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 Poisson-distributed 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 [3] 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 much 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 [4] (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 so-called Nishimori symmetry (of which channel symmetry is a special case).

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 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 occurrence of nodes of very small degrees in significant proportions, which limits the performance. In what follows, we remove this technical barrier and allow 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.

II. PRELIMINARIES

A. Simple ensembles

We start by describing a simple ensemble of codes, which we call LDPC(N, Λ, K), where N is the number of variable nodes, Λ(x) = \sum_{d=0}^D \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 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 \( \lfloor D/K \rfloor \), where \( D \) is the total number of sockets. 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 \( \Lambda \).

**B. Coupled ensembles**

Intuitively, a coupled ensemble \( \text{LDPC}(N, L, w, \Lambda, K) \) 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 \( \{1, \ldots, L\} \). 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 \( \Lambda \). 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 \( z, z+1, \ldots, z+w-1 \). As before, check nodes have degree \( K \), and they are sampled as follows: first a window is picked uniformly at random, then for each edge, a position uniformly and i.i.d. inside that window, and then 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 \( d \) and \( K \). The steps in this process will be described in more detail in Section \( \nabla \).

Note that the ensembles described so far are built in two stages: first the vertex degrees are sampled from the distribution \( \Lambda \) and sockets are attached, obtaining the configuration pattern; in the second stage, the check nodes and edges are connected within this configuration pattern. Both stages are random, except when the degree distribution \( \Lambda \) is regular, in which case the first stage is deterministic. It will be sometimes helpful to separate the two stages and start 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 \( \text{LDPC}(N, \Lambda, K) \), whereas for \( w = L \), the whole ensemble is equivalent to \( \text{LDPC}(NL, \Lambda, K) \), up to a small number of missing check nodes.

**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 \( a = (a_1, \ldots, a_K) \) of variable nodes, thereby incorporating the edge information in the graph. A code is then specified by the total number 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 \( a = (a_1, \ldots, a_K) \). In general we will use the letters \( u, v \) to describe check constraints, \( a, b, c, \ldots \) to describe variable nodes, and \( G, \tilde{G}, G^\prime, \ldots \) 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

\[
h(y) = \frac{1}{2} \ln \frac{p_{Y|X}(y \mid 1)}{p_{Y|X}(y \mid -1)},
\]

and 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 \( \sigma \in \mathcal{X}^V \), where \( V \) is the set of variable nodes. Instead of the outputs, we use the HLLRs \( h \in \mathbb{R}^V \), given by \( h_v = h(y_v) \), where \( y \) is the output vector.

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

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

(1)

where \( \sigma_a \) is short for the product \( \sigma_{a_1} \cdots \sigma_{a_K} \), and \( Z(G) \) is a normalizing factor, also called partition function, given by

\[
Z(G) = \sum_{\sigma \in \mathcal{X}^V} e^{h \cdot \sigma} \prod_{a \in G} (1 + \sigma_a)/2.
\]

One can easily check that the product \( \prod_{a \in G} (1 + \sigma_a)/2 \) is 1 when \( \sigma \) is any codeword, and 0 otherwise. Also, we have denoted this probability measure by \( \mu \) in order to distinguish it from other randomized parameters that appear, notably the channel and the randomness in the graph \( G \). Note that \( \mu \) depends on both \( G \) and the HLLRs \( h \), and when this is not clear we will make it explicit by adding \( G \) or \( h \) as a subscript.
Consider a smooth family of BMS channels ordered by degradation, indexed by a noise parameter. Without loss of generality, we may assume that the parameter $\epsilon$ is the channel entropy $H(Y_i|X_i)$, which varies between 0 (the perfect channel) and 1 (the useless channel). Then there exists a value $\epsilon_{\text{MAP}}$ (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 2) 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:

\[
\epsilon_{\text{MAP}} = \inf \left\{ \epsilon : \liminf_{N \to \infty} \frac{1}{N} \mathbb{E}_{G,\text{LDPC}(N,L)} H(X|Y) > 0 \right\},
\]

\[
\epsilon_{w}^{L,w} = \inf \left\{ \epsilon : \liminf_{N \to \infty} \frac{1}{NL} \mathbb{E}_{G,\text{LDPC}(N,L,w)} H(X|Y) > 0 \right\}.
\]

These definitions employ lim inf 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 and we could have replaced lim inf by lim. The theorem obviously implies the equality of the two MAP thresholds.

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

**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 $\epsilon_{\text{Area}}$ be the area threshold defined as that value so that the integral of the BP-GEXIT curve over the interval $[\epsilon_{\text{Area}}, 1]$ equals the design rate $1-d/K$ (for more details, see [3]). The Maxwell conjecture states that $\epsilon_{\text{Area}} = \epsilon_{\text{MAP}}$.

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 $\epsilon_{\text{Area}}$ (of the simple ensemble) in the following sense:

\[
\epsilon_{\text{Area}} - O\left( \frac{1}{w^{1/2}} \right) \leq \epsilon_{w}^{L,w,\text{open}} \leq \epsilon_{\text{Area}} + O\left( \frac{w}{L} \right),
\]

(3)

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

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 5 the difference of the entropies is at most $O(w/L)$, which goes to 0 as $L \to \infty$. As a consequence,

\[
\lim_{L \to \infty} \epsilon_{w}^{L,w,\text{open}} = \lim_{L \to \infty} \epsilon_{w}^{L,w,\text{MAP}}.
\]
Thus by (3) and Corollary [3] we deduce that in fact $\epsilon_{\text{MAP}}$ equals $\epsilon_{\text{Area}}$, 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 (3) 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 LDPC$(N, \Lambda, K)$. Also, in order to make clear that the channel output depends on the channel entropy parameter $\epsilon$, we will write the former as $Y(\epsilon)$. The MAP-GEXIT function $g_{\text{MAP}}$ is defined as

$$g_{\text{MAP}}(\epsilon) = \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_G \left[ \sum_v H(X_v | Y_{\sim v}(\epsilon)) \right],$$

where $\sim v$ represents the set of all nodes except $v$. Equivalently [12, Section III], the MAP-GEXIT function also takes the form

$$g_{\text{MAP}}(\epsilon) = \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_G \left[ \frac{\partial}{\partial \epsilon} H(X | Y(\epsilon)) \right].$$

The latter formulation can then be employed to lower bound the area below $g_{\text{MAP}}$ above the MAP threshold as follows:

$$\int_{g_{\text{MAP}}} \frac{1}{N} \mathbb{E}_G \left[ \frac{\partial}{\partial \epsilon} H(X | Y(\epsilon)) \right] d\epsilon = \int_{g_{\text{MAP}}} \left( \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_G \left[ \frac{\partial}{\partial \epsilon} H(X | Y(\epsilon)) \right] \right) d\epsilon$$

$$\geq \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_G \left[ \frac{\partial}{\partial \epsilon} H(X | Y(\epsilon)) \right]$$

$$= \limsup_{N \to \infty} \left( \frac{1}{N} \mathbb{E}_G H(X | Y(\epsilon)) - \frac{1}{N} \mathbb{E}_G H(X | Y(\epsilon_{\text{MAP}})) \right)$$

$$\geq R - \liminf_{N \to \infty} \frac{1}{N} \mathbb{E}_G H(X | Y(\epsilon_{\text{MAP}}))$$

$$\geq R,$$

where in step (a) we used the Reverse Fatou Lemma (note that the integrand on the r.h.s. is bounded, see for example [4]), and in step (b), since at $\epsilon = 1$ the channel is completely useless, we have that $H(X | Y(1)) = H(X)$ and $\frac{\partial}{\partial \epsilon} H(X | Y(\epsilon))$ is the rate of the code in the large blocklength limit. For step (c), by the definition of the MAP threshold for any $\epsilon < \epsilon_{\text{MAP}}$, we have $\lim\inf_{N \to \infty} \frac{1}{N} \mathbb{E}_G H(X | Y(\epsilon_{\text{MAP}})) = 0$, and we will now argue for left-continuity at $\epsilon_{\text{MAP}}$. The conditional entropy term can be expanded using

$$H(X | Y(\epsilon)) = H(X) + H(Y(\epsilon) | X) - H(Y(\epsilon)).$$

In the large blocklength limit, the first two terms on the r.h.s. are related to the rate and the one-bit channel entropy, i.e.

$$\lim_{N \to \infty} \frac{1}{N} \mathbb{E}_G [H(X) + H(Y(\epsilon) | X)] = R + \epsilon$$

for any entropy parameter $\epsilon \in [0, 1]$. Then by taking limits in (7), we get

$$\liminf_{N \to \infty} \frac{\mathbb{E}_G H(X | Y(\epsilon))}{N} = R + \epsilon - \limsup_{N \to \infty} \frac{\mathbb{E}_G H(Y(\epsilon))}{N}.$$ 

We now show that the l.h.s. as a function of $\epsilon$ is left-continuous at $\epsilon_{\text{MAP}}$. Suppose it is not; then clearly it must have a positive jump at $\epsilon_{\text{MAP}}$. According to (3), there is a negative jump in $\limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_G H(Y(\epsilon))$. Because $H(Y(\epsilon))$ is an increasing function of $\epsilon$ for any graph, the contradiction becomes evident. Thus, we obtain that $\lim\inf_{N \to \infty} \frac{1}{N} \mathbb{E}_G H(X | Y(\epsilon_{\text{MAP}})) = 0$ and we conclude the discussion of inequality (6).

We now define the BP-GEXIT curve in a similar way, except that now in the calculation of the “extrinsic” entropy of bit $X_v$ we only consider the information in the computation tree of depth $\ell$ rather than all of $Y_{\sim v}$. Denoting the subset of outputs occurring in the computation tree of length $\ell$ by $Y_{\sim v}$ (see [12, Section III]), we can formalize the previous statement as

$$g_{\text{BP}}(\epsilon) = \lim_{\ell \to \infty} \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_G \left[ \sum_v H(X_v | Y_{\sim v}^\ell(\epsilon)) \right].$$

We can easily see then by the data processing inequality that (see also Lemma 9 in [12])

$$g_{\text{MAP}}(\epsilon) \leq g_{\text{BP}}(\epsilon), \text{ for all } \epsilon \in [0, 1].$$

The area threshold mentioned before is defined as the solution $\epsilon_{\text{area}}$ to the equation

$$\int_{g_{\text{area}}} \frac{1}{N} \mathbb{E}_G \left[ \frac{\partial}{\partial \epsilon} H(X | Y(\epsilon)) \right] d\epsilon = R.$$ 

Using then the equality of the MAP and area thresholds established in the previous subsection for the above-mentioned class of LDPC codes and using (6) and (11) we obtain

$$\int_{g_{\text{area}}} \left( g_{\text{BP}}(\epsilon) - g_{\text{MAP}}(\epsilon) \right) d\epsilon \leq R - R = 0.$$ 

The positivity of the integrand (cf. (10)) entails the equality of the two curves almost everywhere above the MAP threshold.

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 [3], [13], 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.

Lemma 4 (Nishimori Identity). Fix a graph $G$ (it can be general, no constraints on the check node degrees needed here). For any odd positive integer $m$ we have

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

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 $\sigma_b = \sigma_{b_1} \cdots \sigma_{b_J}$. The channel used for transmission needs to be BMS, symmetry being the crucial ingredient.
thus we can safely take odd, (equation (1)) to replace terms in the sum are equal, so the expression simplifies to the l.h.s. of (13) can be written as

$$\mathbb{E}_h[\langle \sigma_b \rangle^m] = \int \langle \sigma_b \rangle^m \prod_{v \in V} e^{h_v f(|h_v|)} dh_v.$$ (14)

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

We now perform an average over all codewords $\tau$, obtaining

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

where $C(G)$ is the set of all codewords.

Note that the Gibbs bracket above averages over $\sigma$, and thus we can safely take $\tau_v$ out of the bracket. Since $m$ is odd, $\tau_v^2 = \tau_v$. Next we use the definition of Gibbs measure (equation (1)) to replace $\sum_{\tau \in C(G)} e^{h_v \tau_v} \tau_v$ with $Z(G) \langle h \rangle_{\tau}$. We obtain

$$\mathbb{E}_h[\langle \sigma_b \rangle^m] = \frac{1}{|C(G)|} \int Z(G) \langle h \rangle^{m+1} \prod_{v \in V} f(|h_v|) dh_v.$$ (15)

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

$$\mathbb{E}_h[\langle \sigma_b \rangle^m] = \frac{1}{|C(G)|} \sum_{\lambda \in C(G)} \int \langle \sigma_b \rangle^{m+1} \prod_{v \in V} e^{h_v \lambda_v} f(|h_v|) dh_v.$$ (16)

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

$$\mathbb{E}_h[\langle \sigma_b \rangle^m] = \int \langle \sigma_b \rangle^{m+1} \prod_{v \in V} e^{h_v f(|h_v|)} dh_v.$$ (17)

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.

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

$$Z(G \cup b) = \sum_{\sigma \in X^V} e^{h \sigma} \prod_{a \in G} \frac{1 + \sigma_a}{2} = Z(G) \langle \frac{1 + \sigma_b}{2} \rangle_G.$$ (18)

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

$$\ln(1 + \langle \sigma_b \rangle) = \sum_{j \geq 1} \frac{(-1)^{j+1}}{j} \langle \sigma_b \rangle^j.$$ (19)

We now use the Nishimori Identities (Lemma 4) with $\mathbb{E}_h[\langle \sigma_b \rangle^{m-1}] = \mathbb{E}_h[\langle \sigma_b \rangle^j]$, 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 $\langle \sigma_b \rangle^j$ 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 $\sigma^{(j)} \in X^V$. Because the product measure is just the measure of $r$ independent copies of the measure (henceforth called replicas), it is easy to check that

$$\langle \sigma_b \rangle^r_G = \langle \sigma_b^{(1)} \cdots \sigma_b^{(r)} \rangle^r_G.$$

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 6. Given any graph $G$ and an additional check constraint $b$, we have that

$$\mathbb{E}_h[\ln Z(G \cup b) - \ln Z(G)] = -\ln 2 + \mathbb{E}_h \sum_{r \in \mathbb{Z}_+} \frac{\langle \sigma_b^{(1)} \cdots \sigma_b^{(r)} \rangle}{r^2 - r}.$$ (20)

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 1 is already fixed, i.e., it has been properly sampled at an earlier stage, and there are at least $N \tilde{d}(1 - N^{-\eta})$ and at most $N \tilde{d}(1 + N^{-\eta})$ 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 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 \( \text{var}(s) \) denote the variable node that it is part of; by \( \sigma_v \) we understand \( \sigma_{\text{var}(s)} \). Let \( \text{pos}(v) \) denote the position of the variable \( v \), with the notation extending to sockets in the obvious manner: \( \text{pos}(s) = \text{pos}(\text{var}(s)) \). We also set \( S_z = \{ s \in S : \text{pos}(s) = z \} \), 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 \( \alpha_j = \text{pos}(a_j) \), 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 \( \text{conn} \).

- **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 \( \text{disc} \).

- **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 \( \text{coup} \).

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

Given a multiset of types \( \Gamma \) (a set of types where duplicates can appear), we extend the definition of the positional occupation vector to \( \text{occ}_{\Gamma} = \sum_{\alpha \in \Gamma} \text{occ}_\alpha \). We call a multiset of types \( m \)-admissible if \( \text{occ}_{\Gamma}(z) \leq |S_z| - m \), for all positions \( z \). In other words, an \( m \)-admissible set of types \( \Gamma \) ensures that there exists a graph \( G \) whose check constraints match one-to-one the types in \( \Gamma \) (we say that \( G \) is compatible with \( \Gamma \)), 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 \( \Gamma \) is simply given by the uniform measure over all graphs that are compatible with \( \Gamma \). 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 \( \Gamma \) (the order is immaterial), pick distinct \( a_i \) uniformly at random from the free sockets at position \( \alpha_i \), 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 type \( \alpha \) and the part \( G \) of the graph that is already in place. Let \( B_\alpha \) be the set of check constraints that are compatible with \( \alpha \) 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 \( B_\alpha \).

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 \( F_z \subset S_z \) 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

\[
Q_z(\sigma^{(1)}, \ldots, \sigma^{(r)}) = \frac{1}{|F_z|} \sum_{s \in F_z} \sigma_s^{(1)} \cdots \sigma_s^{(r)}. \tag{19}
\]

The next statement describes the link between the overlap functions and the replica averages introduced by Lemma 5.

**Lemma 7.** Given a number \( m > K^2 \), a fixed channel realization, a fixed graph \( G \) whose associated type set is \( m \)-admissible and fixed type \( \alpha \), we have

\[
\mathbb{E}_{\nu(\alpha, G)} \left\langle \sigma^{(1)}_\alpha \cdots \sigma^{(r)}_\alpha \right\rangle_G = \left( \prod_{j=1}^{K} Q_{\alpha_j}(\sigma^{(1)}_\alpha, \ldots, \sigma^{(r)}_\alpha) \right) + O \left( \frac{1}{m} \right). \tag{20}
\]

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

\[
\frac{1}{|B_\alpha|} \sum_{a \in B_\alpha} \left\langle \sigma^{(1)}_a \cdots \sigma^{(r)}_a \right\rangle.
\]

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 \( B'_\alpha \) be the set of all (pseudo-)check constraints that are compatible with \( \alpha \), and where sockets are allowed to appear multiple times. Then \( B'_\alpha \) can be written as a product:

\[
B'_\alpha = F_{\alpha_1} \times \cdots \times F_{\alpha_K}.
\]
where the set $F_z$ is the set of free sockets at position $z$. The idea is now that we can replace $B_a$ with $B'_a$ in the average 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

$$
\frac{1}{|B_a|} \sum_{a \in B_a} \langle \sigma_a(1) \ldots \sigma_a(r) \rangle = \frac{1}{|B'_a|} \sum_{a \in B'_a} \langle \sigma_a(1) \ldots \sigma_a(r) \rangle + O\left( \frac{1}{m} \right). \tag{22}
$$

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 $B'_a$ in the following way:

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

Taking the bracket outside and factorizing, we obtain

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

which we can identify as the bracketed product of positional overlap functions on the right hand side of (20).

Lemma 8. Let $G$ be a graph whose type multiset is admissible, and fix the channel realization $h$. Then the following inequalities hold:

$$
\mathbb{E}_{a:z(a,G)} \langle \sigma_a(1) \ldots \sigma_a(r) \rangle_G \geq \mathbb{E}_{a:z(a,G)} \langle \sigma_a(1) \ldots \sigma_a(r) \rangle_{G^T} + O(1/m), \tag{23}
$$

$$
\mathbb{E}_{a:z(a,G)} \langle \sigma_a(1) \ldots \sigma_a(r) \rangle_G \geq \mathbb{E}_{a:z(v(a,G))} \langle \sigma_a(1) \ldots \sigma_a(r) \rangle_{G^T} + O(1/m). \tag{24}
$$

Proof: The claim follows by Lemma 7 if we manage to show the following two inequalities:

$$
\mathbb{E}_{a:z(a,G)} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle \geq \mathbb{E}_{a:z(a,G)} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle, \tag{25}
$$

$$
\mathbb{E}_{a:z(a,G)} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle \geq \mathbb{E}_{a:z(a,G)} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle, \tag{26}
$$

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:

$$
\mathbb{E}_{a:z(a,G)} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle = \frac{1}{L} \sum_{z \in [L]} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle = \left( \frac{1}{L} \sum_{z \in [L]} Q_z \right)^K, \tag{27}
$$

$$
\mathbb{E}_{a:z(a,G)} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle = \frac{1}{L} \sum_{z \in [L]} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle = \left( \frac{1}{L} \sum_{z \in [L]} Q_z \right)^K, \tag{28}
$$

$$
\mathbb{E}_{a:z(a,G)} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle = \frac{1}{L} \sum_{z \in [L]} \langle Q_{\alpha_1} \ldots Q_{\alpha_K} \rangle = \left( \frac{1}{L} \sum_{z \in [L]} Q_z \right)^K. \tag{29}
$$

Both inequalities (28) and (26) 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 $G : \{ t_1 \times z \} \rightarrow \{ \{ t_2 \times \text{disc} \} \}$ 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, $\text{conn}$ 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) $NLD/K$. However, there is a fluctuation ($\pm N^{1-\eta/2}$ 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 $T = NLD(1 - N^{-\gamma})/K$, so in case the graph is admissible there will be $O(N^{1-\gamma})$ free sockets left at each position. The exponent $\gamma$ 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 9.** Let $\alpha^1, \ldots, \alpha^T$ be random types, each drawn from a distribution that is either $\text{conn}$, $\text{disc}$ or $\text{coup}$ (could be different for each type). Then with high probability (more precisely $1 - O(\exp(-\kappa n^{1-2\gamma}))$, for some positive constant $\kappa$) the resulting multiset of types is $dN^{1-\gamma}/2$-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 $TK/L + dN^{1-\gamma}/2$ with a very small probability. Next, by the union bound over all positions $z$, we upper bound the probability that the graph is not $dN^{1-\gamma}/2$-admissible and the lemma is proved.

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

$$\Pr \left[ \sum_{t=1}^{T} X_t \geq TK/L + \frac{1}{2}dN^{1-\gamma} \right] \leq \exp \left( -\frac{d^2N^{2-2\gamma}}{2K^2T} \right),$$

(30)

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 10.** The following two inequalities hold:

$$\mathbb{E}_{h,G} \ln Z(G) \leq \mathbb{E}_{h,G} \ln Z(G) + O(N^\gamma),$$

(31)

and

$$\mathbb{E}_{h,G} \ln Z(G) \leq \mathbb{E}_{h,G} \ln Z(G) + O(N^\gamma).$$

(32)

**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

$$\mathbb{E}_{h,G} \ln Z(G) \leq \mathbb{E}_{h,G} \ln Z(G) + O(N^\gamma).$$

(33)

The two ensembles involved in inequality (31) 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 $\{T-t+1\times \text{conn}\}$ and $\{T-t\times \text{conn}\}$ and in case $G$ is not null, adding an extra random check constraint sampled according to $\text{conn}$ and $\text{coup}$, respectively. The plan is to show that the inequality holds also when $G$ is fixed, and then to average over $G$.

Let us fix $m = dN^{1-\gamma}/2$, and let us first deal with the case when the realization of the ensemble $\{T-t+1\times \text{conn}\}$ is not $m$-admissible. This event occurs with a very small probability, subexponential according to Lemma 9. Since $\ln Z(G) = O(N)$ (according to Lemma 5), the error obtained by not considering this case is extremely small and fits in the tolerated term $O(1/m^{1/4})$.

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

$$\mathbb{E}_{h,G} \ln Z(G) \leq \mathbb{E}_{h,G} \ln Z(G).$$

We subtract $\ln Z(G)$ on both sides and then use Lemma 5 to write the difference of log partition functions as a linear combination of brackets of the from $\langle \sigma^{(1)} \ldots \sigma^{(r)} \rangle_G$, after which we can readily apply Lemma 5 and the claim follows.

**VII. RETRIEVING THE ORIGINAL LDPC ENSEMBLES**

We will now investigate further the connection between the ensembles $\{T \times \text{conn}\}$ and $\{T \times \text{disc}\}$. In fact, they are both variants of the uncoupled ensembles introduced in the beginning of Section II. The first one is very similar to LDPC$(NL, \Lambda, K)$, and the second one is similar to $L$ copies of 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 multiset of types generated by $\{T \times \text{conn}\}$ and $\{T \times \text{disc}\}$ 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 II, namely that the number of sockets at each position is $Nd/K \pm O(N^\gamma)$.

We can easily see that the ensemble $\{T \times \text{disc}\}$, 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 $\{T \times \text{conn}\}$ 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 LDPC$(NL, \Lambda, K)$ has a certain type. To see the crux of the problem, in the $\{T \times \text{conn}\}$ 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 LDPC(\(NL, \Lambda, K\)) (the simple ensemble) into \(\{T \times \text{conn}\}\) (the connected ensemble) through only a small amount of check additions and deletions. Let \(X_\alpha\) be the number of check nodes of type \(\alpha\) that occur in a realization of the simple ensemble. For every type \(\alpha\), let \(Y_\alpha\) be a random variable sampled according to \(\text{Bin}(T, L^{-K})\). If \(X_\alpha \geq Y_\alpha\), then exactly \(X_\alpha - Y_\alpha\) check nodes of type \(\alpha\) selected uniformly at random from the existing ones are deleted from the simple ensemble. Otherwise, exactly \(Y_\alpha - X_\alpha\) check nodes of type \(\alpha\) 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 \(Y_\alpha\)) 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 \(X_\alpha = \sum_i R^a_{\alpha i}\), where \(R^a_{\alpha i}\) is the indicator random variable of the event that check node \(\alpha\) has type \(a\), and the sum ranges over all \(M\) check nodes.

**Proposition 11.** The expectation and variance of \(X_\alpha - Y_\alpha\) are given by
\[
\mathbb{E}[X_\alpha - Y_\alpha] = O(N^{1-\gamma}), \quad (34)
\]
\[
\text{Var}[X_\alpha - Y_\alpha] = O(N^{2-\gamma}). \quad (35)
\]

**Proof:** We determine first the probability \(\mathbb{E}R^a_\alpha\) that a check node \(\alpha\) has type \(a\). This event happens if and only if all sockets \(a_i\) to which \(\alpha\) is connected are placed at positions \(\alpha_i\). 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 \(\alpha\) is the first check node to be allocated). The number of sockets at any position is between \(N\hat{d}(1-N^{-\eta})\) and \(N\hat{d}(1+N^{-\eta})\); the number of occupied sockets is at most \(K-1\) (from previous edges). Thus, the probability that \(\text{pos}(a_i) = \alpha_i\) is lower-bounded by
\[
\frac{N\hat{d}(1-N^{-\eta}) - K}{NLd(1+N^{-\eta})} = \frac{1}{L} - O(N^{-\eta}),
\]
and, likewise, upper-bounded by
\[
\frac{N\hat{d}(1+N^{-\eta})}{NLd(1-N^{-\eta})} = \frac{1}{L} + O(N^{-\eta}).
\]
It then follows that
\[
\mathbb{E}R^\alpha_\alpha = \left(\frac{1}{L} + O(N^{-\eta})\right)^K = \frac{1}{L^K} + O(N^{-\eta}). \quad (36)
\]
For the second moments we need \(\mathbb{E}[R^a_\alpha R^b_\beta]\), i.e. the probability that \(a\) and \(b\) have types \(\alpha\) and \(\beta\) 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
\[
\mathbb{E}[R^a_\alpha R^b_\beta] = \left(\frac{1}{L} + O(N^{-\eta})\right)^{2K} = \frac{1}{L^{2K}} + O(N^{-\eta}). \quad (37)
\]
By summing over all check nodes, we get \(\mathbb{E}X_\alpha = \frac{M}{L} + O(N^{1-\eta})\) and elementary calculations, \(\text{Var}X_\alpha = O(N^{2-\gamma})\). Since \(Y_\alpha\) is binomially distributed, and using \(T = M + O(N^{1-\gamma})\), we have
\[
\mathbb{E}Y_\alpha = T\frac{1}{L^K} = \frac{M}{L^K} + O(N^{1-\gamma}),
\]
and also
\[
\text{Var}Y_\alpha = T\frac{1}{L^K}(1 - \frac{1}{L^K}) = O(N),
\]
which is much smaller than \(\text{Var}X_\alpha\).

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 \(\zeta\) to be fixed shortly, reads
\[
\mathbb{P}\left[X_\alpha - Y_\alpha - O(N^{1-\gamma}) \right] \leq N^\zeta O\left(N^{1-\frac{\gamma}{2}}\right) \leq \frac{1}{N^{2\eta}}.
\]
We fix the values \(\zeta = \frac{2}{3}\) and \(\gamma = \frac{2}{3}\) (these choices are somewhat arbitrary), and simplifying we obtain
\[
\mathbb{P}\left[X_\alpha - Y_\alpha \right] \leq N^{\frac{2}{3}}.
\]

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 \(O(N^{-\eta/2})\). 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 5). This proves the following lemma.

**Lemma 12.** Transmitting over a BMS channel, we have
\[
\mathcal{E}_{h,G}\left(\text{LDPC}(NL, \Lambda, K)\right) \ln Z(G) \geq \mathcal{E}_{h,G}(T \{\text{conn}\}) \ln Z(G) + O\left(N^{1-\frac{\gamma}{2}}\right).
\]

**VIII. THE LARGE N LIMIT**

This section wraps up the proof of Theorem 2. The main ingredient is the content of Lemma 10 which can be written as
\[
\mathcal{E}_{h,G}(T \{\text{conn}\}) \ln Z(G) - O\left(N^{1-\gamma}\right) \leq \mathcal{E}_{h,G}(T \{\text{coup}\}) \ln Z(G) \leq \mathcal{E}_{h,G}(T \{\text{disc}\}) \ln Z(G) + O\left(N^{1-\gamma}\right). \quad (38)
\]
Using the results from the previous section on the comparison with the simple ensembles and scaling everything by \(NL\),
we obtain
\[
\frac{1}{NL} \mathbb{E}_{h,G:\text{LDPC}(NL,\Lambda,K)} \ln Z(G) - O(N^{1-\gamma}) \leq \\
\leq \frac{1}{NL} \mathbb{E}_{h,G:\{T\times\text{coup}\}} \ln Z(G) \leq \\
\leq \frac{1}{N} \mathbb{E}_{h,G:\text{LDPC}(N,\Lambda,K)} \ln Z(G) + O(N^{1-\gamma}). \tag{39}
\]

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 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
\[
\lim_{N \to \infty} \frac{1}{N} \mathbb{E}_{h,G:\text{LDPC}(N,\Lambda,K)} \ln Z(G)
\]
exists, we use the following result, whose proof can be found in the Appendix of [7].

**Lemma 13** (The modified superadditivity theorem). Given \(\alpha \in (0, 1)\), suppose a non-negative sequence \(\{a_{N,N \geq 1}\}\) satisfies
\[
a_{N_1+N_2} \geq a_{N_1} + a_{N_2} - O((N_1 + N_2)^\alpha) \tag{40}
\]
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 \(\ln Z(G)\) is 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
\[
\nu_1 = \frac{N_1}{N_1+N_2} \quad \text{and} \quad \nu_2 = \frac{N_2}{N_1+N_2},
\]
respectively.

The only difference appears in the reasoning of Lemma 3 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 \(\nu_1\), the other \(\nu_2\). The weight of the type is the product of the weights of the symbols it contains. If \(\alpha\) is a type, let \(\nu(\alpha)\) be the weight of that type. Then Equations (27) and (29) become
\[
\mathbb{E}_{\text{conn}}(Q_1, \cdots, Q_K) = \\
= \sum_{\alpha \in \{1,2\}^K} \nu(\alpha) (Q_1, \cdots, Q_K) = \langle (\nu_1 Q_1 + \nu_2 Q_2)^K \rangle,
\]
\[
\mathbb{E}_{\text{disc}}(Q_1, \cdots, Q_K) = \\
= \sum_{z \in \{1,2\}} \nu_z (Q_1, \cdots, Q_K) = \langle \nu_1 Q_1^K + \nu_2 Q_2^K \rangle,
\]
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 \(P(x) = \sum_{K \leq 0} p_K x^K\) 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 \(x \geq 0\), 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 [14]).

**ACKNOWLEDGEMENTS**

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

**APPENDIX**

**Proposition 14.** 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 2 we have that
\[
\frac{1}{|B_a\rangle} \sum_{a \in B_a} \langle \sigma^{(1)}_a \cdots \sigma^{(r)}_a \rangle = \\
= \frac{1}{|B'_a\rangle} \sum_{a \in B'_a} \langle \sigma^{(1)}_a \cdots \sigma^{(r)}_a \rangle + O\left(\frac{1}{m}\right). \tag{41}
\]

**Proof:** Rewrite the left hand side as
\[
\frac{1}{|B'_a\rangle} \sum_{a \in B'_a} \langle \sigma^{(1)}_a \cdots \sigma^{(r)}_a \rangle - \sum_{a \in B'_a \setminus B_a} \langle \sigma^{(1)}_a \cdots \sigma^{(r)}_a \rangle. \tag{42}
\]

We will first find an estimate of the quantity \(|B'_a \setminus B_a|\), 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 \(B'_a\) where \(a_i = a_j\) (i.e. edges \(i\) and \(j\) connect to the same socket), for some distinct \(i, j\) with \(1 \leq i, j \leq K\). The
cardinality $q_{i,j}$ of this subset is 0 if $\alpha_i \neq \alpha_j$, and is equal to $|B'_{\alpha_i}|/|F_i|$ 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 $|B'_{\alpha} \setminus B_{\alpha}|$ appearing in (42) by $m/(m - K^2)$. Indeed, this follows from $|B'_{\alpha} \setminus B_{\alpha}|/|B_{\alpha}| = |B'_{\alpha}|/|B_{\alpha}| - |B'_{\alpha} \setminus B_{\alpha}|$.

The absolute value of the second sum in (42) is clearly upper-bounded by $|B'_{\alpha} \setminus B_{\alpha}|$, since the bracket takes values between 0 and 1. Putting everything together, we obtain

$$\frac{1}{|B_{\alpha}|} \sum_{a \in B_{\alpha}} \langle \sigma_{a}^{(1)} \cdots \sigma_{a}^{(r)} \rangle$$

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

$$\frac{1}{|B_{\alpha}|} \sum_{a \in B_{\alpha}} \langle \sigma_{a}^{(1)} \cdots \sigma_{a}^{(r)} \rangle$$

$$\geq \frac{1}{|B'_{\alpha}|} \sum_{a \in B'_{\alpha}} \langle \sigma_{a}^{(1)} \cdots \sigma_{a}^{(r)} \rangle - \frac{K^2}{m - K^2}.$$

**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. Montanari, “The glassy phase of Gallager codes,” Eur. Phys. J. B, vol. 23, pp. 121–136, Sept. 2001.

[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] N. Macris, “Griffith-Kelly-Sherman correlation inequalities: a useful tool in the theory of error correcting codes,” IEEE Trans. Inform. Theory, vol. 55, pp. 664–683, Feb. 2007.

[14] S. Hamed Hassani, N. Macris, and R. Urbanke, “Threshold saturation in spatially coupled constraint satisfaction problems,” Journal of Statistical Physics, pp. 1–44, 2012.