell_{i_1 \dots i_k}] = \quad (\text{F.4})$$

$$= \sum_k \frac{P_k}{P'(1)} \sum_{i_1 \dots i_k} \mathbb{E}_J \mathbb{E} \{ w_{i_1} \dots w_{i_k} X_{\{i_1 \dots i_k\}} \log [1 + \tanh J \tanh \ell_{i_1 \dots i_k}] \} = \quad (\text{F.5})$$

$$= \sum_k \frac{P_k}{P'(1)} \sum_{i_1 \dots i_k} \mathbb{E}_J \mathbb{E} \{ w_{i_1} X_{i_k} \dots w_{i_k} X_{i_k} \log [1 + \tanh J \tanh \ell_{i_1 \dots i_k}] \} + O(1/n).$$

In passing from Eq. (F.4) to Eq. (F.5), we used the general identity $\mathbb{E}^{\{i_1 \dots i_k\}}[A] = \mathbb{E}[X_{\{i_1 \dots i_k\}}A]$, where $X_{\{i_1 \dots i_k\}}$ is defined in Eq. (E.11). We then used Eq. (E.12) to approximate $X_{\{i_1 \dots i_k\}}$ with an error of order $O(1/n)$.

Now we can Taylor expand the logarithm as in App. D. We obtain

$$R_{a,n}(t_*, s) = \sum_{m=1}^{\infty} c_m R_{a,n}(t_*, s; m), \quad (\text{F.6})$$

where $c_m = 1/(2m - 1) - 1/2m$ and

$$R_{a,n}(t_*, s; m) \equiv \frac{1}{P'(1)} \mathbb{E}_J[(\tanh J)^{2m}] \mathbb{E}_s \langle P(Q_{2m}) \rangle_{* \dots} \quad (\text{F.7})$$

The unique difference with respect to Poisson ensemble is in the definition of ‘multi-overlaps’. Now in fact we have (with an abuse we use the same notation as for the simple Poisson ensemble):

$$Q_{2m}(\underline{\sigma}^{(1)}, \dots, \underline{\sigma}^{(2m)}) = \sum_{i=1}^n w_i X_i \sigma_i^{(1)} \dots \sigma_i^{(2m)}. \quad (\text{F.8})$$

Notice that we no longer have $Q_{2m} \in [-1, +1]$ because of the terms X_i . However Eq. (E.13) implies $|X_i| \leq \exp(m/L)$. Recall that m is the number of variable nodes chosen during stage $t_* + 1$, which is exponentially concentrated around its expectation $n\gamma$. On the other hand $L = \sum_i [d_i]_+ \geq \sum_i d_i$, and the last quantity is exponentially concentrated around $n\gamma(t_{\max} - t_*) \geq n\gamma$. Therefore, for any $\delta > 0$, we have $|X_i| \leq e(1 + \varepsilon)$ for any $i \in [n]$ with high probability. As a consequence $|Q_{2m}| \leq e(1 + \varepsilon)$ with high probability.

The other terms in Eq. (F.1) are treated analogously. We finally get

$$R_n(t_*, s) = \frac{1}{P'(1)} \sum_{m=1}^{\infty} c_m \mathbb{E}_J[(\tanh J)^{2m}] \mathbb{E}_t[\langle f(Q_{2m}, q_{2m}) \rangle], \quad (\text{F.9})$$

with $f(X, x)$ defined as in (D.24). Positivity follows from the assumption that $P(x)$ is convex in $[-e(1 + \varepsilon), e(1 + \varepsilon)]$.

References

- [1] R. G. Gallager, *Low Density Parity-Check Codes* (MIT Press, Cambridge, MA, 1963)
- [2] M. G. Luby, M. Mitzenmacher, A. Shokrollahi, and D. A. Spielman, “Efficient Erasure Codes.” *IEEE Trans. on Inform. Theory*, **47**, 569-584 (2001)
- [3] T. Richardson and R. Urbanke, “The capacity of low-density parity check codes under message-passing decoding.” *IEEE Trans. Inform. Theory*, **47**, 599–618 (2001).
- [4] T. Richardson and R. Urbanke, “An Introduction to the Analysis of Iterative Coding Systems”, in *Codes, Systems, and Graphical Models*, edited by B. Marcus and J. Rosenthal (Springer, New York, 2001).
- [5] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman. “Improved Low-Density Parity-Check Codes Using Irregular Graphs.” *IEEE Trans. on Inform. Theory*, **47** 585-598 (2001)

- [6] S.-Y. Chung, G. D. Forney, Jr., T. J. Richardson and R. Urbanke, *IEEE Comm. Letters*, “On the design of low-density parity-check codes within 0.0045 dB from the Shannon limit,” **5**, 58–60 (2001).
- [7] S. Aji, H. Jin, A. Khandekar, R. J. McEliece, and D. J. C. MacKay. “Thresholds for Code Ensembles based on ‘Typical pairs’ decoding. ” *Proceedings of IMA meeting 1999*
- [8] G. Miller and D. Burshtein, *IEEE Trans. Inform. Theory*, **47**, 2696-2710, (2001).
- [9] D. Burshtein, M. Krivelevich, S. Litsyn and G. Miller, “Upper Bounds on the Rate of LDPC Codes”, *IEEE Trans. Inform. Theory*, **48**, 2437-2449, (2002).
- [10] S. Shamai and I. Sason, “Variations on the Gallager bounds, connections and applications,” *IEEE Trans. on Information Theory*, **48**, p. 3029 (2002)
- [11] H. Pishro-Nik and F. Fekri. “On Decoding of Low-Density Parity-Check Codes over the Binary Erasure Channel”. *IEEE Trans. Inform. Theory*, **50**, 439-454, (2004).
- [12] M. Mezard, G. Parisi and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
- [13] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians* (Springer, New York, 2003)
- [14] N. Sourlas. “Spin-glass models as error-correcting codes.” *Nature* **339**, 693-694 (1989)
- [15] N. Sourlas, “Statistical Mechanics and Error-Correcting Codes”, in *Statistical Mechanics of Neural Networks Lecture Notes in Physics* **368**, edited by L. Garrido (Springer, New York, 1990).
- [16] N. Sourlas, “Statistical Mechanics and Error-Correcting Codes” in *From Statistical Physics to Statistical Inference and Back*, edited by P. Grassberger and J.-P. Nadal (Kluwer Academic, Dordrecht, 1994).
- [17] Y. Kabashima, T. Murayama, and D. Saad, “Typical Performance of Gallager-Type Error-Correcting Codes,” *Phys. Rev. Lett.* **84**, 1355 (2000)
- [18] T. Murayama, Y. Kabashima, D. Saad, and R. Vicente, “Statistical Physics of Regular Low-Density Parity-Check Error-Correcting Codes,” *Phys. Rev. E* **62**, 1577 (2000)
- [19] A. Montanari. “The Glassy Phase of Gallager Codes,” *Eur. Phys. J. B* **23**, 121 (2001)
- [20] S. Franz, M. Leone, A. Montanari, and F. Ricci-Tersenghi, “Dynamic phase transition for decoding algorithms”, *Phys. Rev. E* **66**, 046120 (2002)
- [21] A. Montanari and N. Sourlas, “The Statistical Mechanics of Turbo Codes,” *Eur. Phys. J. B* **18**, 107-119 (2000).
- [22] A. Montanari, “Turbo Codes: the Phase Transition.” *Eur. Phys. J. B* **18**, 121-136 (2000).
- [23] F. Guerra. “Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model,” *Commun. Math. Phys.* **233**, 1 (2003)
- [24] F. Guerra nad F. L. Toninelli, “The Thermodynamic Limit in Mean Field Spin Glass Models” *Commun. Math. Phys.* **230** (2002) 1, 71-79
- [25] S. Franz and M. Leone. “Replica Bounds for Optimization Problems and Diluted Spin Systems,” *J. Stat. Phys.* **111**, 535 (2003)
- [26] S. Franz, M. Leone and F.L. Toninelli. “Replica Bounds for Diluted Non-Poissonian Spin Systems.” *J. Phys. A: Math Gen* **36** (2003): 10967

- [27] M. Luby, “LT-codes,” Proceedings of FOCS-43, 271, 2002
- [28] R. M. Tanner, “A Recursive Approach to Low Complexity Codes.” IEEE Trans. Inform. Theory, **27**, 533-547 (1981).
- [29] M. Luby, M. Mitzenmacher, A. Shokrollahi, D.A. Spielman and V. Stemann, “Practical Loss-Resilient Codes”, Proceedings of STOC-29, 1997, 150-159.
- [30] T. Richardson and R. Urbanke, *Modern Coding Theory*, draft available at <http://lthcwww.epfl.ch/index.php>
- [31] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm”, IEEE Trans. Inform. Theory **47**, 498 (2001).
- [32] J. Pearl, *Probabilistic reasoning in intelligent systems: network of plausible inference* (Morgan Kaufmann, San Francisco, 1988).
- [33] T. Richardson, A. Shokrollahi, and R. Urbanke. “Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes”. IEEE Trans. Inform. Theory, **47**, 619-637 (2001).
- [34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, (Wiley, New York, 1991).
- [35] K. Azuma, “Weighted sums of certain dependent random variables”, Tohoku Mathematical Journal, **19** (1967), 357–354.
- [36] C. Di, T. Richardson, and R. Urbanke, “Weight Distribution of Low-Density Parity-Check Codes”. IEEE Trans. Inform. Theory, *submitted*, 2004. Available at <http://lthcwww.epfl.ch/index.php>.
- [37] C. Measson and R. Urbanke. “Further Analytic Properties of EXIT-Like Curves and Applications.” Proceedings of ISIT-2003.
- [38] C. Measson, A. Montanari and C. Measson. “Maxwell Construction: The Hidden Bridge between Maximum-Likelihood and Iterative Decoding”. Proceedings of ISIT-2004.
- [39] J. S. Yedidia, W. T. Freeman, and Y. Weiss, “Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms”, IEEE Trans. Inform. Theory, *submitted*.
- [40] D. Aldous, and J. A. Fill, “Reversible Markov chains and random walks on graphs”. Available at www.stat.berkeley.edu/users/aldous/book.html
- [41] C. Measson, A. Montanari, T. Richardson, and R. Urbanke. “Life Above Threshold: From List Decoding to Area Theorem and MSE” 2004 IEEE Information Theory Workshop, San Antonio, Texas. Available at <http://arxiv.org/abs/cs.IT/0410028>