Spatial Coupling as a Proof Technique

Andrei Giurgiu, Nicolas Macris and Rüdiger Urbanke
School of Computer and Communication Sciences,
EPFL, Lausanne, Switzerland
{andrei.giurgiu, nicolas.macris, rudiger.urbanke}@epfl.ch

***Abstract*—The aim of this paper is to show that spatial coupling can be viewed not only as a means to build better graphical models, but also as a tool to better understand uncoupled models. The starting point is the observation that some asymptotic properties of graphical models are easier to prove in the case of spatial coupling. In such cases, one can then use the so-called interpolation method to transfer results known for the spatially coupled case to the uncoupled one.**

Our main application of this framework is to LDPC codes, where we use interpolation to show that the average entropy of the codeword conditioned on the observation is asymptotically the same for spatially coupled as for uncoupled ensembles. We use this fact to prove the so-called Maxwell conjecture for a large class of ensembles.

In a first paper last year, we have successfully implemented this strategy for the case of LDPC ensembles where the variable node degree distribution is Poisson. In the current paper we now show how to treat the practically more relevant case of general variable degree distributions. In particular, regular ensembles fall within this framework. As we will see, a number of technical difficulties appear when compared to the simpler case of Poissondistributed degrees.

For our arguments to hold we need symmetry to be present. For coding, this symmetry follows from the channel symmetry; for general graphical models the required symmetry is called Nishimori symmetry.

I. Introduction

Spatially coupled codes were introduced in [1] under the name of convolutional LDPC codes. It was recently proved in [2] that spatial coupling can be used as a paradigm to build graphical models on which belief-propagation algorithms perform essentially optimally. The list of applications of this paradigm has expanded in the past years, to include coding and compressed sensing, to name two of the most important ones (see [2] for a review of history and references). But spatial coupling can also become useful in a different way: as a theoretical tool that improves understanding of uncoupled systems. More specifically, sometimes it is easier to prove that (i) a property of a graphical model holds under spatial coupling than for the uncoupled version. If that is the case, and if (ii) the coupled and the uncoupled scenarios are equivalent with respect to that property, then we obtain a proof that the uncoupled graphical system has the said property.

In this paper we prove a statement of type (ii) in the case of LDPC codes. Namely, we prove that the conditional entropy in the infinite blocklength limit is the same for the coupled and uncoupled versions of the code. This enables us to derive the equality of the MAP thresholds for coupled and uncoupled codes and allows us to conclude that the Maxwell Conjecture [3] (a result of type (i), which we already know holds for coupled ensembles) also holds for uncoupled systems. Our treatment is general enough to provide a recipe for similar results for many types of graphical models that exhibit socalled Nishimori symmetry (of which channel symmetry is a special case). Using the freshly-proven Maxwell Conjecture, we can derive two results that hold above the MAP threshold: the equality of the BP and MAP GEXIT curves and the exactness of the replica-symmetric formula.

Our proof succeeds by using the interpolation method, which was introduced in statistical physics by Guerra and Toninelli for the Sherrington-Kirkpatrick spin glasses [4] and gradually found its way to constraint satisfaction problems [5]–[7] and coding theory [8], [9]. The version we use here employs a discrete interpolation between the coupled and two versions of the uncoupled scenarios. An error-tolerating version of the superadditivity lemma is also borrowed from Bayati et al. [7] to show that the conditional entropy has a limit for large blocklengths (called *thermodynamic limit* limit in physics terminology).

The purpose of this paper is to extend the *proof of concept* presented at ISIT 2012 [10] to arbitrary variable-node degree distributions. The technique presented there was only amenable to ensembles with Poisson-distributed degrees, whose range of applicability in coding is limited. This is due to the occurence of nodes of very small degrees in significant proportions, which limits the performance. In what follows, we remove this technical barrier and allow for a wide choice of degree distributions, including regular graphs. However, we keep the restrictions (see [10]) that the check node degrees have to be even and that the channel must be symmetric. The core of the proof rests on the interplay of symmetry and evenness. A summary appeared in ISIT 2013 [11].

II. Preliminaries

A. Simple ensembles

We start by describing a simple ensemble of codes, which we call $\mathrm{LDPC}(N, \Lambda, K)$, where N is the number of variable nodes, $\Lambda(x) = \sum_{d\geq 0} \Lambda_d x^d$ is the variable-node degree distribution, and the integer K is the fixed check-node degree. The distribution Λ must be supported on a finite subset of the positive integers. The average with respect to this distribution will be denoted by $\bar{d}$. In case Λ is supported on a single value, we will call the ensemble *regular*. Next, for each of the N variable nodes, the *target degree* is drawn i.i.d. from Λ, and each variable node is labeled with that many *sockets*.

The purpose of a socket is to receive at most one edge from a check node, and all edges must be connected to sockets on the variable-node side. The number of sockets D will thus be a random variable which concentrates around $N\bar{d}$.

The check nodes and the connections are placed in the following way: As long as there are at least K free sockets (initially all sockets are free), add one new check node connected to K free sockets chosen uniformly at random, without replacement. The chosen sockets then become occupied. The final number of check nodes that are added is exactly $\lfloor D/K \rfloor$. Note that there will be at most $K-1$ unconnected sockets at the end of this process, so the resulting variable node degrees will not in general match the target degrees. However, we will be interested in the limit $N \to \infty$, where the distribution of the resulting degrees matches Λ.

B. Coupled ensembles

Intuitively, a coupled ensemble $\mathrm{LDPC}(N, L, w, \Lambda, K)$ consists of a number L of copies of a simple ensemble, with interaction between copies allowed, in the sense that a check node can be connected to nodes in neighboring copies. More precisely, the variable nodes are distributed into L groups, which lie on a closed circular chain. The positions are indexed by integers modulo L, and we employ the set of representatives $\{1, \dots, L\}$. It will be useful later to also consider open-ended chains (this is in fact the case for spatial coupling used in applications).

Just as for simple ensembles, each node is assigned a number of sockets drawn i.i.d. from the distribution Λ. The check nodes, however, are restricted in the following way: they are only allowed to connect to sockets whose positions lie inside an interval (called *window*) of length w somewhere on the chain, i.e. there exists a position z such that all edges are connected to nodes at positions $z, z+1, \dots, z+w-1$. As before, check nodes have degree K, and they are sampled as follows: first choose a window uniformly at random, then for each edge, choose a position uniformly and i.i.d. inside that window, and then choose uniformly a free socket at that position. In case there are no free sockets in the chosen position, the process is stopped. Note that it is possible to stop with a lot of empty sockets in the chain: for example in a very unlucky case, the same position might be picked all the time. However, with high probability, only a small number of sockets will be free at the end of the process, and it is easy to see that in the limit where $N \to \infty$ the rate of the code only depends on $\bar{d}$ and K. The steps in this process will be described in more detail in Section V.

Note that the ensembles described so far are built in two stages: first the vertices are allotted a number of empty sockets, which is determined by sampling from the distribution Λ, thereby establishing the *configuration pattern*; in the second stage, the edges of the graph are connected to free sockets in the configuration pattern. It will be sometimes helpful to separate the two stages and start at the place where the configuration pattern is already given.

This is a good place to observe that the cases where $w = 1$ and $w = L$ yield instances of the single ensemble in the following ways: for $w = 1$, there are L different, non-interacting copies of $\mathrm{LDPC}(N, \Lambda, K)$, whereas for $w = L$, the whole ensemble is equivalent to $\mathrm{LDPC}(NL, \Lambda, K)$, up to $O(\sqrt{N})$ missing check nodes.

The reader will notice that the ensemble we have just constructed is circular and thus the coupling chain has no boundaries. It is a boundary that makes available all the useful properties of LDPC codes like threshold saturation. We simply find it easier to work with the circular ensemble and we shall see later that we can add a boundary condition with little cost.

C. Graphical notation

Traditionally, the Tanner graph is pictured as a bipartite graph, with edges linking the variable nodes to the check nodes. Here we will consider an equivalent rendering, namely as a hypergraph, where the variable nodes are the only nodes, and check nodes correspond to K-ary hyperedges, i.e. K-tuples of variable nodes.

The check nodes have fixed even degree K, and we think of them as vectors $a = (a_1, \dots, a_K)$ of variable nodes, thereby incorporating the edge information in the graph. A code is then specified by the totality of check node connections corresponding to its Tanner graph. Thus, abusing a bit the standard terminology, we will say that a graph G is just a set of check constraints of the type $a = (a_1, \dots, a_K)$. In general we will use the letters a, b, c, ... to describe check constraints, u, v, ... to describe variable nodes, and G, $\tilde{G}$ G', ... to describe graphs.

D. Transmission over channel

We use these codes to transmit over a binary memoryless symmetric channel $p_{Y|X}(y|x)$, where the input symbol set is $\mathcal{X} = \{+1, -1\}$. For just one use of the channel, it is enough to consider the half-log-likelihood-ratios (HLLR) $h(y)$ instead of the actual outputs y, since they form a sufficient statistic. They are defined (bit-wise) as

$$h(y) = \frac{1}{2} \ln \frac{p_{Y|X}(y|+1)}{p_{Y|X}(y|-1)}, \tag{1}$$

with the possibility of taking infinite values. From $h(y)$ one can recover the posterior probability that the bit x was sent. The latter is easily seen to be proportional to $e^{h(y)x}$.

We now consider sending the whole input vector, which will be denoted usually by $\underline{\sigma} \in \mathcal{X}^V$, where V is the set of variable nodes. Instead of the outputs, we use the HLLRs $\underline{h} \in (\mathbb{R} \cup \{\pm\infty\})^V$, given by $h_v = h(y_v)$, where y is the output vector.

The posterior probability that the codeword $\underline{\sigma}$ was sent is proportional to $e^{\underline{h} \cdot \underline{\sigma}}$, where $\underline{h} \cdot \underline{\sigma}$ stands for the dot product $\sum_{v \in V} h_v \sigma_v$. The full expression for the posterior probability is given by

$$\mu(\sigma) = \frac{e^{\underline{h} \cdot (\underline{\sigma} - \underline{1})} \prod_{a \in G} (1 + \sigma_a)/2}{Z}, \tag{2}$$

where σ_a is short for the product $\sigma_{a_1} \cdots \sigma_{a_K}$, and Z is a normalizing factor, also called *partition function*, given by

$$Z = \sum_{\sigma \in \mathcal{X}^V} e^{\underline{h} \cdot (\underline{\sigma} - \underline{1})} \prod_{a \in G} \frac{1 + \sigma_a}{2}.$$

One can easily check that the product $\prod_{a \in G}(1 + \sigma_a)/2$ is 1 when σ is any codeword, and 0 otherwise. The scaling provided by shifting $\underline{\sigma}$ by 1 downward helps to keep the values involved finite in the case $h = +\infty$. We will see shortly that the case $h = -\infty$ will never occur in our calculations, since by symmetry we can assume the codeword sent is the all-$+1$ codeword.

We have denoted the above probability measure by μ in order to distinguish it from other randomized parameters that appear, notably the channel and the randomness in the graph G. Note that μ depends on both G and the HLLRs $\underline{h}$, and when this is not clear we will make it explicit by adding G or $\underline{h}$ as a subscript: $\mu_{G,\underline{h}}$, $Z(G, \underline{h})$. This measure is akin to the *Gibbs measure* in Statistical Physics, and we will call it as such, in order to distinguish it from other measures.

The average with respect to the measure μ will appear quite often in the rest of the paper, and we use the *Gibbs brackets* $\langle \cdot \rangle$ to indicate it. In other words,

$$\langle f(\sigma) \rangle = \sum_{\sigma \in \mathcal{X}^V} f(\sigma) \mu(\sigma).$$

Regarding notation, the same subscript conventions apply as for μ apply for the bracket.

Because of symmetry, the channel is characterized by the distribution of the HLLR h computed from the output of the channel by (1) assuming the input of the channel is set to $+1$. We will view this distribution as a measure c on $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$, which due to channel symmetry has the property

$$d\mathsf{c}(-h) = d\mathsf{c}(h) e^{-2h}.$$

For this reason we call this property *symmetry of measures*, we denote all symmetric measures on $\overline{\mathbb{R}}$ by $\mathcal{X}$ and we identify $\mathcal{X}$ with the set of BMS channels. There is a partial ordering, called *degradation* defined on $\mathcal{X}$ which expresses the fact that one channel is better or worse with respect to another one. We say that a channel c_1 is degraded w.r.t. a channel c_2 and write $\mathsf{c}_1 \succ \mathsf{c}_2$ if there exists a third channel that can transform the output of c_2 (the better channel) into the output of c_1 (the worse channel). For properties of symmetric measures and alternative definitions of degradedness, we refer the reader to Chapter 4 of [3].

There are three types of randomness that are involved in our construction: (i) the random graph which is picked from an LDPC ensemble; (ii) the randomness induced by the channel and (iii) the Gibbs measure. The expectation in the first case is denoted by $\mathbb{E}_{G:\mathcal{G}}[\cdot]$, where $\mathcal{G}$ denotes the ensemble. The expectation with respect to the channel is written as $\mathbb{E}_h[\cdot] = \int \cdot d\mathsf{c}(h)$. As seen before, the average with respect to the Gibbs measure is denoted by angular brackets. The symbols $\mathbb{E}_{G:\mathcal{G}}$ and $\mathbb{E}_h$ commute, since the graph and the channel are independent. The angular bracket, however, depends on both h and the graph G and thus cannot commute with the $\mathbb{E}$ symbols. In the language of Statistical Physics, the graph and the channel are said to be quenched.

There is a deep and useful connection between $\ln Z(G)$ and the conditional entropy $H(\underline{X}|\underline{Y})$ (where $\underline{X}$ is the input vector and $\underline{Y}$ the output vector). In fact, in our case they are equal, because of the downward shift we added to $\underline{\sigma}$. We would like to express our results in terms of the latter, which carries more information-theoretic intuition, but we find it more natural to work with the former.

Lemma 1. *For a linear binary code of block length N represented by a graph G, we have*

$$H(\underline{X}|\underline{Y}) = \mathbb{E}_{\underline{h}} \ln Z(G, \underline{h}).$$

Proof: We use successively the definition of entropy, channel symmetry, and the fact that $\mu(\underline{1}) = Z^{-1}$:

$$\begin{aligned} H(\underline{X}|\underline{Y}) &= \sum_{\underline{x} \in \mathcal{C}(G)} \int \int d\underline{y} \prod_v p_{Y|X}(y_v|x_v) \log p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) \\ &= \int \int d\underline{y} \prod_v p_{Y|X}(y_v|\underline{1}) \log p_{\underline{X}|\underline{Y}}(\underline{1}|\underline{y}) \\ &= \int \int \prod_v d\mathsf{c}(h_v) \log \mu_{G,\underline{h}}(\underline{1}) \\ &= \mathbb{E}_{\underline{h}} \ln Z(G, \underline{h}), \end{aligned}$$

where $\mathcal{C}(G)$ is the set of codewords. ∎

III. Outline of the results

A. Comparison of entropies

We will set up the machinery of the interpolation method and direct it at proving the following theorem (for the proof, see Section VIII), which states that the entropies of the simple and coupled ensembles are asymptotically the same in the large N limit.

Theorem 2. *Let L, w, K be integers such that $L \geq w \geq 1$ and K is even and Λ be a degree distribution with finite support. Then for a fixed BMS channel we have*

$$\begin{aligned} &\lim_{N \to \infty} \frac{1}{N} \mathbb{E}_{G:\mathrm{LDPC}(N,\Lambda,K)} H(\underline{X}|\underline{Y}) = \\ &\qquad = \lim_{N \to \infty} \frac{1}{LN} \mathbb{E}_{G:\mathrm{LDPC}(N,L,w,\Lambda,K)} H(\underline{X}|\underline{Y}), \end{aligned} \tag{3}$$

and in particular the two limits exist.

Consider a family of channels $\{\mathsf{c}_\epsilon\}$ indexed by a parameter $\epsilon \in [\underline{\epsilon}, \overline{\epsilon}]$. Such a family is called *smooth* if for all continuously differentiable functions $f : \overline{\mathbb{R}} \to \mathbb{R}$ such that $f(h)e^h$ is bounded, its expectation $\int f(h) d\mathsf{c}_\epsilon(h)$ exists and is continuously differentiable with respect to ϵ in $[\underline{\epsilon}, \overline{\epsilon}]$.

A family of channels $\{\mathsf{c}_\epsilon\}$ is said to be *ordered by degradation* if $\mathsf{c}_\epsilon \prec \mathsf{c}_{\epsilon'}$ whenever $\epsilon < \epsilon'$. For a smooth family of channels ordered by degradation we may assume without loss of generality that the parameter ϵ is the channel entropy $H(Y_v|X_v)$, which varies between 0 (the perfect channel) and 1

(the useless channel). We will henceforth denote this universal parameter by h. The change of parameter is possible since the entropy h is given by the functional (12), which has the form $\int f(h)d\mathsf{c}_\epsilon(h)$ with $f(h) = \log_2(1+e^{-2h})$ and is by smoothness continuously differentiable as a function of ϵ.

Given a smooth family of channels ordered by degradation and parameterized by h in the whole interval $[0,1]$, there exists a value h_{MAP} (called *MAP threshold*) such that for channel parameters below this value, the scaled average conditional entropy (quantities of the kind appearing on both sides of (3)) converges to zero in the infinite block length limit, while above this value it is positive.

More formally, for the two kinds of LDPC ensembles, we define the MAP threshold in the following manner:

$$\mathrm{h}_{\text{MAP}} = \inf\left\{\mathrm{h} : \lim_{N\to\infty}\frac{1}{N}\mathbb{E}_{G:\text{LDPC}(N,\Lambda,K)}H(\underline{X}|\underline{Y}) > 0\right\},$$
$$\mathrm{h}_{\text{MAP}}^{L,w} = \inf\left\{\mathrm{h} : \lim_{N\to\infty}\frac{1}{NL}\mathbb{E}_{G:\underset{(N,L,w,\Lambda,K)}{\text{LDPC}}}H(\underline{X}|\underline{Y}) > 0\right\}.$$

These definitions usually employ $\liminf$ and are meaningful even when the existence of limits is not guaranteed. However, in our case, the existence of limits is part of the result of Theorem 2. The theorem obviously implies the equality of the two MAP thresholds.

Corollary 3. *With the same assumptions as in Theorem 2, we have* $\mathrm{h}_{MAP} = \mathrm{h}_{MAP}^{L,w}$.

B. The proof of the Maxwell Conjecture

As an application of this, we will prove the Maxwell conjecture for a large class of degree distributions. Let us recall the statement of the conjecture. Let h_{Area} be the area threshold defined as that value so that the integral of the BP-GEXIT curve over the interval $[\mathrm{h}_{\text{Area}}, 1]$ equals the design rate $1-\bar{d}/K$ (for more details, see [3]). The *Maxwell conjecture* states that $\mathrm{h}_{\text{Area}} = \mathrm{h}_{\text{MAP}}$.

The following was recently proved in [2]. For a large class of LDPC ensembles, if we consider the corresponding coupled ensemble, then the BP threshold (and hence, by threshold saturation, the MAP threshold) is very well approximated by h_{Area} (of the simple ensemble) in the following sense:

$$\mathrm{h}_{\text{Area}} - O(\frac{1}{w^{1/2}}) \le \mathrm{h}_{\text{MAP}}^{L,w,\text{open}} \le \mathrm{h}_{\text{Area}} + O(\frac{w}{L}). \tag{4}$$

The threshold $\mathrm{h}_{\text{MAP}}^{L,w,\text{open}}$ is the one of an *open* coupled chain, which is constructed such that the positions on the chain are from $\{1,\dots,L\}$, but the windows do not "wrap around". Instead we add *ghost* variable nodes at positions $-w+2,\dots,-1,0$ and $L+1,\dots,L+w-1$, whose input bits will always be fixed to $+1$. The windows are of the form $\{z,\dots,z+w-1\}$, where $z=-w+2,\dots,L$.

The only difference in the average conditional entropy of the open and closed chains comes from the check nodes that lie at the boundary of the chain. The proportion of these check-nodes is $O(w/L)$. By an application of the second statement in Lemma 9, the difference of the entropies is at most $O(w/L)$, which goes to 0 as $L\to\infty$. As a consequence,

$$\lim_{L\to\infty}\mathrm{h}_{\text{MAP}}^{L,w,\text{open}} = \lim_{L\to\infty}\mathrm{h}_{\text{MAP}}^{L,w}.$$

Thus by (4) and Corollary 3, we deduce that in fact h_{MAP} equals h_{Area}, by first taking the limit $L\to\infty$ and then $w\to\infty$. This completes the proof of the Maxwell conjecture for all those LDPC ensembles for which (4) is known.

C. Proof of the equality of the MAP- and the BP-GEXIT curves above the MAP threshold

In the rest of this section we will only work with uncoupled systems, so the ensemble over which we average is always $\text{LDPC}(N,\Lambda,K)$. Also, in order to make clear that the channel output depends on the channel entropy parameter h, we will write the former as $Y(\mathrm{h})$. The MAP-GEXIT function g^{MAP} is defined [12, Definitions 3 and 6] as

$$g^{\text{MAP}}(\mathrm{h}) = \limsup_{N\to\infty}\frac{1}{N}\,\mathbb{E}_G\left[\frac{d}{d\mathrm{h}}H(\underline{X}|\underline{Y}(\mathrm{h}))\right]. \tag{5}$$

The derivative $\frac{d}{d\mathrm{h}}H(\underline{X}|\underline{Y}(\mathrm{h}))$ for any binary linear code along a smooth BMS channel family parametrized by h is strictly increasing and takes values in the $[0,1]$ interval [13, Theorem 5.2, Corollary 5.1]. Thus the conditional entropy $H(\underline{X}|\underline{Y}(\mathrm{h}))$ is a convex function, so is its average over the code ensemble, and so is the scaled limit of these functions as $N\to\infty$. This also establishes the continuity of $\lim_{N\to\infty}\frac{1}{N}\,\mathbb{E}_G[H(\underline{X}|\underline{Y}(\mathrm{h}))]$ as a function of h.

We lower bound the area below g^{MAP} above the MAP threshold as follows:

$$\begin{aligned}
&\int_{\mathrm{h}_{\text{MAP}}}^1 g^{\text{MAP}}(\mathrm{h})d\mathrm{h} = \\
&= \int_{\mathrm{h}_{\text{MAP}}}^1 \left(\limsup_{N\to\infty}\frac{1}{N}\,\mathbb{E}_G\left[\frac{d}{d\mathrm{h}}H(\underline{X}|\underline{Y}(\mathrm{h}))\right]\right)d\mathrm{h} \\
&\overset{(a)}{\ge} \limsup_{N\to\infty}\int_{\mathrm{h}_{\text{MAP}}}^1 \frac{1}{N}\,\mathbb{E}_G\left[\frac{d}{d\mathrm{h}}H(\underline{X}|\underline{Y}(\mathrm{h}))\right]d\mathrm{h} \\
&\overset{(b)}{=} \lim_{N\to\infty}\left(\frac{1}{N}\mathbb{E}_G H(\underline{X}|\underline{Y}(1)) - \frac{1}{N}\mathbb{E}_G H(\underline{X}|\underline{Y}(\mathrm{h}_{\text{MAP}}))\right) \\
&\overset{(c)}{=} R - 0 = R,
\end{aligned} \tag{6}$$

where in step (a) we used the Reverse Fatou Lemma (note that the integrand on the r.h.s. is bounded), in step (b) we used the existence of limits provided by Theorem 2, and in step (c), since at $\mathrm{h}=1$ the channel is completely useless, we have that $H(\underline{X}|\underline{Y}(1)) = H(\underline{X})$ and $\frac{1}{N}\mathbb{E}_G[H(\underline{X})]$ is the rate of the code in the large blocklength limit; also for $\mathrm{h}=\mathrm{h}_{\text{MAP}}$, we have that $\lim_{N\to\infty}\frac{1}{N}\,\mathbb{E}_G[H(\underline{X}|\underline{Y}(\mathrm{h}))] = 0$ because of continuity.

The BP-GEXIT curve is defined [12, Definition 6] by

$$g^{\text{BP}}(\mathrm{h}) = \lim_{\ell\to\infty}\limsup_{N\to\infty}\frac{1}{N}\,\mathbb{E}_G\left[\sum_v g_{G,v}^{\text{BP}}(\mathrm{h})\right], \tag{7}$$

$$g_{G,v}^{\text{BP}}(\mathrm{h}) = \frac{\partial H(X_v|Y_v(\mathrm{h}_v),\Phi_v^\ell(\mathrm{h}))}{\partial \mathrm{h}_v}\Big|_{\mathrm{h}_v=\mathrm{h}}, \tag{8}$$

where $\Phi_v^\ell(\mathrm{h})$ is the BP estimate based on a computation tree of depth ℓ. An equivalent form is given by Equation (47) in the Appendix.

It is known that (see Lemma 9 in [12])

$$g^{\text{MAP}}(\mathrm{h}) \leq g^{\text{BP}}(\mathrm{h}), \text{ for all } \mathrm{h} \in [0,1]. \tag{9}$$

The area threshold mentioned before is defined as the solution h_{area} to the equation

$$\int_{\mathrm{h}_{\text{area}}}^{1} g^{\text{BP}}(\mathrm{h}) d\mathrm{h} = R. \tag{10}$$

Using then the equality of the MAP and area tresholds established in the previous subsection for the above-mentioned class of LDPC codes and using (6) and (10) we obtain

$$\int_{\mathrm{h}_{\text{MAP}}}^{1} (g^{\text{BP}}(\mathrm{h}) - g^{\text{MAP}}(\mathrm{h})) d\mathrm{h} \leq R - R = 0. \tag{11}$$

The positivity of the integrand (cf. (9)) entails the following result.

Theorem 4. *Given an* $\mathrm{LDPC}(N, \Lambda, K)$ *ensemble and a smooth family of channels indexed by the entropy parameter* h*, the two curves* g^{MAP} *and* g^{BP} *are equal almost everywhere above the MAP threshold.*

The discussion of (6) also entails the following result, which will be useful subsequently. Among others, this allows us to exchange the $\liminf$ with $\lim$ in the expresssion for the MAP threshold.

Proposition 5. *The limit* $\lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{\mathrm{LDPC}(N,\Lambda,K)} H(\underline{X}|\underline{Y}(\mathrm{h}))$ *exists for all values of* h*, and furthermore*

$$\int_{\mathrm{h}_0}^{1} g^{MAP}(\mathrm{h}) d\mathrm{h} = R - \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{\mathrm{LDPC}(N,\Lambda,K)} H(\underline{X}|\underline{Y}(\mathrm{h}_0)),$$

where $R = 1 - \Lambda'(1)/K$ *is the rate of the code.*

D. Exactness of the replica-symmetric functional

The previous result, namely the equality of the BP and MAP GEXIT curves, allows us to derive another useful identity. We can prove that under certain conditions (above the MAP threshold) the potential functional, also called replica-symmetric (RS) formula, is in fact equal to the conditional entropy $H(\underline{X}|\underline{Y})$. Note that while the former is a quantity derived by message passing, the latter is related to combinatorial optima. Also, unlike GEXIT curves, these quantities make sense already without considering the channel as part of a smooth family and thus appear to be more natural. This section contains a result which is very similar to Lemma 26 from [2]. We offer a somewhat different proof for the sake of completeness.

In order to define the potential functional, we need to introduce the density evolution operations. The beliefs that are transmitted during BP have distributions that are symmetric measures. We use two operations that act on measures, $\circledast, \boxtimes : \mathcal{X} \times \mathcal{X} \to \mathcal{X}$, which correspond to the operations carried on beliefs in BP: the measure $\mathsf{z}_1 \circledast \mathsf{z}_2$ is the distribution of the sum of two independent random variables with laws z_1 and z_2, respectively; the measure $\mathsf{z}_1 \boxtimes \mathsf{z}_2$ is the distribution of $\tanh^{-1}(\tanh h_1 \tanh h_2)$, where $h_1 \sim \mathsf{z}_1$ and $h_2 \sim \mathsf{z}_2$ are independent random variables. The operations can be generalized straightforwardly to apply to any finite signed measures. For a more complete exposition of the properties of the two operations $\circledast$ and $\boxtimes$, please refer to [14].

The two operations on measures are commutative, associative and bilinear with respect to addition of measures. By $\mathsf{z}^{\circledast n}$ we mean the $\circledast$-product of z with itself n times. Given a polynomial $\lambda(u) = \sum_{n=0}^{\deg\lambda} \lambda_n u^n$, we define $\lambda^{\circledast}(\mathsf{z}) = \sum_{n=0}^{\deg\lambda} \lambda_n \mathsf{z}^{\circledast n}$. The definitions of $\mathsf{z}^{\boxtimes n}$ and $\lambda^{\boxtimes}(\mathsf{z})$ are similar.

The conditional entropy $H(X|Y)$ of the input bit given the output is given in terms of the HLLR distribution c of the channel for any $\mathsf{c} \in \mathcal{X}$ by the linear functional

$$H(\mathsf{c}) = \int \log_2(1 + e^{-2h}) d\mathsf{c}(h). \tag{12}$$

We restrict ourselves now to regular LDPC ensembles with left and right degrees d_l and d_r, respectively. However, since much of the derivations hold more generally, we will work with the polynomials Λ, P and λ, ρ as left and right degrees from the node and from the edge perspective, respectively. For us, they take the simple forms $\lambda(u) = u^{d_l - 1}$, $\rho(u) = u^{d_r - 1}$, $\Lambda(u) = u^{d_l}$ and $P(u) = u^{d_r}$.

The density evolution (DE) equation can then be written as $\mathsf{x}^{\ell+1} = \mathsf{c} \circledast \lambda^{\circledast}(\rho^{\boxtimes}(\mathsf{x}^\ell))$. The fixpoint that can be reached by starting with $\mathsf{x}^0 = \Delta_0$ will be called *forward DE fixpoint* and will be denoted by x_c.

We are now ready to define the replica-symmetric functional, which depends on the channel c and the message density x as

$$\begin{aligned} U(\mathsf{x}, \mathsf{c}) = &-\frac{L'(1)}{R'(1)} H(R^{\boxtimes}(\mathsf{x})) - L'(1) H(\rho^{\boxtimes}(\mathsf{x})) \\ &+ L'(1) H(\mathsf{x} \boxtimes \rho^{\boxtimes}(x)) + H(\mathsf{c} \circledast L^{\circledast}(\rho^{\boxtimes}(\mathsf{x}))). \end{aligned} \tag{13}$$

The next theorem states that in a region of channels above the MAP threshold characterized by a regularity condition, this functional (which is algorithmic in nature as it comes from message passing) is equal to the conditional entropy, which is combinatorial in nature. In other words, the Bethe approximation of the entropy is exact in the said region.

To express the regularity constraint, we first define the region of channels above the MAP threshold:

$$\mathcal{C}_0 = \{\mathsf{c} \in \mathcal{X} : \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{\mathrm{LDPC}(N,d_l,d_r)} H(\underline{X}|\underline{Y}(\mathsf{c})) > 0.\}$$

Ideally, we would like our result to hold in the whole of this region, but, unfortunately, we need to add a Lipschitz type of restriction. Let

$$\begin{aligned} \mathcal{C}_1 = \big\{ &\mathsf{c}_0 \in \mathcal{X} : \text{there is } \delta > 0 \text{ s.t. for all } \mathsf{c}, \mathsf{c}' \in [\mathsf{c}_0, \Delta_0] \\ &\text{we have that } \left| \frac{\mathcal{B}(\mathsf{x}_\mathsf{c} - \mathsf{x}_{\mathsf{c}'})}{\mathcal{B}(\mathsf{c} - \mathsf{c}')} \right| \leq \frac{1}{\delta} \big\}, \end{aligned} \tag{14}$$

where $\mathcal{B}(\cdot)$ is the Bhattacharyya functional defined by (48), and

$$[\mathsf{c}_0, \mathsf{c}_1] = \{\mathsf{c} : \mathsf{c} = p\mathsf{c}_0 + (1-p)\mathsf{c}_1, \text{ for some } p \in [0,1]\}.$$

Note that the regions $\mathcal{C}_0$ and $\mathcal{C}_1$ depend on the parameters of the code.

Theorem 6. *Given the regular ensemble* $\mathrm{LDPC}(N, d_l, d_r)$ *with even* d_r*, for any channel* $\mathsf{c} \in \mathcal{C}_1 \cap \mathcal{C}_1$ *we have that*

$$U(\mathsf{x}_\mathsf{c}, \mathsf{c}) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{\mathrm{LDPC}(N,d_l,d_r)} H(\underline{X}, \underline{Y}(\mathsf{c})).$$

As the proof is fairly technical, we defer it to Appendix B. We show now that for large degree pairs, $\mathcal{C}_0 \subseteq \mathcal{C}_1$, i.e. the theorem holds everywhere above the MAP threshold. This is made precise by Lemma 18 from [2], reproduced below, which states that all channels with entropy above a value that goes to 0 as the right degree increases are in $\mathcal{C}_1$.

Lemma 7. *Let* d_l *and* d_r *be fixed numbers. There is a constant* [1] $\bar{\mathrm{h}}$ *depending only on the degrees* d_l *and* d_r *satisfying*

$$\bar{\mathrm{h}} < \frac{e^{1/4}\sqrt{2}}{d_r^{1/4}} \tag{15}$$

such that $\{\mathsf{c} \in \mathcal{X} : H(\mathsf{c}) > \bar{\mathrm{h}}\} \subseteq \mathcal{C}_1$.

We can readily see that for large degrees the right hand side of condition (15) approaches 0. Also, for large degrees, the MAP threshold approaches capacity and is bounded away from 0 uniformly for all channel families. This implies that $\mathcal{C}_0 \subseteq \mathcal{C}_1$. We also believe that the theorem remains true without this somewhat technical condition, but such a result remains unproven.

IV. Some useful lemmas

We present in this section two results that are quite general in nature, meaning that they are true for any linear code. They already appear in [8], [15], but we reproduce short proofs here in order to make the paper self-contained. The symmetry of the channel is a property that seems indispensable for the proofs in the rest of this paper, and we will need it in the form of the Nishimori Identity. The channel used for transmission needs to be BMS, symmetry being the crucial ingredient.

Lemma 8 (Nishimori Identity)**.** *Fix a graph* G *(no constraints on the check node degrees needed here) and a channel* $\mathsf{c} \in \mathcal{X}$*. For any odd positive integer* m *we have*

$$\mathbb{E}_h[\langle\sigma_b\rangle^m] = \mathbb{E}_h\left[\langle\sigma_b\rangle^{m+1}\right], \tag{16}$$

where $b = (b_1, \ldots, b_J)$ *is a vector of variables (which need not be interpreted as a check constraint) of arbitrary length, and* $\sigma_b = \sigma_{b_1} \cdots \sigma_{b_J}$.

Proof: We will assume here that the measure c does not contain mass at infinity. Extending to the general case can easily be done by considering the point mass at $+\infty$ separately. Because of symmetry, the measure defined by $d\mathsf{s}(\mathsf{h}) = e^{-h} d\mathsf{c}(h)$ has the property $d\mathsf{s}(h) = d\mathsf{s}(-h)$. Using the memoryless property of the channel, the l.h.s. of (16) can be written as

$$\mathbb{E}_h[\langle\sigma_b\rangle^m] = \int \langle\sigma_b\rangle^m \prod_{v\in V} e^{h_v} d\mathsf{s}(h_v). \tag{17}$$

We now observe that due to channel symmetry the above quantity is preserved under the transformation $h_v \mapsto h_v\tau_v$, $\sigma_v \mapsto \sigma_v\tau_v$, if τ is a codeword. As a matter of fact, the transformed HLLRs $h_v\tau_v$ are those received when the codeword τ was transmitted, instead of the all-$+1$ codeword.

We now perform an average over all codewords τ, obtaining

$$\mathbb{E}_h[\langle\sigma_b\rangle^m] = \frac{1}{|\mathcal{C}(G)|} \sum_{\tau\in\mathcal{C}(G)} \int \langle\sigma_b\tau_b\rangle^m \prod_{v\in V} e^{h_v\tau_v} d\mathsf{s}(h_v),$$

where $\mathcal{C}(G)$ is the set of all codewords.

Note that the Gibbs bracket above averages over σ, and thus we can safely take τ_b out of the bracket. Since m is odd, $\tau_b^m = \tau_b$. Next we use the definition of Gibbs measure (equation (2)) to replace $\sum_{\tau\in\mathcal{C}(G)} e^{h\cdot\tau}\tau_b$ with $Z(G)\langle\tau_b\rangle$. We obtain

$$\mathbb{E}_h[\langle\sigma_b\rangle^m] = \frac{1}{|\mathcal{C}(G)|} \int Z(G) \langle\sigma_b\rangle^{m+1} \prod_{v\in V} d\mathsf{s}(h_v). \tag{18}$$

Expanding $Z(G,h)$ into $\sum_{\lambda\in\mathcal{C}(G)} e^{h\cdot\lambda}$ we get

$$\mathbb{E}_h[\langle\sigma_b\rangle^m] = \frac{1}{|\mathcal{C}(G)|} \sum_{\lambda\in\mathcal{C}(G)} \int \langle\sigma_b\rangle^{m+1} \prod_{v\in V} e^{h_v\lambda_v} d\mathsf{s}(h_v).$$

A second gauge transformation $h_v \mapsto h_v\lambda_v$, $\sigma_v \mapsto \sigma_v\lambda_v$ allows us to cancel all λ factors, since $\lambda_v^2 = 1$. All $|\mathcal{C}(G)|$ terms in the sum are equal, so the expression simplifies to

$$\mathbb{E}_h[\langle\sigma_b\rangle^m] = \int \langle\sigma_b\rangle^{m+1} \prod_{v\in V} e^{h_v} d\mathsf{s}(h_v), \tag{19}$$

and thus the claim follows. ∎

The next result quantifies the effect on $\ln Z$ of one extra check node added to some general linear code. This is the main reason why we chose to work with $\ln Z$ instead of the conditional entropy.

Lemma 9. *Given any graph* G *and an additional check constraint* b*, we have that*

$$\mathbb{E}_h[\ln Z(G \cup b) - \ln Z(G)] = -\ln 2 + \sum_{r\in 2\mathbb{Z}_+} \frac{\mathbb{E}_h[\langle\sigma_b\rangle_G^r]}{r^2 - r}.$$

In particular, $-\ln 2 \le \ln Z(G \cup b) - \ln Z(G) \le 0$.

The second part of the statement shows that the contribution of one extra check node gives only a finite variation in $\ln Z$, and it turns out to be very useful for the cases where we need to show that two similar ensembles have log-partition functions that are asymptotically identical.

[1] An expression for $\bar{\mathrm{h}}$ can be found in Lemma 18 of [2].

Proof: Using the definition of the partition function $Z(G \cup b)$, we are able to write

$$Z(G \cup b) = \sum_{\underline{\sigma} \in \mathcal{X}^V} e^{\underline{h} \cdot (\underline{\sigma} - \underline{1})} \frac{1+\sigma_b}{2} \prod_{a \in G} \frac{1+\sigma_a}{2}$$
$$= Z(G) \left\langle \frac{1+\sigma_b}{2} \right\rangle_G .$$

Then $\ln Z(G \cup b) - \ln Z(G) = -\ln 2 + \ln(1 + \langle \sigma_b \rangle)$. Expanding the logarithm into power series, we obtain

$$\ln(1 + \langle \sigma_b \rangle) = \sum_{j \geq 1} \frac{(-1)^{j+1}}{j} \langle \sigma_b \rangle^j . \tag{20}$$

We now use the Nishimori Identities (Lemma 8) with $\mathbb{E}_h \left[\langle \sigma_b \rangle^{j-1} \right] = \mathbb{E}_h \left[\langle \sigma_b \rangle^j \right]$, for even j. This allows us to merge each odd-index term with the following term, proving the claim. ∎

Let us now analyze the terms of the form $\langle \sigma_b \rangle_G^r$ that appear in the last lemma. For this purpose, we will work with the product measure $\mu^{\otimes r}$. The measure space here is the one of r-tuples $(\sigma^{(1)}, \ldots, \sigma^{(r)})$, where $\sigma^{(j)} \in \mathcal{X}^V$. Because the product measure is just the measure of r independent copies of the measure (henceforth called *replicas*), it is easy to check that

$$\langle \sigma_b \rangle_G^r = \left\langle \sigma_b^{(1)} \cdots \sigma_b^{(r)} \right\rangle_G^{\otimes r} .$$

The $\otimes r$ sign at the top right of the bracket is just to remind us that we deal with the product measure $\mu^{\otimes r}$. Since this is evident from context, we will drop this sign in the future. We are then able to restate the last lemma as follows.

Corollary 10. *Given any graph G and an additional check constraint b, we have that*

$$\mathbb{E}_h[\ln Z(G \cup b) - \ln Z(G)] =$$
$$= -\ln 2 + \mathbb{E}_h \sum_{r \in 2\mathbb{Z}_+} \frac{\left\langle \sigma_b^{(1)} \cdots \sigma_b^{(r)} \right\rangle_G}{r^2 - r}. \tag{21}$$

V. The configuration model

In this section we introduce the language needed to describe and dissect all the kinds of ensembles that we need. We assume that the configuration pattern introduced in Section II-B is already fixed, i.e., it has been properly sampled at an earlier stage, and there are at least $N\bar{d}(1 - N^{-\eta})$ and at most $N\bar{d}(1 + N^{-\eta})$ sockets at every position. By a straightforward application of a Azuma-Hoeffding type of inequality and the union bound for all positions, this happens with high probability[2] in the first stage, as long as $0 < \eta < \frac{1}{2}$. The fixed underlying configuration pattern is always of the coupled kind, i.e., there are L groups of N variable nodes each; the simple kind will arise from the conditions $w = 1$ and $w = L$. Given the fixed configuration pattern, each variable node v has a *target degree* $d(v)$, and exactly $d(v)$ *sockets* numbered from 1 to $d(v)$. Given a socket s, let $\mathrm{var}(s)$ denote the variable node that it is part of; by σ_s we understand $\sigma_{\mathrm{var}(s)}$. Let $\mathrm{pos}(v)$ denote the position of the variable v, with the notation extending to sockets in the obvious manner: $\mathrm{pos}(s) = \mathrm{pos}(\mathrm{var}(s))$. We also set S to be the set of all sockets and put $S_z = \{s \in S : \mathrm{pos}(s) = z\}$, i.e. the set of sockets at a particular position.

Check nodes will connect to sockets, so a check node a will have the form of a K-tuple $(a_1, \ldots, a_K)$, where the components a_j are sockets. Note that the ordering of the edges leaving the check-node matters, so the check also "stores" this information. We say that a check node a has *type* $\alpha = (\alpha_1, \ldots, \alpha_K)$ if $\alpha_j = \mathrm{pos}(a_j)$, for all $1 \leq j \leq K$. In other words, the type records the positions of the variable nodes to which the check node a connects.

We now consider random types, of which there are three kinds that are important to us:

- **The connected random type.** This random type is uniformly distributed over the set of all L^K possible types. We denote this distribution by $\mathbf{conn}$.
- **The disconnected random type.** This type is uniformly distributed over the set of all types whose entries are all equal, i.e., types of the form $(z, z, \ldots, z)$. We denote this distribution by $\mathbf{disc}$.
- **The coupled random type.** We choose a position z uniformly at random and the result is a type uniformly distributed over the set of all types whose entries lie in the set $\{z, \ldots z + w - 1\}$. We denote this distribution by $\mathbf{coup}$.

We now define the *positional occupation vector* occ_α of a type α to be a vector whose z entry counts the number of occurences of position z in type α. As an example, if $K = 6$ and $\alpha = (1, 3, 2, 5, 1, 3)$ and assuming there are $L = 5$ positions, then $\mathrm{occ}_\alpha = (2, 1, 2, 0, 1)$.

Given a multiset of types Γ (a set of types where duplicates can appear), we extend the definition of the positional occupation vector to $\mathrm{occ}_\Gamma = \sum_{\alpha \in \Gamma} \mathrm{occ}_\alpha$. We call a multiset of types *m-admissible* if $\mathrm{occ}_\Gamma(z) \leq |S_z| - m$, for all positions z. In other words, an m-admissible set of types Γ ensures that there exists a graph G whose check constraints match one-to-one the types in Γ (we say that G is *compatible* with Γ), and in addition, there are at least m sockets at each position that remain free. We will also use the word *admissible* to mean 0-admissible. One should think about the multiset of types as being a kind of "pre-graph", where only the positions of the edges are decided, but not yet the actual sockets.

The random graph generated by an admissible multiset of types Γ is simply given by the uniform measure over all graphs that are compatible with Γ. To sample this random graph, the algorithm is as follows: start with the empty graph; for each type $\alpha = (\alpha_1, \ldots, \alpha_K)$ in the multiset Γ (the order is immaterial), pick *distinct* a_i uniformly at random from the free sockets at position α_i, and add check constraint $(a_1, \ldots, a_K)$ to the graph. We will use this check-generating procedure often, so we will say that check constraint a is chosen according to distribution $\nu(\alpha, G)$ that depends on the

[2]By *with high probability* we mean that the event in question happens with probability $1 - o(1/\mathrm{poly}(N))$. The parameters L and w are considered constant for this purpose

type α, and the part G of the graph that is already in place. Let B_α be the set of check constraints that are compatible with α and are connected to free sockets (sockets that do not appear in G). Note that a socket must never be used twice, so they are chosen without replacement. Then $\nu(\alpha, G)$ is the uniform measure on B_α.

We also trivially extend this definition to the case of a random graph generated by a *random* multiset of types. This latter random object will be typically a list of independent random types of one of the three kinds *connected*, *disconnected* and *coupled*. For the sake of precision, in case the multiset of types is not admissible (by this we mean m-admissible, where m will be fixed later), we define the generated random graph to be the empty one.

We now introduce a quantity inspired from Statistical Physics that plays an important role in what comes next, namely the *positional overlap functions*. Fix a configuration graph G, a channel realization h, and the number r of replicas of the measure $\mu_{G,h}$. Let $F_z \subseteq S_z$ be the set of free sockets at position z (free sockets being those that do not appear in any check constraint of G). The *positional overlap functions* Q_z, indexed by a position z, are defined by

$$Q_z(\sigma^{(1)}, \dots, \sigma^{(r)}) = \frac{1}{|F_z|} \sum_{s \in F_z} \sigma_s^{(1)} \cdots \sigma_s^{(r)}. \tag{22}$$

The next statement describes the link between the overlap functions and the replica averages introduced by Lemma 9.

Lemma 11. *Given a number $m > K^2$, a fixed channel realization, a fixed graph G whose associated type set is m-admissible and fixed type α, we have*

$$\mathbb{E}_{a:\nu(\alpha,G)} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle_G = \\ = \left\langle \prod_{j=1}^{K} Q_{\alpha_j}(\sigma^{(1)}, \dots, \sigma^{(r)}) \right\rangle + O\left(\frac{1}{m}\right). \tag{23}$$

Proof: The left hand side is nothing else than the average over all possible a that are compatible with the type α and connect to free sockets. In other words,

$$\frac{1}{|B_\alpha|} \sum_{a \in B_\alpha} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle. \tag{24}$$

The goal is to somehow factorize the sum, but the fact that sockets are not replaced makes it a bit harder. Suppose that, contrary to our current model, free sockets are allowed to be chosen with replacement, that is, it is possible to have $a_i = a_j$ for $i \neq j$. Let B'_α be the set of all (pseudo-)check constraints that are compatible with α, and where sockets are allowed to appear multiple times. Then B'_α can be written as a product:

$$B'_\alpha = F_{\alpha_1} \times \ldots \times F_{\alpha_K},$$

where the set F_z is the set of free sockets at position z. The idea is now that we can replace B_α with B'_α in the average (24) without losing too much, while gaining the ability to factorize the sum.

The relation between the two, which is proven in the Appendix A, is

$$\frac{1}{|B_\alpha|} \sum_{a \in B_\alpha} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle = \\ = \frac{1}{|B'_\alpha|} \sum_{a \in B'_\alpha} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle + O\left(\frac{1}{m}\right). \tag{25}$$

Now we are in a better position, since on the r.h.s. any entry a_i is chosen independently of the others. We rewrite the sum over B'_α in the following way:

$$\frac{1}{|F_{\alpha_1}|} \sum_{a_1 \in F_{\alpha_1}} \cdots \frac{1}{|F_{\alpha_K}|} \sum_{a_K \in F_{\alpha_K}} \left\langle \sigma_{a_1}^{(1)} \cdots \sigma_{a_K}^{(1)} \cdots \sigma_{a_1}^{(r)} \cdots \sigma_{a_K}^{(r)} \right\rangle.$$

Taking the bracket outside and factorizing, we obtain

$$\left\langle \left(\frac{1}{|F_{\alpha_1}|} \sum_{a_1 \in F_{\alpha_1}} \sigma_{a_1}^{(1)} \cdots \sigma_{a_1}^{(r)} \right) \cdots \left(\frac{1}{|F_{\alpha_K}|} \sum_{a_K \in F_{\alpha_K}} \sigma_{a_K}^{(1)} \cdots \sigma_{a_K}^{(r)} \right) \right\rangle,$$

which we can identify as the bracketed product of positional overlap functions on the right hand side of (23). ∎

Lemma 12. *Let G be a graph whose type multiset is m-admissible, and fix the channel realization h. Then the following inequalities hold:*

$$\mathbb{E}_{\substack{\alpha:\mathbf{conn} \\ a:\nu(\alpha,G)}} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle_G \leq \\ \leq \mathbb{E}_{\substack{\alpha:\mathbf{coup} \\ a:\nu(\alpha,G)}} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle_G + O(1/m), \tag{26}$$

$$\mathbb{E}_{\substack{\alpha:\mathbf{coup} \\ a:\nu(\alpha,G)}} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle_G \leq \\ \leq \mathbb{E}_{\substack{\alpha:\mathbf{disc} \\ a:\nu(\alpha,G)}} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle_G + O(1/m). \tag{27}$$

Proof: The claim follows by Lemma 11 if we manage to show the following two inequalities:

$$\mathbb{E}_{\alpha:\mathbf{conn}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle \leq \mathbb{E}_{\alpha:\mathbf{coup}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle, \tag{28}$$

$$\mathbb{E}_{\alpha:\mathbf{coup}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle \leq \mathbb{E}_{\alpha:\mathbf{disc}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle, \tag{29}$$

where the dependence of the positional overlap functions on the spin systems $\sigma^{(j)}$ has been dropped in order to lighten notation.

We rewrite the quantities above as follows:

$$\mathbb{E}_{\alpha:\mathbf{conn}}\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle = \\ = \frac{1}{L^K}\sum_{\substack{(\alpha_1,\ldots,\alpha_K)\\ \in[L]^K}}\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle = \left\langle\left(\frac{1}{L}\sum_{z\in[L]}Q_z\right)^K\right\rangle, \quad (30)$$

$$\mathbb{E}_{\alpha:\mathbf{coup}}\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle = \\ = \frac{1}{L}\sum_{z'\in[L]}\frac{1}{w^K}\sum_{\substack{(\alpha_1,\ldots,\alpha_K)\\ \in\{z',\ldots,z'+w-1\}^K}}\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle \\ = \left\langle\frac{1}{L}\sum_{z'\in[L]}\left(\frac{1}{w}\sum_{z=z'}^{z'+w-1}Q_z\right)^K\right\rangle, \quad (31)$$

$$\mathbb{E}_{\alpha:\mathbf{disc}}\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle = \\ = \frac{1}{L}\sum_{z\in[L]}\langle Q_z\cdots Q_z\rangle = \left\langle\frac{1}{L}\sum_{z\in[L]}Q_z^K\right\rangle. \quad (32)$$

Both inequalities (28) and (29) are proved by an application of Jensen's Inequality using the convexity of the function $x\mapsto x^K$, for even K. ∎

VI. The interpolation

We now move a bit further and consider random ensembles of graphs. These are obtained in the following way: first we prescribe the numbers of random types of each kind that we want, i.e. how many types should be connected, disconnected and coupled. Afterwards, the random types are sampled according to the distributions prescribed. Finally the graph is chosen uniformly to match the multiset of types, in the spirit of the previous section.

We use the notation $G:\left\{{t_1\times\mathbf{coup}\atop t_2\times\mathbf{disc}}\right\}$ to say that G is sampled in the way outlined above, where t_1 and t_2 are the number of random types of the coupled kind and disconnected kind, respectively. Of course, we could specify any combination of the three kinds, **conn** included.

Now we need to set the number of check nodes in the ensemble. There are two conflicting constraints we would like to satisfy: first, the set of types needs to be admissible with high probability — so that the sampled graph exists in the form we want; second, the number of free sockets that remain should be small, in the sense that the proportion of free sockets needs to vanish in the limit.

The average amount of check nodes needed to use all available sockets is (ideally) $NL\bar{d}/K$. However, there is a fluctuation ($\pm N^{1-\eta}\bar{d}$ at each position) of the amount of available sockets and it might not be possible to connect actual check nodes to all sockets (for example, because of window constraints). As a consequence, we choose the actual size of the graph (by this we mean the number of multi-edges, i.e. check nodes) to be $T = NL\bar{d}(1-N^{-\gamma})/K$, so in case the graph is admissible there will be $O(N^{1-\gamma})$ free sockets left at each position. The exponent γ is arbitrary, as long as $0<\gamma<\eta$. The next lemma confirms that by using this value for T, the resulting set of types is admissible with high probability.

Lemma 13. *Let $\alpha^1,\ldots,\alpha^T$ be random types, each drawn from a distribution that is either* **conn**, **disc** *or* **coup** *(could be different for each type). Then with high probability (more precisely $1-O(\exp(-\kappa N^{1-2\gamma}))$, for some positive constant κ) the resulting multiset of types is $\bar{d}N^{1-\gamma}/2$-admissible.*

Proof: The plan is the following: fix a position z, and show that the number of appearances of z as entries of $\alpha^1,\ldots,\alpha^T$ exceeds $TK/L+\bar{d}N^{1-\gamma}/2$ with a very small probability. Next, by the union bound over all positions z, we upper bound the probability that the graph is not $\bar{d}N^{1-\gamma}/2$-admissible and the lemma is proved.

We concentrate on the above claim, and define X_t to be the number of entries in α^t equal to z, for $1\le t\le T$. Clearly the X_t are independent, bounded and their expectation equals K/L (the choice of distribution of α^t is immaterial as long as it is one of **conn**, **disc** or **coup**). Then by Hoeffding's Inequality, the probability that $\sum X_t$ deviates from its expectation TK/L decays very fast. More exactly,

$$\mathbb{P}\left[\sum_{t=1}^{T}X_t\ge\frac{TK}{L}+\frac{1}{2}\bar{d}N^{1-\gamma}\right]\le\exp\left(-\frac{\bar{d}^2N^{2-2\gamma}}{2K^2T}\right), \quad (33)$$

which proves the claim. ∎

The previous lemma essentially allows us to take the expectation over an ensemble of graphs without caring too much about non-admissibility. We are now ready to prove a key result, expressed as the following lemma.

Lemma 14. *The following two inequalities hold:*

$$\mathbb{E}_{h,G:\{T\times\mathbf{conn}\}}\ln Z(G)\le \\ \le\mathbb{E}_{h,G:\{T\times\mathbf{coup}\}}\ln Z(G)+O\left(N^{\gamma}\right), \quad (34)$$

$$\mathbb{E}_{h,G:\{T\times\mathbf{coup}\}}\ln Z(G)\le \\ \le\mathbb{E}_{h,G:\{T\times\mathbf{disc}\}}\ln Z(G)+O\left(N^{\gamma}\right). \quad (35)$$

Proof: We only discuss the first of the two inequalities, since the proof of the other is identical. We will set up a chain of inequalities, at the ends of which sit the two quantities that we need to compare. This is the main idea of the *interpolation method*: finding a sequence of objects that transition "smoothly" between two objects that can differ significantly. In our case, it is easily seen that the claim follows if we are able to show that

$$\mathbb{E}_{h,G:\left\{{(t+1)\times\mathbf{conn}\atop(T-t-1)\times\mathbf{coup}}\right\}}\ln Z(G)\le \\ \le\mathbb{E}_{h,G:\left\{{t\times\mathbf{conn}\atop(T-t)\times\mathbf{coup}}\right\}}\ln Z(G)+O\left(N^{\gamma-1}\right). \quad (36)$$

The two ensembles involved in inequality (34) lie at the endpoints of a chain of T inequalities of the form above, with t moving from 0 to $T-1$. The crucial observation here is that the two ensembles $\left\{{(t+1)\times\mathbf{conn}\atop(T-t-1)\times\mathbf{coup}}\right\}$ and $\left\{{t\times\mathbf{conn}\atop(T-t)\times\mathbf{coup}}\right\}$ can both be obtained by sampling a graph $\widetilde{G}$ from their common part, $\left\{{t\times\mathbf{conn}\atop(T-t-1)\times\mathbf{coup}}\right\}$ and in case G is not null, adding an

extra random check constraint sampled according to $\mathbf{conn}$ and $\mathbf{coup}$, respectively. The plan is to show that the inequality (36) holds also when $\widetilde{G}$ is fixed, and then to average over $\widetilde{G}$.

Let us fix $m = \bar{d}N^{1-\gamma}/2$, and let us first deal with the case when the realization of the ensemble $\left\{ {}^{t\times\mathbf{conn}}_{(T-t-1)\times\mathbf{coup}} \right\}$ is not m-admissible. This event occurs with a very small probability, subexponential according to Lemma 13. Since $\ln Z(G) = O(N)$ (according to Lemma 9), the error obtained by not considering this case is extremely small and fits in the tolerated term $O\left(\frac{1}{N^{1-\gamma}}\right)$.

Otherwise, $\widetilde{G}$ is such that there are at least m free sockets at every position, and we need to show that

$$\mathbb{E}_h \mathbb{E}_{\substack{\alpha:\mathbf{conn} \\ a:\nu(\alpha,\widetilde{G})}} \ln Z(\widetilde{G} \cup a) \leq \mathbb{E}_h \mathbb{E}_{\substack{\alpha:\mathbf{coup} \\ a:\nu(\alpha,\widetilde{G})}} \ln Z(\widetilde{G} \cup a).$$

We substract $\ln Z(\widetilde{G})$ on both sides and then use Lemma 9 to write the difference of log partition functions as a linear combination of brackets of the form $\left\langle \sigma^{(1)} \cdots \sigma^{(r)} \right\rangle_{\widetilde{G}}$, after which we can readily apply Lemma 12 and the claim follows. ∎

VII. Retrieving the original LDPC ensembles

We will now investigate further the connection between the ensembles $\{T \times \mathbf{conn}\}$ and $\{T \times \mathbf{disc}\}$. In fact, they are both variants of the uncoupled ensembles introduced in the beginning of Section II. The first one is very similar to $\mathrm{LDPC}(NL, \Lambda, K)$, and the second one is similar to L copies of $\mathrm{LDPC}(N, \Lambda, K)$. The only differences that occur are related to the case where there is a large deviation in the number of sockets generated in the first stage, or when the multisets of types generated by $\{T \times \mathbf{conn}\}$ and $\{T \times \mathbf{disc}\}$ are not admissible. Also since the first stage of the ensemble generation, where we obtain the configuration pattern, is the same in all cases, we condition on the event that the configuration pattern is known and that it satisfies the condition stated at the beginning of Section V, namely that the number of sockets at each position is $N\bar{d}/K \pm O(N^{\eta})$.

We can easily see that the ensemble $\{T \times \mathbf{disc}\}$, conditioned on the fact that its realization is admissible, can be extended to L copies of the simple ensemble on N variable nodes by adding $O(N^{1-\gamma})$ extra check constraints. Thus the scaled log partition function is the same up to a sublinear term.

Can we say the same about the ensemble $\{T \times \mathbf{conn}\}$ and the simple ensemble on NL variable nodes? Yes, but it requires a lengthier argument. Let us look closer at the latter. This ensemble is not generated using types (since positions play no role here), but we can still count the occurrences of various types that appear in it. There are exactly L^K different types, and the next proposition estimates the probability that a particular random check constraint in the simple ensemble $\mathrm{LDPC}(NL, \Lambda, K)$ has a certain type. To see the crux of the problem, in the $\{T \times \mathbf{conn}\}$ ensemble, the types are generated uniformly. Whereas in the simple ensemble, a position with considerably more occupied sockets than other positions has a lesser chance to be picked.

We will proceed by transforming the ensemble $\mathrm{LDPC}(NL, \Lambda, K)$ (the *simple* ensemble) into $\{T \times \mathbf{conn}\}$ (the *connected* ensemble) through only a small amount of check additions and deletions. Let X_α be the number of check nodes of type α that occur in a realization of the simple ensemble. For every type α, let Y_α be a random variable sampled according to $\mathrm{Bin}(T, L^{-K})$. If $X_\alpha > Y_\alpha$, then exactly $X_\alpha - Y_\alpha$ check nodes of type α selected uniformly at random from the existing ones are deleted from the simple ensemble. Otherwise, exactly $Y_\alpha - X_\alpha$ check nodes of type α are chosen uniformly at random from all possible combinations of compatible free sockets and inserted in the graph without replacement. All insertions of check nodes must occur after all deletions have been performed (the order of the types is important). If at any stage there are no free sockets at a particular position to choose from, it just means the underlying multiset of types (which here is given by the numbers Y_α) is not T-admissible, and we produce the trivial code.

In order to bound the number of check node insertions and deletions, we compute the first and second moments of $X_\alpha - Y_\alpha$. The total number of check nodes M in the simple ensemble is fixed for our purposes (depends only on the configuration pattern), so we can write $X_\alpha = \sum_a R_\alpha^a$, where R_α^a is the indicator random variable of the event that check node a has type α, and the sum ranges over all M check nodes.

Proposition 15. *The expectation and variance of $X_\alpha - Y_\alpha$ are given by*

$$\mathbb{E}[X_\alpha - Y_\alpha] = O(N^{1-\gamma}), \tag{37}$$

$$\mathrm{Var}[X_\alpha - Y_\alpha] = O(N^{2-\eta}). \tag{38}$$

Proof: We determine first the probability $\mathbb{E}R_\alpha^a$ that a check node a has type α. This event happens if and only if all sockets a_i to which a is connected are placed at positions α_i. For this, we need to evaluate the proportion of free sockets at each position (all sockets are free initially, because w.l.o.g. we can say that a is the first check node to be allocated). The number of sockets at any position is between $N\bar{d}(1 - N^{-\eta})$ and $N\bar{d}(1 + N^{-\eta})$; the number of occupied sockets is at most $K - 1$ (from previous edges). Thus, the probability that $\mathrm{pos}(a_i) = \alpha_i$ is lower-bounded by

$$\frac{N\bar{d}(1 - N^{-\eta}) - K}{NL\bar{d}(1 + N^{-\eta})} = \frac{1}{L} - O(N^{-\eta}),$$

and, likewise, upper-bounded by

$$\frac{N\bar{d}(1 + N^{-\eta})}{NL\bar{d}(1 - N^{-\eta})} = \frac{1}{L} + O(N^{-\eta}).$$

It then follows that

$$\mathbb{E}R_\alpha^a = \left(\frac{1}{L} + O(N^{-\eta})\right)^K = \frac{1}{L^K} + O(N^{-\eta}). \tag{39}$$

For the second moments we need $\mathbb{E}\left[R_\alpha^a R_\beta^b\right]$, i.e. the probability that a and b have types α and β at the same

time. The reasoning is essentially similar to the previous case, only now there are $2K$ edges to connect and at most $2K-1$ occupied sockets (by symmetry we can arrange that a and b are the first two check nodes to be allocated). Then we have

$$\mathbb{E}\big[R_\alpha^a R_\beta^b\big] = \left(\frac{1}{L} + O(N^{-\eta})\right)^{2K} = \frac{1}{L^{2K}} + O(N^{-\eta}). \quad (40)$$

By summing over all check nodes, we get $\mathbb{E}X_\alpha = \frac{M}{L^K} + O(N^{1-\eta})$ and after elementary calculations, $\mathrm{Var}X_\alpha = O(N^{2-\eta})$. Since Y_α is binomially distributed, and using $T = M + O(N^{1-\gamma})$, we have

$$\mathbb{E}Y_\alpha = \frac{T}{L^K} = \frac{M}{L^K} + O(N^{1-\gamma}),$$

and also

$$\mathrm{Var}Y_\alpha = T\frac{1}{L^K}\left(1 - \frac{1}{L^K}\right) = O(N),$$

which is much smaller than $\mathrm{Var}X_\alpha$. ∎

To show that the amount of inserted and deleted check nodes is small, we employ now the Chebyshev Inequality, which, for some value of the parameter ζ to be fixed shortly, reads

$$\mathbb{P}\Big[\big|X_\alpha - Y_\alpha - O\left(N^{1-\gamma}\right)\big| \geq N^\zeta O\left(N^{1-\frac{\eta}{2}}\right)\Big] \leq \frac{1}{N^{2\zeta}}.$$

We fix the values $\zeta = \frac{\eta}{4}$ and $\gamma = \frac{\eta}{2}$ (these choices are somewhat arbitrary), and simplifying we obtain

$$\mathbb{P}\Big[|X_\alpha - Y_\alpha| \geq O\left(N^{1-\frac{\eta}{4}}\right)\Big] \leq N^{-\frac{\eta}{2}}.$$

Using the union bound over all L^K possible types, the bound on the probability that the number of insertions and deletions is sublinear in the way depicted above remains $O\left(N^{-\eta/2}\right)$. In case the the number of insertions and deletions is too large, we use the $O(N)$ we use the fact that $\ln Z(G)$ is always $O(N)$ (see Lemma 9). This proves the following lemma.

Lemma 16. *Transmitting over a BMS channel, we have*

$$\mathbb{E}_{h,G:\mathrm{LDPC}(NL,\Lambda,K)} \ln Z(G) \geq \mathbb{E}_{h,G:\{T\times\mathbf{conn}\}} \ln Z(G) + O\left(N^{1-\frac{\eta}{4}}\right).$$

VIII. The large N limit

This section wraps up the proof of Theorem 2. The main ingredient is the content of Lemma 14, which can be written as

$$\begin{aligned}\mathbb{E}_{h,G:\{T\times\mathbf{conn}\}} \ln Z(G) - O\left(N^{1-\gamma}\right) \leq \\ \leq \mathbb{E}_{h,G:\{T\times\mathbf{coup}\}} \ln Z(G) \leq \\ \leq \mathbb{E}_{h,G:\{T\times\mathbf{disc}\}} \ln Z(G) + O\left(N^{1-\gamma}\right). \quad (41)\end{aligned}$$

Using the results from the previous section on the comparison with the simple ensembles and scaling everything by NL, we obtain

$$\begin{aligned}\frac{1}{NL}\mathbb{E}_{h,G:\mathrm{LDPC}(NL,\Lambda,K)} \ln Z(G) - O\left(N^{1-\gamma}\right) \leq \\ \leq \frac{1}{NL}\mathbb{E}_{h,G:\{T\times\mathbf{coup}\}} \ln Z(G) \leq \\ \leq \frac{1}{N}\mathbb{E}_{h,G:\mathrm{LDPC}(N,\Lambda,K)} \ln Z(G) + O\left(N^{1-\gamma}\right). \quad (42)\end{aligned}$$

The next step is to take the $N \to \infty$ limit, and in case it exists for the outer terms, which we are about to show, we can apply the "sandwich rule" to obtain Theorem 2. Note that the ensemble appearing in the middle is what we call $\mathrm{LDPC}(N, L, w, \Lambda, K)$ — we are of course not obliged to pick it as such: we could do another level of processing in the style of the previous section; however the current form is known to fulfill the Maxwell conjecture, so we need not go any further.

To show that the limit

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{h,G:\mathrm{LDPC}(N,\Lambda,K)} \ln Z(G)$$

exists, we use the following result, whose proof can be found in the Appendix of [7].

Lemma 17 (The modified superadditivity theorem)**.** *Given $\alpha \in (0,1)$, suppose a non-negative sequence $\{a_{N,N\geq 1}\}$ satisfies*

$$a_{N_1+N_2} \geq a_{N_1} + a_{N_2} - O((N_1+N_2)^\alpha) \quad (43)$$

for every $N_1, N_2 \geq 1$. Then the limit $\lim_{N\to\infty} \frac{a_N}{N}$ exists (it may be $+\infty$).

The claim then follows by setting the sequence a_N to be the negative of the sequence we study (since $\ln Z(G)$ are negative). It remains to be shown that superadditivity indeed holds.

Since this part is a somewhat simpler variation of the interpolation we have already seen, we only present the proof sketch. We consider a coupled ensemble consisting of only two positions($L = 2$) and interpolate between the cases $w = 1$ (disconnected case) and $w = 2$ (connected case). The novelty is that the number of variables at the first and second positions differ, they are N_1 and N_2, respectively. For the connected case, when edges from check nodes are connected, we do not pick the position at random, but rather weigh the choice by $\nu_1 = \frac{N_1}{N_1+N_2}$ and $\nu_2 = \frac{N_2}{N_1+N_2}$, respectively.

The only difference appears in the reasoning of Lemma 12, where the types are not uniformly distributed anymore. The types are now binary strings of length K, with the two symbols appearing denoting the position, one having weight ν_1, the other ν_2. The weight of the type is the product of the weights of the symbols it contains. If α is a type, let $\nu(\alpha)$ be the weight of that type. Then Equations (30) and (32) become

$$\mathbb{E}_{\alpha:\mathbf{conn}}\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle =$$
$$= \sum_{\alpha\in\{1,2\}^K} \nu(\alpha)\,\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle = \left\langle (\nu_1 Q_1+\nu_2 Q_2)^K\right\rangle,$$
$$\mathbb{E}_{\alpha:\mathbf{disc}}\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle =$$
$$= \sum_{z\in\{1,2\}} \nu_z\,\langle Q_{\alpha_1}\cdots Q_{\alpha_K}\rangle = \left\langle \nu_1 Q_1^K+\nu_2 Q_2^K\right\rangle,$$

and clearly the lemma remains true in this case as well.

IX. Conclusions

The present analysis can be extended with almost no change to arbitrary check-node degree distributions whose generating polynomial $P(x)=\sum_{K\le 0}\rho_K x^K$ is convex for $x\in[-1,1]$. Experimental evidence suggests that even this condition can be relaxed, but new ideas seem to be required for the proofs. A possible route would be to show self-averaging properties for overlap functions, which would allow to use the convexity of $x\mapsto P(x)$ for $x\ge 0$, which holds for any degree distributions (see [9] for a related approach).

The idea of using spatial coupling as a proof technique potentially goes beyond coding theory. We can use it to analyze the free energy of general spin glass models and find exact characterizations or bounds on their phase transition thresholds. We plan to come back to this problem in a forthcoming publication.

Finally, let us also mention that recently, algorithmic lower bounds to thresholds of constraint-satisfaction problems were derived by comparing simple and spatially-coupled constraint-satisfaction models (see [16]).

Acknowledgements

Andrei Giurgiu acknowledges support from the Swiss National Science Foundation grant No. 200020-140388.

Appendix A
Proof of (25)

Proposition 18. *Given a fixed configuration graph G whose underlying type set is m-admissible for $m>K^2$ and a fixed channel realisation h, then with the notation from the proof of Lemma 11 we have that*

$$\frac{1}{|B_\alpha|}\sum_{a\in B_\alpha}\left\langle \sigma_a^{(1)}\cdots\sigma_a^{(r)}\right\rangle =$$
$$= \frac{1}{|B'_\alpha|}\sum_{a\in B'_\alpha}\left\langle \sigma_a^{(1)}\cdots\sigma_a^{(r)}\right\rangle + O\left(\frac{1}{m}\right). \qquad (44)$$

Proof: Rewrite the left hand side as

$$\frac{1}{|B'_\alpha|}\frac{|B'_\alpha|}{|B_\alpha|}\left(\sum_{a\in B'_\alpha}\left\langle \sigma_a^{(1)}\cdots\sigma_a^{(r)}\right\rangle - \sum_{a\in B'_\alpha\setminus B_\alpha}\left\langle \sigma_a^{(1)}\cdots\sigma_a^{(r)}\right\rangle\right). \qquad (45)$$

We will first find an estimate of the quantity $|B'_\alpha\setminus B_\alpha|$, i.e. the number of (pseudo-)check constraints that connect to at least one socket multiple times. To do this, let us look at the subset of B'_α where $a_i=a_j$ (i.e. edges i and j connect to the same socket), for some distinct i,j with $1\le i,j\le K$. The cardinality $q_{i,j}$ of this subset is 0 if $\alpha_i\neq\alpha_j$, and is equal to $|B'_\alpha|/|F_i|\le|B'_\alpha|/m$ if $\alpha_i=\alpha_j$.

A (rough) upper bound for $|B'_\alpha\setminus B_\alpha|$ is given then by sum $\sum_{i\neq j}q_{i,j}$, which in turn never exceeds $K^2|B'_\alpha|/m$.

We are now able to bound the ratio $|B'_\alpha|/|B_\alpha|$ appearing in (45) by $m/(m-K^2)$. Indeed, this follows from

$$\frac{|B'_\alpha|}{|B_\alpha|}=\frac{|B'_\alpha|}{|B'_\alpha|-|B'_\alpha\setminus B_\alpha|}.$$

The absolute value of the second sum in (45) is clearly upper-bounded by $|B'_\alpha\setminus B_\alpha|$, since the bracket takes values between 0 and 1. Putting everything together, we obtain

$$\frac{1}{|B_\alpha|}\sum_{a\in B_\alpha}\left\langle \sigma_a^{(1)}\cdots\sigma_a^{(r)}\right\rangle \le$$
$$\le\left(\frac{m}{m-K^2}\right)\frac{1}{|B'_\alpha|}\sum_{a\in B'_\alpha}\left\langle \sigma_a^{(1)}\cdots\sigma_a^{(r)}\right\rangle + \frac{K^2}{m-K^2},$$
$$\frac{1}{|B_\alpha|}\sum_{a\in B_\alpha}\left\langle \sigma_a^{(1)}\cdots\sigma_a^{(r)}\right\rangle \ge$$
$$\ge \frac{1}{|B'_\alpha|}\sum_{a\in B'_\alpha}\left\langle \sigma_a^{(1)}\cdots\sigma_a^{(r)}\right\rangle - \frac{K^2}{m-K^2}.$$

■

Appendix B
Proof of Theorem 6

We construct a smooth family of channels by interpolating between the given channel c^* and the worst channel, denoted by Δ_0 (since in the log-likelihood representation it consists of a point mass at 0):

$$\mathsf{c}_\mathrm{h}=\frac{\mathrm{h}-\mathrm{h}^*}{1-\mathrm{h}^*}\Delta_0+\frac{1-\mathrm{h}}{1-\mathrm{h}^*}\mathsf{c}^*,$$

where $\mathrm{h}^*=H(\mathsf{c}^*)$ and the parameter h has been chosen in such a way that it coincides with $H(\mathsf{c})$, varying from h^* to 1. Also, to ease notation, for the DE fixpoint we use x_h as a shorthand for $\mathsf{x}_{\mathsf{c}_\mathrm{h}}$.

The plan is as follows: first we will show that

$$\frac{d}{d\mathrm{h}}U(\mathsf{x}_\mathrm{h},\mathsf{c}_\mathrm{h})=g^{\mathrm{BP}}(\mathrm{h}). \qquad (46)$$

Then by Theorem 4, we can replace $g^{\mathrm{BP}}(\mathrm{h})$ with $g^{\mathrm{MAP}}(\mathrm{h})$. By integrating the two sides between h^* and 1 and checking that for the worst channel

$$U(\mathsf{x}_1,\Delta_0)=R=\lim_{N\to\infty}\frac{1}{N}\mathbb{E}H(\underline{X}|\underline{Y}(1)),$$

thereby ending the proof of Theorem 6.

It remains to check (46). Note that an equivalent form of (7) written in the density evolution language is

$$g^{\mathrm{BP}}(\mathrm{h})=\left[\frac{d}{d\mathrm{h}}H(\mathsf{c}_\mathrm{h}\circledast\Lambda^{\circledast}(\rho^{\boxplus}(\mathsf{x}_{\mathrm{h}'})))\right]_{\mathrm{h}'=\mathrm{h}}. \qquad (47)$$

In the ensuing calculations, we will replace x_h by x whenever its meaning is clear from context. It can be easily checked

that this form is very similar to the left hand side of (46), except that the differential operator only affects c and not x (i.e. it is a partial derivative). We will subsequently show that since x is the forward DE fixpoint, the partial derivative equals the total derivative.

We will compute the derivative of each term in (13) separately. The treatment is somewhat similar to the calculation of directional derivatives of the potential function in [14]. Each of the first three terms is of the form

$$\begin{aligned}\frac{d}{d\mathrm{h}}H(f^{\boxcircledast}(\mathsf{x}_\mathrm{h})) &= \lim_{\Delta\mathrm{h}\to 0}\frac{H(f^{\boxcircledast}(\mathsf{x}_{\mathrm{h}+\Delta\mathrm{h}})) - H(f^{\boxcircledast}(\mathsf{x}_\mathrm{h}))}{\Delta\mathrm{h}}\\ &= \lim_{\Delta\mathrm{h}\to 0}\frac{H(f^{\boxcircledast}(\mathsf{x}+\Delta\mathsf{x})) - H(f^{\boxcircledast}(\mathsf{x}))}{\Delta\mathrm{h}},\end{aligned}$$

where $f(u) = \sum_k f_k u^k$ is some polynomial and $\Delta\mathsf{x}$ is a shorthand for $\mathsf{x}_{\mathrm{h}+\Delta\mathrm{h}} - \mathsf{x}_\mathrm{h}$. To keep the formulas uncluttered, in all expressions containing the limit $\Delta\mathsf{x}\to 0$ we suppress the h indices. Expanding, we obtain

$$\begin{aligned}&\frac{d}{d\mathrm{h}}H(f^{\boxcircledast}(\mathsf{x}_\mathrm{h})) = \lim_{\Delta\mathrm{h}\to 0}\frac{H\left(\sum_k f_k \sum_{j\geq 1}\binom{k}{j}\Delta\mathsf{x}^{\boxcircledast j}\boxcircledast \mathsf{x}^{k-j}\right)}{\Delta\mathrm{h}}\\ &= \lim_{\Delta\mathrm{h}\to 0}\frac{H\left(\sum_k k f_k \Delta\mathsf{x}\boxcircledast \mathsf{x}^{k-1}\right)}{\Delta\mathrm{h}} + \lim_{\Delta\mathrm{h}\to 0}\frac{H\left(\Delta\mathsf{x}^{\boxcircledast 2}\boxcircledast g(\mathsf{x},\Delta\mathsf{x})\right)}{\Delta\mathrm{h}}\\ &= \lim_{\Delta\mathrm{h}\to 0}\frac{H(\Delta\mathsf{x}\boxcircledast f'^{\boxcircledast}(\mathsf{x}))}{\Delta\mathrm{h}},\end{aligned}$$

where in the last step all the higher order terms (i.e. those containing a $\circledast$-power of $\Delta\mathsf{x}$ higher than 1 disappear. The polynomial g was introduced just to collect those terms, and the fact that they vanish is shown below in Lemma 24. Explicitly, the derivatives of the first three terms are:

$$\begin{aligned}\frac{d}{d\mathrm{h}}\left[-\frac{\Lambda'(1)}{P'(1)}H(P^{\boxcircledast}(\mathsf{x}_\mathrm{h}))\right] &= -\Lambda'(1)\lim_{\Delta\mathrm{h}\to 0}\frac{H\left(\Delta\mathsf{x}\boxcircledast\rho^{\boxcircledast}(\mathsf{x})\right)}{\Delta\mathrm{h}},\\ \frac{d}{d\mathrm{h}}\left[-\Lambda'(1)H(\rho^{\boxcircledast}(\mathsf{x}_\mathrm{h}))\right] &= -\Lambda'(1)\lim_{\Delta\mathrm{h}\to 0}\frac{H\left(\Delta\mathsf{x}\boxcircledast\rho'^{\boxcircledast}(\mathsf{x})\right)}{\Delta\mathrm{h}},\\ \frac{d}{d\mathrm{h}}\left[\Lambda'(1)H(\mathsf{x}_\mathrm{h}\boxcircledast\rho^{\boxcircledast}(\mathsf{x}_\mathrm{h}))\right] &=\\ = \Lambda'(1)\lim_{\Delta\mathrm{h}\to 0}&\frac{H\left(\Delta\mathsf{x}\boxcircledast\rho^{\boxcircledast}(\mathsf{x})\right) + H\left(\Delta\mathsf{x}\boxcircledast\mathsf{x}\boxcircledast\rho'^{\boxcircledast}(\mathsf{x})\right)}{\Delta\mathrm{h}}.\end{aligned}$$

Using Lemma 22, we replace $H\left(\mathsf{x}\boxcircledast\rho'^{\boxcircledast}(\mathsf{x})\boxcircledast\Delta\mathsf{x}\right)$ with $H\left(\rho'^{\boxcircledast}(\mathsf{x})\boxcircledast\Delta\mathsf{x}\right) - H\left(\mathsf{x}\circledast(\rho'^{\boxcircledast}(\mathsf{x})\boxcircledast\Delta\mathsf{x})\right)$, and we are thus able to cancel the contributions of the first two terms.

The derivative of the last of the four terms in (13) needs to be handled more cautiously, since it contains both kinds of operations on densities. However, the idea remains the same: we examine the quantity

$$H((\mathsf{c}+\Delta\mathsf{c})\circledast\Lambda^{\circledast}(\rho^{\boxcircledast}(\mathsf{x}+\Delta\mathsf{x})) - \mathsf{c}\circledast\Lambda^{\circledast}(\rho^{\boxcircledast}(\mathsf{x})))$$

andi we classify the terms that appear according to the position of $\Delta\mathsf{c}$ and $\Delta\mathsf{x}$. There are two terms that contain once either $\Delta\mathsf{c}$ and $\Delta\mathsf{x}$:

- $\Delta\mathsf{c}\circledast\Lambda^{\circledast}(\rho^{\boxcircledast}(\mathsf{x}))$,
- $\mathsf{c}\circledast\Lambda'^{\circledast}(\rho^{\boxcircledast}(\mathsf{x}))\circledast(\rho'^{\boxcircledast}(\mathsf{x})\boxcircledast\Delta\mathsf{x})$.

The higher order terms (the ones that contain at least two of $\Delta\mathsf{x}$ and $\Delta\mathsf{c}$) are of the types

- $(\Delta\mathsf{x}\boxcircledast\Delta\mathsf{x}\boxcircledast g_1(\mathsf{x},\Delta\mathsf{x}))\circledast g_2(\mathsf{x},\Delta\mathsf{x},\mathsf{c})$,
- $(\Delta\mathsf{x}\boxcircledast g_1(\mathsf{x}))\circledast(\Delta\mathsf{x}\boxcircledast g_2(\mathsf{x}))\circledast g_2(\mathsf{x},\Delta\mathsf{x},\mathsf{c})$,
- $(\Delta\mathsf{x}\boxcircledast g_1(\mathsf{x},\Delta\mathsf{x}))\circledast g_2(\mathsf{x},\Delta\mathsf{x})\circledast\Delta\mathsf{c}$,

where the functions g_1, g_2, g_3 are products involving $\circledast$ and $\boxcircledast$ of their parameters. All the terms above have vanishing contributions in the limit, by Lemma 24.

We are now able to collect all the terms that remain and assemble them in the form

$$\begin{aligned}&\frac{d}{d\mathrm{h}}U(\mathsf{x}_\mathrm{h},\mathsf{c}_\mathrm{h}) =\\ &\lim_{\Delta\mathrm{h}\to 0}\frac{H((\mathsf{x}-\mathsf{c}\circledast\Lambda'^{\circledast}(\rho^{\boxcircledast}(\mathsf{x})))\circledast(\rho'^{\boxcircledast}(\mathsf{x})\boxcircledast\Delta\mathsf{x}))}{\Delta\mathrm{h}}\\ &\quad+\lim_{\Delta\mathrm{h}\to 0}\frac{H(\Delta\mathsf{c}\circledast\Lambda^{\circledast}(\rho^{\boxcircledast}(\mathsf{x})))}{\Delta\mathrm{h}}\\ &= 0 + g^{\mathrm{BP}}(\mathrm{h}),\end{aligned}$$

where in the last step we used the fact that x is the fixpoint of the DE equation, and also the alternative definition of the BP GEXIT curve provided by (47).

The proof is now complete, and we are left to show that the higher order terms do not contribute in the limit.We begin with some definitions and some new notations. Degradation induces a partial ordering on $\mathcal{X}$, which we denote by $z\prec z'$, where z' is degraded with respect to z. Note that density evolution preserves degradation, and the following proposition follows from standard arguments in [3].

Proposition 19. *If* $\mathsf{c},\mathsf{c}'\in\mathcal{X}$ *and* $\mathsf{c}\prec\mathsf{c}'$ *then* $\mathsf{x}_\mathsf{c}\prec\mathsf{x}_{\mathsf{c}'}$.

For any $\mathsf{z}\in\mathcal{X}$, the Bhattacharyya functional [3] is given by

$$\mathcal{B}(\mathsf{z}) = \int \mathsf{z}(h)e^{-h}d\mathsf{z}(h). \tag{48}$$

There is a metric defined on $\mathcal{X}$, the *Wasserstein distance (on the $|D|$ domain)* [2], that has the following useful properties which we state here without proof. For any $\mathsf{z},\mathsf{z}'\mathsf{y}\in\mathcal{X}$,

$$\begin{aligned}d(\mathsf{z}\circledast\mathsf{y},\mathsf{z}'\circledast\mathsf{y}) &\leq 2d(\mathsf{z},\mathsf{z}'),\\ d(\mathsf{z}\boxcircledast\mathsf{y},\mathsf{z}'\boxcircledast\mathsf{y}) &\leq d(\mathsf{z},\mathsf{z}').\end{aligned}$$

Let $\mathcal{F}$ be the set of functions $f:\mathcal{X}\to\mathcal{X}$ of the form

$$f(\mathsf{z}) = \mathsf{y}_1 *_1 (\mathsf{y}_2 *_2 (\ldots(\mathsf{y}_k *_k \mathsf{z})))$$

for some $\mathsf{y}_1,\ldots,\mathsf{y}_k\in\mathcal{X}$ and $*_1,\ldots,*_k\in\{\circledast,\boxcircledast\}$. We can easily extend f by linearity, in order to define quantities like $f(\mathsf{z}-\mathsf{z}')$. Then for each $f\in\mathcal{F}$ there is a constant M such that for all $\mathsf{z}\prec\mathsf{z}'$ we have that

$$(f(\mathsf{z}),f(\mathsf{z}'))\leq Md(\mathsf{z},\mathsf{z}'). \tag{49}$$

If $\mathsf{z}\prec\mathsf{z}'$, the Wasserstein distance is bounded above and below by powers of the Bhattacharyya functional, in the sense that

$$\frac{1}{4}(\mathcal{B}(\mathsf{z}')-\mathcal{B}(\mathsf{z}))^2\leq d(\mathsf{z},\mathsf{z}')\leq 2\sqrt{\mathcal{B}(\mathsf{z}')-\mathcal{B}(\mathsf{z})}.$$

The following lemma (part of Lemma 21 in [2]) will enable us to factorize the entropy of a $\circledast$-product. The reason why we consider the Bhattacharyya functional is contained in the following lemmas.

Lemma 20. *Let* $\mathsf{z}, \mathsf{z}', \mathsf{y}, \mathsf{y}' \in \mathcal{X}$ *such that* $\mathsf{z} \succ \mathsf{z}'$. *Then*

$$|H((\mathsf{z}-\mathsf{z}') \circledast (\mathsf{y}-\mathsf{y}')| \leq \frac{8}{\log 2} \mathcal{B}(\mathsf{z}-\mathsf{z}') \sqrt{2d(\mathsf{y},\mathsf{y}')}.$$

We are now ready to tackle the higher order contributions. Let $M_1, M_2, \ldots$ denote constants independent of the channel.

Proposition 21. *With the notation from the beginning of this section, for any* $f \in \mathcal{F}$ *(extended by linearity), we have*

$$\lim_{\Delta \mathrm{h} \to 0} \frac{H(\Delta \mathsf{x} \circledast f(\Delta \mathsf{x}))}{\Delta \mathrm{h}} = 0,$$
$$\lim_{\Delta \mathrm{h} \to 0} \frac{H(\Delta \mathsf{c} \circledast f(\Delta \mathsf{x}))}{\Delta \mathrm{h}} = 0.$$

Proof: We concentrate on the first limit, as the second is similar but easier. Applying Lemma 20 we obtain the upper bound

$$\lim_{\Delta \mathrm{h} \to 0} M_1 \frac{\mathcal{B}(\Delta \mathsf{x}) \sqrt{2d(f(\mathsf{x}), f(\mathsf{x}+\Delta \mathsf{x}))}}{H(\Delta \mathsf{c})}.$$

Since the parametrization is just a linear interpolation between c^* and Δ_0 and $H(\cdot)$ and $\mathcal{B}(\cdot)$ are linear functionals, we have that $H(\Delta \mathsf{c}) = M_2 \mathcal{B}(\Delta \mathsf{c})$. Then we can replace the denominator by the Bhattacharyya quantity and use the regularity condition (14). The only thing left to be shown is that $\sqrt{2d(f(\mathsf{x}), f(\mathsf{x}+\Delta \mathsf{x}))} \to 0$. This follows from inequality (49) and the fact that d is a metric. ∎

The main tool to turn $\circledcirc$ into $\circledast$ and vice-versa is the following.

Lemma 22 (Duality lemma, [3])**.** *Let* $\mathsf{z}, \mathsf{z}', \mathsf{y}, \mathsf{y}' \in \mathcal{X}$. *Then*

$$H(\mathsf{z} \circledast \mathsf{y}) + H(\mathsf{z} \circledcirc \mathsf{y}) = H(\mathsf{z}) + H(\mathsf{y}).$$

For differences of densities, because of linearity of H*, this takes the forms*

$$H((\mathsf{z}-\mathsf{z}') \circledast \mathsf{y}) + H((\mathsf{z}-\mathsf{z}') \circledcirc \mathsf{y}) = H(\mathsf{z}-\mathsf{z}'), \quad (50)$$
$$H((\mathsf{z}-\mathsf{z}') \circledast (\mathsf{y}-\mathsf{y}')) + H((\mathsf{z}-\mathsf{z}') \circledcirc (\mathsf{y}-\mathsf{y}')) = 0. \quad (51)$$

Proposition 21 with the identity map as f and (51) implies

$$\lim_{\Delta \mathrm{h} \to 0} \frac{|H(\Delta \mathsf{x} \circledcirc \Delta \mathsf{x})|}{\Delta \mathrm{h}} = 0. \quad (52)$$

Proposition 23 (Proposition 6 in [14])**.** *If* z *is any symmetric measure (not necessarily signed), then*

$$H(\mathsf{z}) = \mathsf{z}(\overline{\mathbb{R}}) - \sum_{k=1}^{\infty} \frac{(\log 2)^{-1}}{2k(2k-1)} M_k(\mathsf{z}),$$

where $M_k(\mathsf{z}) = \int (\tanh h)^{2k} d\mathsf{z}(h)$ *and* $\mathsf{z}(\overline{\mathbb{R}})$ *is the total mass of* z.

Moreover, for any symmetric measures z_1 *and* z_2,

$$M_k(\mathsf{z}_1 \circledcirc \mathsf{z}_2) = M_k(\mathsf{z}_1) M_k(\mathsf{z}_2).$$

Since the quantities $M_k(\Delta \mathsf{x} \circledcirc \Delta \mathsf{x}) = M_k(\Delta \mathsf{x})^2$ are all positive, the previous proposition implies that

$$|H(\Delta \mathsf{x} \circledcirc \Delta \mathsf{x} \circledcirc \mathsf{y})| \leq |H(\Delta \mathsf{x} \circledcirc \Delta \mathsf{x})|, \quad (53)$$

for all $\mathsf{y} \in \mathcal{X}$. By an application of (50), one also obtains

$$|H((\Delta \mathsf{x} \circledcirc \Delta \mathsf{x} \circledcirc \mathsf{y}_1) \circledast \mathsf{y}_2)| \leq 2|H(\Delta \mathsf{x} \circledast \Delta \mathsf{x})|. \quad (54)$$

We are finally ready to state the result proving the vanishing contribution of higher order terms:

Lemma 24. *We have*

$$\lim_{\Delta \mathrm{h} \to 0} \frac{H((\Delta \mathsf{x} \circledcirc \Delta \mathsf{x} \circledcirc g_1(\mathsf{x}, \Delta \mathsf{x})) \circledast g_2(\mathsf{x}, \Delta \mathsf{x}, \mathsf{c}, \Delta \mathsf{c}))}{\Delta \mathrm{h}} = 0, \quad (55)$$

$$\lim_{\Delta \mathrm{h} \to 0} \frac{H((\Delta \mathsf{x} \circledcirc g_1(\mathsf{x})) \circledast (\Delta \mathsf{x} \circledcirc g_2(\mathsf{x})) \circledast g_3(\mathsf{x}, \Delta \mathsf{x}, \mathsf{c}))}{\Delta \mathrm{h}} = 0, \quad (56)$$

$$\lim_{\Delta \mathrm{h} \to 0} \frac{H(\Delta \mathsf{c} \circledast (\Delta \mathsf{x} \circledcirc g_2(\mathsf{x})) \circledast g_3(\mathsf{x}, \Delta \mathsf{x}, \mathsf{c}))}{\Delta \mathrm{h}} = 0. \quad (57)$$

Proof: The limit (55) is a direct consequence of (54). The third one, (57), is a consequence of Proposition 21. The second one can also be reduced to the form appearing in Proposition 21 by using the Duality Lemma twice:

$$\begin{aligned} &H((\Delta \mathsf{x} \circledcirc g_1(\mathsf{x})) \circledast (\Delta \mathsf{x} \circledcirc g_2(\mathsf{x})) \circledast g_3(\mathsf{x}, \Delta \mathsf{x}, \mathsf{c})) \\ &= H(\Delta \mathsf{x} \circledcirc g_1(\mathsf{x}) \circledcirc ((\Delta \mathsf{x} \circledcirc g_2(\mathsf{x})) \circledast g_3(\mathsf{x}, \Delta \mathsf{x}, \mathsf{c}))) \\ &= H(\Delta \mathsf{x} \circledast (g_1(\mathsf{x}) \circledcirc ((\Delta \mathsf{x} \circledcirc g_2(\mathsf{x})) \circledast g_3(\mathsf{x}, \Delta \mathsf{x}, \mathsf{c})))). \end{aligned}$$

∎

REFERENCES

[1] A. J. Felström and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," vol. 45, pp. 2181–2190, Sept. 1999.

[2] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation." E-print arXiv:1201.2999, Jan. 2012.

[3] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, Mar. 2008.

[4] F. Guerra and F. L. Toninelli, "The high temperature region of the Viana-Bray diluted spin glass model," *J. Stat. Phys.*, vol. 115, pp. 501–555, Apr. 2004.

[5] S. Franz and M. Leone, "Replica bounds for optimization problems and diluted spin systems," *J. Stat. Phys.*, vol. 111, pp. 535–564, 2003.

[6] S. Franz, M. Leone, and F. L. Toninelli, "Replica bounds for diluted non-Poissonian spin systems," *J. Phys. A: Math. Gen.*, vol. 36, pp. 10967—-10985, 2003.

[7] M. Bayati, D. Gamarnik, and P. Tetali, "Combinatorial approach to the interpolation method and scaling limits in sparse random graphs," in *Proceedings of the 42nd ACM Symp. on Theory of Comp.*, STOC '10, (New York, NY, USA), pp. 105–114, ACM, June 2010.

[8] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3221–3246, 2005.

[9] S. Kudekar and N. Macris, "Sharp bounds for optimal decoding of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 55, pp. 4635–4650, Oct. 2009.

[10] A. Giurgiu, N. Macris, and R. Urbanke, "How to prove the Maxwell conjecture via spatial coupling – a proof of concept," in *Proceedings of the IEEE Int. Symp. Inf. Theory (ISIT) 2012*, pp. 458–462, 2012.

[11] A. Giurgiu, N. Macris, and R. Urbanke, "And now to something completely different: spatial coupling as a proof technique," in *Proceedings of the IEEE Int. Symp. Inf. Theory (ISIT) 2012*, 2013.

[12] C. Méasson, A. Montanari, T. J. Richardson, and R. Urbanke, "The generalized area theorem and some of its consequences," *IEEE Trans. Inf. Theor.*, vol. 55, pp. 4793–4821, Nov. 2009.

[13] C. Méasson, *Conservation laws for coding, PhD Thesis*. EPFL, 2006.
[14] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "A proof of threshold saturation for spatially-coupled LDPC codes on BMS channels," *Proc. Annual Allerton Conf. on Commun., Control and Comp.*, 2012.
[15] N. Macris, "Griffith-Kelly-Sherman correlation inequalities: a useful tool in the theory of error correcting codes," *IEEE Trans. Inform. Theory*, vol. 53, pp. 664–683, Feb. 2007.
[16] S. Hamed Hassani, N. Macris, and R. Urbanke, "Threshold saturation in spatially coupled constraint satisfaction problems," *Journal of Statistical Physics*, pp. 1–44, 2012.