And Now to Something Completely Different: 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 left 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 [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 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 [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 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.

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 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.

Owing to space constraints, we are not able to present the proofs here with all due rigor and details. For a complete version of this paper, we refer the interested reader to [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 a variable-node degree distribution with finite support, and the integer K is the fixed check-node degree. The average with respect to this distribution will be denoted by $\bar{d}$. 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 D, 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 $\lfloor D/K \rfloor$.

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 $\{1, \dots, L\}$.

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 only allowed to connect to sockets whose positions lie inside an interval $\{z, \dots, z+w-1\}$ (called *window* of length w) somewhere on the chain. 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, but the probability for this to happen is vanishing quickly in the limit $N \to \infty$. The steps in this process will be described in more detail in Section IV.

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 a small number of missing check nodes.

C. 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 $h(y) = \frac{1}{2} \ln \frac{p_{Y|X}(y|+1)}{p_{Y|X}(y|-1)}$, 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 σ was sent is proportional to $e^{h \cdot \sigma}$, where $h \cdot \sigma$ stands for $\sum_{v \in V} h_v \sigma_v$. The full expression for the posterior probability, is given by

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

where a iterates over check nodes, $a_1, \dots, a_K$ represent the neighbors of a, σ_a is short for the product $\sigma_{a_1} \cdots \sigma_{a_K}$, and $Z(G)$ is the *partition function*, the normalizing factor that turns μ into a probability measure.

The average with respect to the measure μ will appear quite often in the rest of the paper, so we use angular brackets (to be referred as *Gibbs brackets*) $\langle \cdot \rangle$ to indicate it.

Because of channel symmetry, we will assume without loss of generality that the all-$+1$ codeword is sent. Thus all randomness in the channel is captured by the i.i.d. HLLRs $h_v, v \in V$. When taking expectations with respect to the channel we will use the symbol $\mathbb{E}_h$, and whenever the expectation is with respect to an ensemble of graphs we will use $\mathbb{E}_{G:\mathcal{G}}$, where $\mathcal{G}$ denotes the ensemble of graphs. It is important to remember that the averages denoted by $\mathbb{E}$ and the bracketed average do not commute. 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 $H(X|Y)$ (where X is the input vector and Y the output vector), which carries more information-theoretic intuition and captured by the following widely known result (for the proof, see, for example, [12]).

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

$$H(X|Y) = \mathbb{E}_h \ln Z(G, h) - N\, \mathbb{E}_h[h] \, .$$

The meaning is that the two quantities are the same, up to a term that only depends on the channel (and not on the code). The $\ln Z(G, h)$ has the property that it is the sum of bounded contributions of individual check-nodes, as the following lemma suggests. The proof, making use of the Nishimori Identities, can be found in [11].

Lemma 2. *Given any graph G and an additional check constraint b, we have that*

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

To better cope with terms of the form $\langle \sigma_b \rangle_G^r$, we will work with the product measure $\mu^{\otimes r}$. The measure space here is the one of r-tuples $(\sigma^{(1)}, \dots, \sigma^{(r)})$, 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 $\langle \sigma_b \rangle_G^r = \left\langle \sigma_b^{(1)} \cdots \sigma_b^{(r)} \right\rangle_G$.

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, which states that the entropies of the simple and coupled ensembles are asymptotically the same in the large N limit.

Theorem 3. *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*

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{G:\mathrm{LDPC}(N,\Lambda,K)} H(X|Y) = \\ = \lim_{N\to\infty} \frac{1}{LN}\mathbb{E}_{G:\mathrm{LDPC}(N,L,w,\Lambda,K)} H(X|Y). \quad (2)$$

Consider a family of BMS channels ordered by degradation, indexed by a noise parameter ϵ. Then there exists a value ϵ_{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.

Corollary 4. *With the same assumptions as in Theorem 3 we have $\epsilon_{MAP} = \epsilon_{MAP}^{L,w}$, where $\epsilon_{MAP}^{L,w}$ is the MAP threshold for the coupled ensemble with L positions and window size 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 ϵ_{Area} be the area threshold defined as that value so that the integral of the BP-GEXIT curve over the interval $[\epsilon_{\mathrm{Area}}, 1]$ equals the design rate $1 - \bar{d}/K$ (for more details, see [3]). The *Maxwell conjecture* states that $\epsilon_{\mathrm{Area}} = \epsilon_{\mathrm{MAP}}$.

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 ϵ_{Area} (of the simple ensemble) in the following sense:

$$\epsilon_{\mathrm{Area}} - O(\frac{1}{w^{1/2}}) \leq \epsilon_{\mathrm{MAP}}^{L,w,\mathrm{open}} \leq \epsilon_{\mathrm{Area}} + O(\frac{w}{L}). \quad (3)$$

The threshold $\epsilon_{\mathrm{MAP}}^{L,w,\mathrm{open}}$ 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".

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 Lemma 2, the difference of the entropies is at most $O(w/L)$, which goes to 0 as $L \to \infty$. As a consequence,

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

Thus by (3) and Corollary 4, we deduce that in fact ϵ_{MAP} equals ϵ_{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.

IV. The configuration model

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. It 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*. Let S be the set of all sockets and S_z the set of sockets at a particular position z.

Check nodes will connect to sockets, so a check node a will have the form of a K-tuple $(a_1, \dots, a_K)$, 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* $\alpha = (\alpha_1, \dots, \alpha_K)$ if socket a_j is placed at position α_j along the chain, for all $1 \leq j \leq K$.

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 of the form $(z, z, \dots, z)$. 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 all types whose entries lie in the set $\{z, \dots z + w - 1\}$. We denote this distribution by $\mathbf{coup}$.

Let Γ be a multiset of types (i.e., for each type we specify how many times it appears). A graph G is said to be *compatible* with Γ if the types of its check nodes are the types in Γ. Γ can also be used to count how many used sockets are at each position, and Γ is said to be m-admissible if there are at least m sockets at each position that remain free.

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 $\alpha = (\alpha_1, \dots, \alpha_K)$ 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 $(a_1, \dots, a_K)$ 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 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 $F_z \subset S_z$ 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

$$Q_z(\sigma^{(1)}, \dots, \sigma^{(r)}) = \frac{1}{|F_z|} \sum_{s \in F_z} \sigma_s^{(1)} \cdots \sigma_s^{(r)}. \tag{4}$$

The next statement describes the link between the overlap functions and the replica averages introduced by Lemma 2.

Lemma 5. *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*

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

Proof sketch: The complete proof of this and all other lemmas can be found in [11]. 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,

$$\frac{1}{|B_\alpha|} \sum_{a \in B_\alpha} \left\langle \sigma_a^{(1)} \cdots \sigma_a^{(r)} \right\rangle. \tag{5}$$

The goal is to somehow factorize the sum, so we will modify the model a bit, namely we will allow sockets to be picked with replacement, independently from each other. The cost of this trick is adding $O(1/m)$ to the quantity (5), which can be roughly thought of as the probability to choose non-unique sockets. Then the result can be written as

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

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. ■

Lemma 6. *Let G be a graph whose type multiset is m-admissible, and fix the channel realization h. Then the following inequalities hold:*

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

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

Proof: The claim follows by Lemma 5 if we manage to show the following two inequalities:

$$\mathbb{E}_{\alpha:\mathbf{conn}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle \leq \mathbb{E}_{\alpha:\mathbf{coup}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle, \tag{6}$$

$$\mathbb{E}_{\alpha:\mathbf{coup}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle \leq \mathbb{E}_{\alpha:\mathbf{disc}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle, \tag{7}$$

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:

$$\mathbb{E}_{\alpha:\mathbf{conn}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle = \\ = \frac{1}{L^K} \sum_{\substack{(\alpha_1,\dots,\alpha_K) \\ \in [L]^K}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle = \left\langle \left(\frac{1}{L} \sum_{z \in [L]} Q_z \right)^K \right\rangle, \tag{8}$$

$$\mathbb{E}_{\alpha:\mathbf{coup}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle = \\ = \frac{1}{L} \sum_{z' \in [L]} \frac{1}{w^K} \sum_{\substack{(\alpha_1,\dots,\alpha_K) \\ \in \{z',\dots,z'+w-1\}^K}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle \\ = \left\langle \frac{1}{L} \sum_{z' \in [L]} \left(\frac{1}{w} \sum_{z=z'}^{z'+w-1} Q_z \right)^K \right\rangle, \tag{9}$$

$$\mathbb{E}_{\alpha:\mathbf{disc}} \langle Q_{\alpha_1} \cdots Q_{\alpha_K} \rangle = \\ = \frac{1}{L} \sum_{z \in [L]} \langle Q_z \cdots Q_z \rangle = \left\langle \frac{1}{L} \sum_{z \in [L]} Q_z^K \right\rangle. \tag{10}$$

Both inequalities (6) and (7) are proved by an application of Jensen's Inequality using the convexity of the function $x \mapsto x^K$, for even K. ■

V. THE INTERPOLATION

We now move a bit further and consider random ensembles of graphs.

We use the notation $G : \left\{ \begin{smallmatrix} t_1 \times \mathbf{coup} \\ t_2 \times \mathbf{disc} \end{smallmatrix} \right\}$ 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) $NL\bar{d}/K$. As a consequence, we choose the size 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 $0 < \gamma < 1/2$. Using Hoeffding's Inequality (since the types are drawn independently), one can easily see that by using this value for T, the resulting set of types is admissible with high probability. This essentially allows us to take the expectation over an ensemble of graphs without caring too much about non-admissibility. The interpolation is done as follows.

Lemma 7. *The following two inequalities hold:*

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

Proof sketch: 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

$$\mathbb{E}_{h,G:\left\{\substack{(t+1)\times\mathbf{conn}\\(T-t-1)\times\mathbf{coup}}\right\}}\ln Z(G) \leq \leq \mathbb{E}_{h,G:\left\{\substack{t\times\mathbf{conn}\\(T-t)\times\mathbf{coup}}\right\}}\ln Z(G) + O\left(N^{\gamma-1}\right). \quad (11)$$

The ensembles appearing in the lemma 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 $\left\{\substack{(t+1)\times\mathbf{conn}\\(T-t-1)\times\mathbf{coup}}\right\}$ and $\left\{\substack{t\times\mathbf{conn}\\(T-t)\times\mathbf{coup}}\right\}$ can both be obtained by sampling a graph $\widetilde{G}$ from their common part, $\left\{\substack{t\times\mathbf{conn}\\(T-t-1)\times\mathbf{coup}}\right\}$ 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 (11) 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 $\left\{\substack{t\times\mathbf{conn}\\(T-t-1)\times\mathbf{coup}}\right\}$ is not m-admissible. This event is negligible 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 2 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 6 and the claim follows. ■

Scaling everything by NL, and identifying the ensembles at the end of the chain as $\mathrm{LDPC}(NL,\Lambda,K)$ and $L\times\mathrm{LDPC}(N,\Lambda,K)$, we obtain

$$\frac{1}{NL}\mathbb{E}_{h,G:\mathrm{LDPC}(NL,\Lambda,K)}\ln Z(G) - O\left(N^{1-\gamma}\right) \leq \leq \frac{1}{NL}\mathbb{E}_{h,G:\{T\times\mathbf{coup}\}}\ln Z(G) \leq \leq \frac{1}{N}\mathbb{E}_{h,G:\mathrm{LDPC}(N,\Lambda,K)}\ln Z(G) + O\left(N^{1-\gamma}\right). \quad (12)$$

The next step is to take the $N\to\infty$ limit, and in case it exists for the outer terms, which is shown by exhibiting a superadditivity property of $\mathrm{LDPC}(N,\Lambda,K)$, we can apply the "sandwich rule" to obtain Theorem 3. Note that the ensemble appearing in the middle is what we call $\mathrm{LDPC}(N,L,w,\Lambda,K)$, which fits the requirements of [2].

VI. Conclusions

The present analysis can be extended with almost no change to arbitrary check-node degree distributions whose generating polynomial $P(x)=\sum_{K\leq 0}\rho_K x^K$ 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 $x\geq 0$, 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 [13]).

Acknowledgements

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

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, "Quadratic replica coupling for the Sherrington-Kirkpatrick mean field spin glass model," *J. Math. Phys.*, vol. 43, no. 7, pp. 3704–3716, 2002.

[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, "Spatial coupling as a proof technique." E-print arXiv:1301.5676, Jan. 2013.

[12] A. Montanari, "The glassy phase of Gallager codes," *Eur. Phys. J. B*, vol. 23, pp. 121–136, Sept. 2001.

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