Approaching the Rate-Distortion Limit With Spatial Coupling, Belief Propagation, and Decimation

Vahid Aref, Nicolas Macris, *Member, IEEE*, and Marc Vuffray

Abstract—We investigate an encoding scheme for lossy compression of a binary symmetric source based on simple spatially coupled low-density generator-matrix codes. The degree of the check nodes is regular and the one of code-bits is Poisson distributed with an average depending on the compression rate. The performance of a low complexity belief propagation guided decimation algorithm is excellent. The algorithmic rate-distortion curve approaches the optimal curve of the ensemble as the width of the coupling window grows. Moreover, as the check degree grows both curves approach the ultimate Shannon rate-distortion limit. The belief propagation guided decimation encoder is based on the posterior measure of a binary symmetric test-channel. This measure can be interpreted as a random Gibbs measure at a temperature directly related to the noise level of the test-channel. We investigate the links between the algorithmic performance of the belief propagation guided decimation encoder and the phase diagram of this Gibbs measure. The phase diagram is investigated thanks to the cavity method of spin glass theory which predicts a number of phase transition thresholds. In particular, the dynamical and condensation phase transition temperatures (equivalently test-channel noise thresholds) are computed. We observe that: 1) the dynamical temperature of the spatially coupled construction saturates toward the condensation temperature and 2) for large degrees the condensation temperature approaches the temperature (i.e., noise level) related to the information theoretic Shannon test-channel noise parameter of rate-distortion theory. This provides heuristic insight into the excellent performance of the belief propagation guided decimation algorithm. This paper contains an introduction to the cavity method.

Index Terms—Lossy source coding, rate-distortion bound, low-density generator matrix codes, belief propagation, decimation, spatial coupling, threshold saturation, spin glass, cavity method, density evolution, dynamical and condensation phase transitions.

I. INTRODUCTION

LOSSY source coding is one of the oldest and most fundamental problems in communications. The objective is to compress a given sequence so that it can be reconstructed up to some specified distortion. It was established long ago [1] that Shannon's rate distortion bound for binary sources (under Hamming distance) can be achieved using linear codes. However, it is of fundamental importance to find low complexity encoding schemes that achieve the rate distortion limit. An early attempt used trellis codes [2], for memoryless sources and bounded distortion measures. It is possible to approach the Shannon limit as the trellis constraint length increases, but the complexity of this scheme, although linear in the block length N, becomes exponential in the trellis constraint length. In [3] an entirely different scheme is proposed (also with linear complexity and diverging constants) based on the concatenation of a small code and optimal encoding of it. More recently, important progress was achieved thanks to polar codes [4] which were shown to achieve the rate-distortion bound with a successive cancellation encoder of complexity $O(N \ln N)$ [5]. Further work on the efficient construction of such codes followed [6].

Another interesting recent direction uses Low-Density Generator-Matrix (LDGM) codes as first investigated in [7] for binary erasure sources and in [8] for symmetric Bernoulli sources. LDGM based codes with Poisson degrees for code-bit nodes and regular degree for check nodes, achieve the ultimate Shannon rate-distortion limit under optimal encoding when the check degrees grow large. This conclusion was reached (by non-rigorous means) from the replica [8] and cavity [9] methods from statistical physics. This was later proved in [10] by second moment methods. These studies also showed that the gap to the rate-distortion bound vanishes exponentially in the large check degree limit.

In [9] it was recognized that using a plain message passing algorithm without decimation is not effective in lossy compression. Indeed the estimated marginals are either non-converging or non-biased because there exists an exponentially large number of compressed words that lead to roughly the same distortion. One has to supplement Belief Propagation (BP) (or Survey Propagation (SP)) with a decimation process. This yields and encoding scheme of low complexity.[1] In this respect the lossy compression schemes based on random graphs are an incarnation of random constraint satisfaction problems and, from this perspective it is not too surprising that their analysis share common features. The general idea of BP or SP guided-decimation algorithms is to: i) Compute approximate marginals by message passing; ii) Fix bits with the largest bias, and if there is no biased bit take a random decision; iii) Decimate

V. Aref is with Bell labs, Alcatel-Lucent AG, Murray Hill, NJ 07974 USA, and also with the Institute of Telecommunications, Stuttgart University, Stuttgart 70174, Germany (e-mail: vahid.aref@alcatel-lucent.com).

N. Macris is with the School of Computer and Communication Science, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: nicolas.macris@epfl.ch).

M. Vuffray is with the Theory Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87544 USA (e-mail: vuffray@lanl.gov).

[1] $O(N^2)$ or $O(N)$ depending on the exact implementation.

the graph and repeat this process on the smaller graph instance. For naive choices (say regular, or check-regular) of degree distributions the Shannon rate-distortion limit *is not* approached by such algorithms. However it has been observed that *it is* approached for degree distributions that have been optimized for channel LDPC coding [10]–[12]. These observations are empirical: it is not clear how to analyze the decimation process, and there is no real principle for the choice of the degree distribution.

In this contribution we investigate a *simple spatially coupled* LDGM construction. The degree distributions that we consider are regular on the check side and Poisson on the code-bit side. The average of the Poisson distribution is adjusted to achieve the desired compression rate. We explore a low complexity Belief Propagation Guided Decimation (BPGD) encoding algorithm, that takes advantage of spatial coupling, and approaches the Shannon rate-distortion limit for large check degrees and any compression rate. No optimization on the degree distributions is needed. The algorithm is based on the posterior measure of a test binary symmetric channel (BSC). We interpret this posterior as a random Gibbs measure with an inverse temperature parameter equal to the half-log-likelihood parameter of the test-BSC. This interpretation allows us to use the cavity method of spin glass theory in order to investigate the phase diagram of the random Gibbs measure. Although the cavity method is not rigorous, it makes definite predictions about the phase diagram of the measure. In particular it predicts the presence of phase transitions that allow to gain insight into the reasons for the excellent performance of the BPGD encoder on the spatially coupled lossy compression scheme.

Spatially coupled codes were first introduced in the context of channel coding in the form of convolutional LDPC codes [13] and it is now well established that the performance of such ensembles under BP decoding is consistently better than the performance of the underlying ensembles [14]–[16]. This is also true for coupled LDGM ensembles in the context of rateless codes [17]. The key observation is that the BP threshold of a coupled ensemble saturates towards the maximum a posteriori MAP threshold of the underlying ensemble as the width of the coupling window grows. A proof of this *threshold saturation phenomenon* has been accomplished in [18] and [19]. An important consequence is that spatially coupled regular LDPC codes with large degrees universally achieve capacity. Recently, more intuitive proofs based on replica symmetric energy functionals have been given in [20] and [21]. Spatial coupling has also been investigated beyond coding theory in other models such as the Curie-Weiss chain, random constraint satisfaction problems [22]–[24], and compressed sensing [25]–[27].

Let us now describe in more details the main contents of this paper. Summaries have appeared in [28] and [29]. In [28] we had investigated regular spatially coupled graph constructions with constant degrees for both check and code-bits. The performance of the BPGD algorithm are similar to the case of Poisson degree for code-bit nodes, on which we will concentrate here.

In section II we set up the framework for lossy source coding with spatially coupled LDGM ensembles for a binary symmetric Bernoulli source and Hamming distortion. We investigate ensembles with regular check degrees and Poisson code-bit node degrees. Important parameters of the spatial constructions are the number of positions L, the number of nodes n at each position, and the window width w over which we couple the nodes. The infinite block length limit investigated in this paper corresponds to $\lim_{L \to +\infty} \lim_{n \to +\infty}$ in the specified order. Optimal encoding consists in finding the compressed word that minimizes the Hamming distortion between a given source realization and the reconstructed word. Since we will use methods from statistical mechanics, we will translate the problem in this language. Optimal encoding can be viewed as the search for the minimum energy configurations of a random spin Hamiltonian. Although the optimal encoder is computationally impractical, it is important to determine the optimal distortion of the ensemble in order to set a limit on what cannot be achieved algorithmically for the ensemble. In this respect, an important rigorous result that is reviewed in section II is that, in the infinite block length limit $\lim_{L \to +\infty} \lim_{n \to +\infty}$, for any fixed w the optimal distortion for a spatially coupled ensemble is equal to the optimal distortion for the underlying uncoupled ensemble (and is therefore independent of w). This result follows from an equivalent one proved in [24] for the random XORSAT problem for any values of the constraint density. There are various results in the literature about the optimal encoder for the uncoupled ensemble. So we can essentially transfer them directly to our spatially coupled setting.

As explained in section II optimal encoding can be viewed as the study of the *zero temperature limit* of the Gibbs measure associated with a Hamiltonian. This Gibbs measure forms the basis of the BP based algorithms that we use. This Gibbs measure is nothing else than the posterior measure of the dual test-channel problem, and that the inverse temperature is the half-log-likelihood parameter of a test-BSC.[2] The free energies of the spatially coupled and underlying ensembles are the same [24] in the infinite block length limit (fixed w) and therefore their *static* phase transition temperature (the condensation temperature) is also the same (see below).

The Gibbs measure (or posterior measure of the dual test-channel problem) is the basis for setting up the BPGD algorithms. This is explained in detail in Section III. The crucial point is the use of the spatial dimension of the graphical construction. The main idea is that when the biases are small a random bit *from the boundary of the chain* is fixed to a random value, and as long as there exist bits with large biases they are eliminated from the chain by fixing them and decimating the graph. We consider two forms of BPGD. The first one, which as it turns out performs slightly better, is based on hard decisions. The second one uses a randomized rounding rule for fixing the bits.

Section IV reviews the simulation results and discusses the performance for the two versions of the BPGD encoders.

---

[2]More precisely, if p is the flip parameter of the BSC test-channel then the inverse temperature is $\beta = \frac{1}{2} \ln(\frac{1-p}{p})$.

For both algorithms we observe that the rate-distortion curve of the coupled ensemble approaches the Shannon limit when $n >> L >> w >> 1$ and the node degrees get large. We cannot assess if the Shannon limit is *achieved* based on our numerical results. However we observe that in order to avoid finite size effects the degrees have to become large only after the other parameters grow large in the specified order. In practice though $n = 2000$, $L = 64$, $w = 3$ and check degrees equal to $l = 3$ yield good results for a compression rate $1/2$. The performance of the BPGD algorithms depend on the inverse temperature parameter in the Gibbs measure, and one can optimize with respect to this parameter. Interestingly, for the coupled ensemble, we observe that for large degrees (when Shannon's rate-distortion limit is approached) the optimal parameter corresponds to the information theoretic value of the flip probability given by the Shannon distortion. This is non-trivial: indeed it is not true for the uncoupled ensemble.

The behavior of BPGD algorithms is to some extent controlled by the phase transitions in the phase diagram of the Gibbs measure. In section V we review the predictions of the cavity method, and in particular the predictions about the dynamical and condensation phase transition temperatures. At the condensation temperature the free energy displays a singularity and is thus a thermodynamic or static phase transition threshold. The dynamical temperature on the other hand is not a singularity of the free energy. As we will see in section V in the framework of the cavity method it is defined via a "complexity function" which counts the number of "pure states". The dynamical temperature is the value at which the complexity jumps to a non zero value. For a number of models it is known that Markov Chain Monte Carlo algorithms have an equilibration time which diverges at (and below) this dynamical temperature. Similarly, BPGD with randomized rounding correctly samples the Gibbs measure down to temperatures slightly higher than the dynamical threshold. We observe a threshold saturation phenomenon for the spatially coupled construction. First as said above, since the condensation threshold is a singularity of the free energy it is the same for the uncoupled and coupled ensembles for any w. Second, as the window width w grows the dynamical threshold saturates towards the condensation one. In practice we observe this saturation for values of w as low as $w = 3, 4, 5$. Thus for spatially coupled codes the BPGD algorithm is able to correctly sample the Gibbs measure down to a temperature approximately equal to the condensation threshold. This explains why the algorithm performs well, indeed it is able to operate at much lower temperatures than in the uncoupled case. A large degree analysis of the cavity equations shows that the condensation temperature tends to the information theoretic value corresponding to the flip parameter of the BSC test-channel given by Shannon's distortion. These facts, put together, provide insight as to the excellent performance of the BPGD algorithm for the spatially coupled ensemble.

Section VI presents the cavity equations for the coupled ensemble on which the results of the previous paragraph are based. These equations are solved by population dynamics in Section VII. The cavity equations take the form of six fixed point integral equations. However we observe by population dynamics that two of them are satisfied by a trivial fixed point. This is justified by a theoretical analysis in Section VIII. When this trivial fixed point is used the remaining four equations reduce to two fixed point integral equations which have the form of usual density evolution equations for a BSC channel. This simplification is interesting because although the original Gibbs measure does not possess channel symmetry,[3] this symmetry emerges here as a solution of the cavity equations. Within this framework the saturation of the dynamical temperature towards the condensation one appears to be very similar than threshold saturation in the context of channel coding with LDPC codes. A proof of threshold saturation for the present problem is beyond the scope of this paper, but we do give in Section VIII a few insights on possible ways to attack it.

For an introduction to the cavity theory we refer the reader to the book [30]. This theory is not easy to grasp both conceptually and technically. This paper contains a high level introduction of the main concepts in Section V and a summary of the main technical ideas in Appendix A. We hope that this will be helpful for unfamiliar readers. The necessary derivations and adaptations to the present setting of a spatially coupled Gibbs measure are summarized in Appendices B and C. The main sections II-V and the conclusion can be read without explicitly going into the cavity formalism.

II. COUPLED LDGM ENSEMBLES FOR LOSSY COMPRESSION

A. Lossy Compression of Symmetric Bernoulli Sources

Let $\underline{X} = \{X_1, X_2, \ldots, X_N\}$ represent a source of length N, where X_a, $a = 1, \ldots, N$ are i.i.d Bernoulli$(1/2)$ random variables. We compress a source word $\underline{x}$ by mapping it to one of 2^{NR} index words $\underline{u} \in \{0, 1\}^{NR}$, where $R \in [0, 1]$ is the compression rate. This is the encoding operation. The decoding operation maps the stored sequence $\underline{u}$ to a reconstructed sequence $\widehat{\underline{x}}(\underline{u}) \in \{0, 1\}^N$.

For a given pair $(\underline{x}, \widehat{\underline{x}})$, we measure the distortion by the relative Hamming distance

$$d_N(\underline{x}, \widehat{\underline{x}}) = \frac{1}{N} \sum_{a=1}^{N} |x_a - \widehat{x}_a|. \tag{1}$$

The quality of reconstruction is measured by the average distortion

$$D_N(R) = \mathbb{E}_{\underline{X}}[d_N(\underline{x}, \widehat{\underline{x}})] \tag{2}$$

where $\mathbb{E}_{\underline{X}}$ is the expectation with respect to the symmetric Bernoulli source.

For the symmetric Bernoulli source considered here, it is well-known that for any encoding-decoding scheme, the average distortion is lower bounded by Shannon's rate-distortion curve [31]

$$D_{\text{sh}}(R) = h_2^{-1}(1 - R) \tag{3}$$

where $h_2(x) = -x \log_2 x - (1 - x) \log_2(1 - x)$ is the binary entropy function. The rate-distortion curve is convex decreasing with $D_{\text{sh}}(0) = 1/2$ and $D_{\text{sh}}(1) = 0$.

---

[3]In the context of spin glass theory this is the Nishimori gauge symmetry.
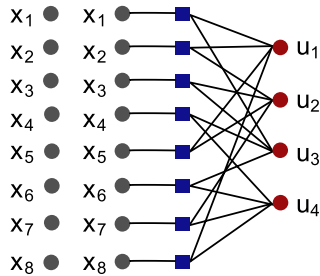
Fig. 1. A bipartite graph from the underlying LDGM $(2, 0.5)$ ensemble. Here $n = 8$, $m = 4$ and $l = 2$. Labels represent code-bits u_i, reconstructed bits $\hat{x}_i$ and source bits x_i.
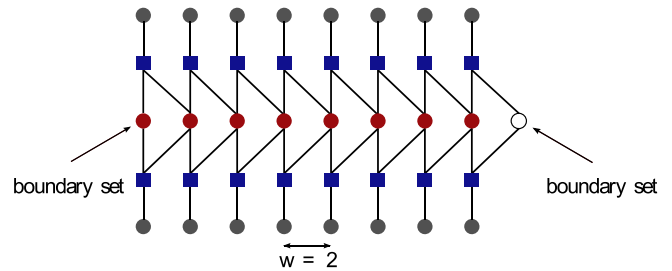

Fig. 2. The "protograph" representation of the spatially coupled LDGM $(2, 0.5, L = 8, w = 2)$ ensemble. The code-bit nodes in boundary sets have smaller degree than the code-bit nodes in the other sets.

B. Spatially Coupled Low-Density Generator Matrix Constructions

Our lossy source coding scheme is based on a spatially coupled LDGM code ensemble. We first describe the *underlying* ensemble.

1) Underlying Poisson LDGM(l, R) Ensemble: These are bipartite graphs with a set C of n check nodes of constant degree l, a set V of m code-bit nodes of variable degree, and a set E of edges connecting C and V. The ensemble of graphs is generated as follows: each edge emanating from a check node is connected uniformly at random to one of the code-bit nodes. The degree of code-bit nodes is a random variable with Binomial distribution Bi($ln, 1/m$). In the asymptotic regime of large n, m with $m/n = R$ the code-bit node degrees are i.i.d Poisson distributed with an average degree l/R. Note that this construction allows the possibility to have multi-edges in the graph.

2) Spatially Coupled LDGM(l, R, L, w, n) Ensemble: We first lay out a set of positions indexed by integers $z \in \mathbb{Z}$ on a one dimensional line. This line represents a "spatial dimension". We fix a "window size" which is an integer $w \geq 1$. Consider L sets of check nodes each having n nodes, and locate the sets in positions 1 to L. Similarly, locate $L + w - 1$ sets of m code-bit nodes each, in positions 1 to $L+w-1$. All checks have constant degree l, and each of the l edges emanating from a check at position $z \in \{1, \ldots, L\}$ is connected uniformly at random to code-bit nodes within the range $\{z, \ldots, z + w - 1\}$. It is easy to see that for $z \in \{w, \ldots, L - w + 1\}$, in the asymptotic limit $n \to +\infty$, the code-bit nodes have Poisson degrees with average l/R. For the remaining positions close to the boundary the average degree is reduced. More precisely for positions on the left side $z \in \{1, \ldots, w - 1\}$ the degree is asymptotically i.i.d Poisson with average $l/R \times z/w$. For positions on the right side $z \in \{L + 1, \ldots, L + w - 1\}$ the degree is asymptotically Poisson with average $l/R \times (L + w - z)/w$. Figures 1 and 2 give a schematic view of an underlying and a spatially coupled graph.

3) Notation: Generic graphs from the ensembles will be denoted by Γ or $\Gamma(C, V, E)$. We will use letters a, b, c for check nodes and letters i, j, k for code-bit nodes of a given graph (from underlying or coupled ensembles). We will often make use of the notation ∂a for the set of all code-bit nodes connected to $a \in C$, i.e. $\partial a = \{i \in V \,|\, (i, a) \in E\}$.

Similarly, for $i \in V$, $\partial i = \{a \in C \,|\, (i, a) \in E\}$. For spatially coupled graphs the sets of nodes at a specified position z are C_z and V_z.

C. Decoding Rule and Optimal Encoding

We "attach" a code bit u_i to each code-bit node $i \in V$. To each check node $a \in C$ we "attach" two type of bits: the reconstructed bit $\hat{x}_a$ and the source bit x_a. By definition the source sequence has length N. So we have $n = N$ for the underlying ensembles, and $nL = N$ for the coupled ensembles. A compressed word $\underline{u}$ has length m for the underlying ensemble, and $m(L + w - 1)$ for the coupled ensemble. Thus the compression design rate is $R = m/n$ for the underlying ensemble, and it is $R_{\text{cou}} = m(L+w-1)/nL = R(1+\frac{w-1}{L})$ for the coupled ensemble. The compression design rate of the coupled ensembles is slightly higher, due to the code-bit nodes at the boundary, but in the asymptotic regime $n, m >> L >> w$ the difference between the design rate R of the underlying ensemble vanishes.

1) Decoding Rule: The reconstruction mapping is given by the linear operation (modulo 2 sum)

$$\hat{x}_a(\underline{u}) = \oplus_{i \in \partial a} u_i. \tag{4}$$

In this paper we do not investigate non-linear decoding rules, although the whole analysis developed here can be adapted to such rules. Source coding with such "non-linear check nodes" have been investigated for underlying LDGM(l, R) ensembles [32].

2) Optimal Encoding: Given a source word $\underline{x}$, the optimal encoder seeks to minimize the Hamming distortion (1), and so searches among all $\underline{u} \in \{0, 1\}^{NR}$ to find a configuration $\underline{u}^*$ such that

$$\underline{u}^* = \text{argmin}_{\underline{u}} d_N\left(\underline{x}, \hat{\underline{x}}(\underline{u})\right). \tag{5}$$

The resulting minimal distortion is

$$d_{N,\text{min}}(\underline{x}) = \min_{\underline{u}} d_N\left(\underline{x}, \hat{\underline{x}}(\underline{u})\right). \tag{6}$$

3) Optimal Distortion of the Ensemble: A performance measure is given by the *optimal distortion of the ensemble* (not to be confused with Shannon's optimal distortion)

$$D_{N,\text{opt}} = \mathbb{E}_{\text{LDGM}, \underline{X}}[d_{N,\text{min}}(\underline{x})] \tag{7}$$

where $\mathbb{E}_{\text{LDGM}, \underline{X}}$ is an expectation over the graphical ensemble at hand and the symmetric Bernoulli source $\underline{X}$.

Finding the minimizers in (5) by exhaustive search takes exponential time in N; and there is no known efficient algorithmic procedure to solve the minimization problem. Nevertheless, the cavity method proposes a formula for the asymptotic value of (7) as $N \to +\infty$. It is conjectured that this formula is exact. We come back to this point at the end of paragraph II-D.

D. Statistical Mechanics Formulation

We equip the configuration space $\{0, 1\}^{NR}$ with the conditional probability distribution (over $\underline{u} \in \{0, 1\}^{NR}$)

$$\mu_\beta(\underline{u} \mid \underline{x}) = \frac{1}{Z_\beta(\underline{x})} e^{-2\beta N d_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))}$$

$$= \frac{1}{Z_\beta(\underline{x})} \prod_{a \in C} e^{-2\beta |x_a - \oplus_{i \in \partial a} u_i|} \quad (8)$$

where $\beta > 0$ is a real number and

$$Z_\beta(\underline{x}) = \sum_{\underline{u}} e^{-2\beta N d_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))} \quad (9)$$

a normalizing factor. The expectation with respect to $\underline{u}$ is denoted by the bracket $\langle - \rangle$. More precisely the average of a function $A(\underline{u})$ is

$$\langle A(\underline{u}) \rangle = \frac{1}{Z} \sum_{\underline{u} \in \{-1, +1\}^N} A(\underline{u}) e^{-2\beta N d_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))}. \quad (10)$$

An important function that we consider below is the distortion of a pair $(\underline{x}, \widehat{\underline{x}}(\underline{u}))$, $A(\underline{u}) = d_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))$.

Note that the minimizer $\underline{u}^*$ in (5) maximizes this conditional distribution,

$$\underline{u}^* = \operatorname{argmax}_{\underline{u}} \mu_\beta \left(\underline{u} \mid \underline{x} \right). \quad (11)$$

The source coding problem can thus be interpreted as an estimation problem where $\underline{x}$ is an observation and $\underline{u}$ has to be estimated.

In this paper we prefer the statistical mechanics interpretation, because we use related methods and concepts. Equation (8) defines the Gibbs distribution associated to a "spin glass" Hamiltonian $2N d_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))$. This Hamiltonian is a cost-function for assignments of "dynamical" variables, the spins (or bits) $u_i \in \{0, 1\}$. The Hamiltonian is random: for each realization of the source sequence $\underline{x}$ and the graph instance we have a different realization of the cost-function. The source and graph instance are qualified as "quenched" or "frozen" random variables, to distinguish them from dynamical variables, because in physical systems - as well as in algorithms - they fluctuate on vastly different time scales. The parameter β is the "inverse temperature" in appropriate units, and the normalizing factor (9) is the partition function.

Finding $\underline{u}^*$ amounts to find the "minimum energy configuration". The minimum energy per node is equal to $2d_{N,\min}$, and it is easy to check the identity (use 6 and 9)

$$2d_{N,\min}(\underline{x}) = - \lim_{\beta \to \infty} \frac{1}{\beta N} \ln Z_\beta(\underline{x}). \quad (12)$$

TABLE I
OPTIMAL DISTORTION FOR LDGM (l, $R = 0.5$) ENSEMBLES COMPUTED IN [9]; SHANNON'S BOUND FOR $R = 0.5$ IS $D_{sh} \approx 0.1100$

l	3	4	5	6
D_{opt}	0.1179	0.1126	0.1110	0.1104

As this identity already shows, a fundamental role is played by the average free energy

$$f_N(\beta) = -\frac{1}{\beta N} \mathbb{E}_{\text{LDGM}, \underline{X}}[\ln Z_\beta(\underline{x})]. \quad (13)$$

For example the average free energy allows to compute the optimal distortion of the ensemble

$$2D_{N,\text{opt}} = \lim_{\beta \to +\infty} f_N(\beta). \quad (14)$$

There exists also another useful relationship that we will use between average distortion and free energy. Consider the "internal energy" defined as

$$u_N(\beta) = 2\mathbb{E}_{\text{LDGM}, \underline{X}}[\langle d_N(\underline{x}, \widehat{\underline{x}}(\underline{u})) \rangle] \quad (15)$$

It is straightforward to check that the internal energy can be computed from the free energy (use (9), (13), (15))

$$u_N(\beta) = \frac{\partial}{\partial \beta}(\beta f_N(\beta)) \quad (16)$$

and that in the zero temperature limit it reduces to the average minimum energy or optimal distortion (use (6), (7), (15))

$$2D_{N,\text{opt}} = \lim_{\beta \to +\infty} u_N(\beta). \quad (17)$$

What is the relation between the quantities $f_N(\beta)$, $u_N(\beta)$, and $D_{N,\text{opt}}$ for the underlying and coupled ensembles? The following theorem states that they are equal in the infinite block length limit. This limit is defined as

$$\lim_{N \to +\infty} = \lim_{n \to +\infty}$$

with m/n fixed for the underlying ensemble; and as

$$\lim_{N \to +\infty} = \lim_{L \to +\infty} \lim_{n \to +\infty}$$

with m/n fixed for the coupled ensemble. We stress that for the coupled ensemble the order of limits is important.

Theorem 1: Consider the two ensembles LDGM(l, R, n) and LDGM(l, R, L, w, n) for an even l and R. Then the respective limits $\lim_{N \to +\infty} f_N(\beta)$, $\lim_{N \to +\infty} u_N(\beta)$ and $\lim_{N \to +\infty} D_{N,\text{opt}}$ exist and have identical values for the two ensembles.

This theorem is proved in [24] for the max-XORSAT problem. The proof in [24] does not depend on the constraint density, so that it applies verbatim to the present setting. We conjecture that this theorem is valid for a wider class of graph ensembles. In particular we expect that it is valid for odd l and also for the regular LDGM ensembles (see [33] for similar results concerning LDPC codes).

It is conjectured that the one-step-replica-symmetry-breaking-formulas (1RSB), obtained from the cavity

method [34], for the $N \rightarrow +\infty$ limit of the free, internal and ground state energies are exact. Remarkably, it has been proven [35], using an extension of the Guerra-Toninelli interpolation bounds [36], that these formulas are upper bounds. The 1RSB formulas allow to numerically compute [9], using population dynamics, $D_{\text{opt}} \equiv \lim_{N \rightarrow +\infty} D_{N,\text{opt}}$. As an illustration, Table I reproduces D_{opt} for increasing check degrees. Note that D_{opt} approaches D_{sh} as the degrees increase. One observes that with increasing degrees the optimal distortion of the ensemble attains Shannon's rate-distortion limit.

III. BELIEF PROPAGATION GUIDED DECIMATION

Since the optimal encoder (5) is intractable, we investigate suboptimal low complexity encoders. In this contribution we focus on two encoding algorithms based on the belief propagation (BP) equations supplemented with a decimation process.

1) Belief Propagation Equations: Instead of estimating the block $\underline{u}$ (as in (5)) we would like to estimate bits u_i with the help of the marginals

$$\mu_i(u_i \mid \underline{x}) = \sum_{\underline{u} \setminus u_i} \mu_\beta(\underline{u} \mid \underline{x}) \tag{18}$$

where the sum is over $u_1, \ldots u_N$ with u_i omitted. However computing the exact marginals involves a sum with an exponential number of terms and is also intractable. For sparse random graphs, when the size of the graph is large, any finite neighborhood of a node i is a tree with high probability. As is well known, computing the marginals on a tree-graph can be done exactly and leads to the BP equations. It may therefore seem reasonable to compute the BP marginal distribution in place of (18),

$$\mu_i^{\text{BP}}(u_i \mid \underline{x}) = \frac{1}{2 \cosh \beta \eta_i} e^{\beta(-1)^{u_i} \eta_i} \tag{19}$$

where the biases η_i are computed from solutions of the BP equations. The later are a set of fixed point equations involving $2|E|$ real valued messages $\eta_{i \rightarrow a}$ and $\widehat{\eta}_{a \rightarrow i}$ associated to the edges $(i, a) \in E$ of the graph. We have

$$\begin{cases} \widehat{\eta}_{a \rightarrow i} = (-1)^{x_a} \beta^{-1} \tanh^{-1}\big(\tanh \beta \prod_{j \in \partial a \setminus i} \tanh \beta \eta_{j \rightarrow a}\big) \\ \eta_{i \rightarrow a} = \sum_{b \in \partial i \setminus a} \widehat{\eta}_{b \rightarrow i} \end{cases} \tag{20}$$

and

$$\eta_i = \sum_{a \in \partial i} \widehat{\eta}_{a \rightarrow i}. \tag{21}$$

The derivation of these equations can be worked out by reducing the general BP equations (64) (Appendix A) with the parameterization (86) (Appendix B).

For any solution of the BP equations one may consider the estimator

$$\begin{aligned} \widehat{u}_i^{\text{BP}} &= \text{argmax}_{u_i} \mu_i^{\text{BP}}(u_i \mid \underline{x}) \\ &= \begin{cases} \frac{1}{2}(1 - \text{sign} \eta_i), & \text{if } \eta_i \neq 0 \\ \text{Bernoulli}(\frac{1}{2}), & \text{if } \eta_i = 0 \end{cases} \end{aligned} \tag{22}$$

One may then use the decoding rule (4) to determine a reconstructed word and the corresponding distortion.

To solve the BP equations one uses an iterative method. A set of initial messages $\eta_{i \rightarrow a}^{(0)}$ are fixed at time $t = 0$ and updated according to

$$\begin{cases} \widehat{\eta}_{a \rightarrow i}^{(t)} = (-1)^{x_a} \beta^{-1} \tanh^{-1}\big(\tanh \beta \prod_{j \in \partial a \setminus i} \tanh \beta \eta_{j \rightarrow a}^{(t)}\big) \\ \eta_{i \rightarrow a}^{(t+1)} = \sum_{b \in \partial i \setminus a} \widehat{\eta}_{b \rightarrow i}^{(t)} \end{cases}$$

The bias at time t is simply given by $\eta_i^{(t)} = \sum_{a \in \partial i} \widehat{\eta}_{a \rightarrow i}^{(t)}$.

Unfortunately, even when the BP updates are converging they are not always biased. This is because there exist an exponentially large (in N) number of compressed words that lead to roughly the same distortion. This has an undesirable consequence: it is not possible to pick the relevant solution by a plain iterative solution of the BP equations. To get around this problem, the BP iterations are equipped with a heuristic decimation process explained in the next paragraph. We note that here BP always has to be equipped with a decimation process for all values of parameters of the algorithm, whether the BP fixed point is unique or non-unique. The problem here is akin to the class of constraint satisfaction problems.

2) Decimation Process: We start with a description of the *first round* of the decimation process. Let Γ, $\underline{x}$ be a graph and source instance. Fix an initial set of messages $\eta_{i \rightarrow a}^{(0)}$ at time $t = 0$. Iterate the BP equations (20) to get a set of messages $\eta_{i \rightarrow a}^{(t)}$ and $\widehat{\eta}_{a \rightarrow i}^{(t)}$ at time $t \geq 0$. Let $\epsilon > 0$ be some small positive number and T some large time. Define a *decimation instant* t_{dec} as follows:

- i) If the total variation of messages does not change significantly in two successive iterations,

$$\frac{1}{|E|} \sum_{(i,a) \in E} |\widehat{\eta}_{a \rightarrow i}^{(t)} - \widehat{\eta}_{a \rightarrow i}^{(t-1)}| < \epsilon \tag{23}$$

for some $t < T$, then $t_{\text{dec}} = t$.
- ii) If (23) does not occur for all $t \leq T$ then $t_{\text{dec}} = T$.

At instant t_{dec} each code-bit has a bias given by $\eta_i^{(t_{\text{dec}})}$. *Select* and *fix one particular code-bit* i_{dec} according to a decision rule

$$(i_{\text{dec}}, u_{i_{\text{dec}}}) \leftarrow \mathcal{D}(\underline{\eta}^{(t_{\text{dec}})}). \tag{24}$$

The precise decision rules that we investigate are described in the next paragraph. At this point, update $x_a \leftarrow x_a \oplus u_{i_{\text{dec}}}$ for all $a \in \partial i_{\text{dec}}$, and decimate the graph $\Gamma \leftarrow \Gamma \setminus i_{\text{dec}}$. This defines a new graph and source instance, on which we repeat a *new round*. The initial set of messages of the new round is the one obtained at time t_{dec} of the previous round.

3) Belief-Propagation Guided Decimation: The decision rule (24) involves two choices. One has to choose i_{dec} and then set $u_{i_{\text{dec}}}$ to some value. Let us first describe the choice of i_{dec}.

We evaluate the maximum bias

$$B_{t_{\text{dec}}} = \max_{i \in V} |\eta_i^{(t_{\text{dec}})}| \tag{25}$$

at each decimation instant. If $B_{t_{\text{dec}}} > 0$, we consider the set of nodes that maximize (25), we choose one of them uniformly at random, and call it i_{dec}. If $B_{t_{\text{dec}}} = 0$ and we have a graph of the *underlying ensemble*, we choose a node uniformly at random from $\{1, \ldots m\}$, and call it i_{dec}. If $B_{t_{\text{dec}}} = 0$ and we have a

graph of the *coupled ensemble*, we choose a node uniformly at random from the w left-most positions of the current graph, and call it i_{dec}. Note that because the graph gets decimated the w left-most positions of the current graph form a moving boundary.

With the above choice of decimation node the encoding process is seeded at the boundary each time the BP biases fail to guide the decimation process. We have checked that if we choose i_{dec} uniformly at random from the whole chain (for coupled graphs) the performance is not improved by coupling. In [28] we adopted periodic boundary conditions and the seeding region was set to an arbitrary window of length w at the beginning of the process, which then generated its own boundary at a later stage of the iterations.

We now describe two decision rules for setting the value of $u_{i_{\text{dec}}}$ in (24).

1) **Hard Decision**

$$u_{i_{\text{dec}}} = \begin{cases} \frac{1}{2}(1 - \text{sign}\eta_{i_{\text{dec}}}^{(t_{\text{dec}})}), & \text{if } B_{t_{\text{dec}}} > 0 \\ \text{Bernoulli}(\frac{1}{2}), & \text{if } B_{t_{\text{dec}}} = 0 \end{cases} \quad (26)$$

where $\theta(.)$ is the Heaviside step function. We call this rule and the associated algorithm BPGD-h.

2) **Randomized Decision**

$$u_{i_{\text{dec}}} = \begin{cases} 0, & \text{with prob } \frac{1}{2}(1 + \tanh\beta\eta_{i_{\text{dec}}}^{(t_{\text{dec}})}) \\ 1, & \text{with prob } \frac{1}{2}(1 - \tanh\beta\eta_{i_{\text{dec}}}^{(t_{\text{dec}})}). \end{cases} \quad (27)$$

In other words, we fix a code-bit randomly with a probability given by its BP marginal (19). We call this rule and the associated algorithm BPGD-r.

Algorithm 1 summarizes the BPGD algorithms for all situations.

4) Initialization and Choice of Parameters ϵ, T: We initialize $\eta_{i \to a}^{(0)}$ to zero just at the beginning of the algorithm. After each decimation step, rather than resetting messages to zero we continue with the previous messages. We have observed that resetting the messages to zero does not lead to very good results.

The parameters ϵ and T are in practice set to $\epsilon = 0.01$ and $T = 10$. The simulation results do not seem to change significantly when we take ϵ smaller and T larger.

5) Choice of β: Let us now clarify the role of β. It may seem from the discussion of the statistical mechanical formulation in section II that β should be taken to $+\infty$. This is the case for the computation of the optimal ensemble performance. However for the BPGD algorithm this *is not* the best choice for β. The reason being that for large values of β the BP iterations do not converge and therefore one does not obtain a reliable bias.

We indeed observe that the performance of the BPGD algorithm does depend on the choice of β which enters in the BP equations (20) and in the randomized decision rule (27). It is possible to optimize on β. This is important in order to approach (with coupled codes) the optimal distortion of the ensemble, and furthermore to approach the Shannon bound in the large degree limit.

While we do not have a first principle theory for the optimal choice of β we provide empirical observations in section IV.

Algorithm 1 BP Guided Decimation Algorithm

1 Generate a graph instance $\Gamma(C, V, E)$ from the underlying or coupled ensembles. ;

2 Generate a Bernoulli symmetric source word $\underline{x}$.;

3 Set $\eta_{i \to a}^{(0)} = 0$ for all $(i, a) \in E$.;

4 **while** $V \neq \emptyset$ **do**

5 $\quad$ Set $t = 0$.;

6 $\quad$ **while** *Convergence* (23) *is not satisfied and $t < T$* **do**

7 $\quad\quad$ Update $\hat{\eta}_{a \to i}^{(t)}$ according to (20) for all $(a, i) \in E$.;

8 $\quad\quad$ Update $\eta_{i \to a}^{(t+1)}$ according to (20) for all $(i, a) \in E$.;

9 $\quad\quad$ $t \leftarrow t + 1$.;

10 $\quad$ Compute bias $\eta_i^{(t)} = \sum_{a \in \partial i} \hat{\eta}_{a \to i}^{(t)}$ for all $i \in V$;

11 $\quad$ Find $B = \max_{i \in V} |\eta_i^{(t)}|$.;

12 $\quad$ **if** $B = 0$ **then** For an instance from the underlying ensemble randomly pick a code-bit i from V. For a graph from the coupled ensemble randomly pick a code-bit from the w left-most positions of Γ and fix it randomly to 0 or 1.;

13 $\quad$ **else**

14 $\quad\quad$ Select $i = \arg\max_{i \in V} |\eta_i^{(t)}|$.;

15 $\quad\quad$ Fix a value for u_i according to rule (26) or (27).;

16 $\quad$ Update $x_a \leftarrow x_a \oplus u_i$ for all $a \in \partial i$.;

17 $\quad$ Reduce the graph $\Gamma \leftarrow \Gamma \setminus \{i\}$.;

We observe that knowing the dynamical and condensation (inverse) temperatures predicted by the cavity method allows to make an educated guess for an estimate of the optimal β. Two results (discussed at more length in the next section) are noteworthy: (i) for coupled instances we can take larger values of β; and (ii) for coupled codes with large degrees the best β approaches the information theoretic test-channel value.

6) Computational Complexity: It is not difficult to see that the complexity of the plain BPGD algorithm 1 is $O(N^2)$, in other words $O(n^2)$ for underlying and $O(n^2 L^2)$ for coupled ensembles. By employing window decoding [37], [38], one can reduce the complexity of the coupled ensemble to $O(n^2 L)$ with almost the same performance. This can be further reduced to $O(nL)$ by noticing that the BP messages do not change significantly between two decimation steps. As a result, we may decimate δn code-bits at each step for some small δ, so that the complexity becomes $O(nL/\delta)$. To summarize, it is possible to get linear in block length complexity without significant loss in performance.

IV. SIMULATIONS

In this section we discuss the performance of the BPGD algorithms. The comparison between underlying ensembles LDGM(l, R, n), coupled ensembles LDGM(l, R, w, L, n) and the Shannon rate-distortion curve is illustrated. The role played by the parameter β is investigated.

A. BPGD Performance and Comparison to the Shannon Limit

Fig. 3 and 4 display the average distortion $D_{\text{BPGD}}(R)$ obtained by the BPGD algorithms (with hard and randomized
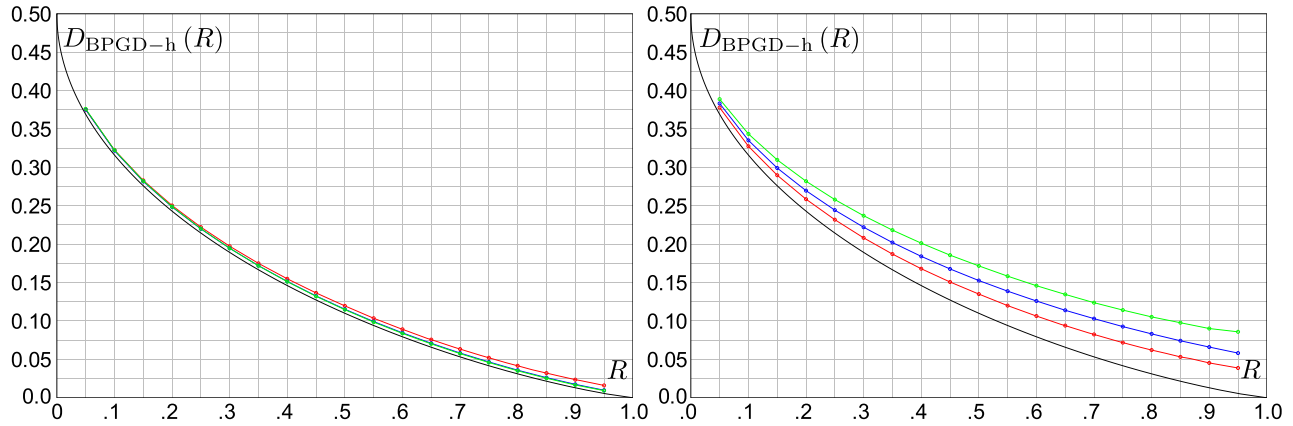
Fig. 3. The BPGD-h algorithmic distortion versus compression rate R compared to the Shannon rate-distortion curve at the bottom. Points are obtained by optimizing over β and averaging over 50 instances. Left: spatially coupled LDGM ($l, R, L = 64, w = 3, n = 2000$) ensembles for $l = 3, 4, 5$ (top to bottom). Right: LDGM ($l, R, N = 128000$) ensembles for $l = 3, 4, 5$ (bottom to top).
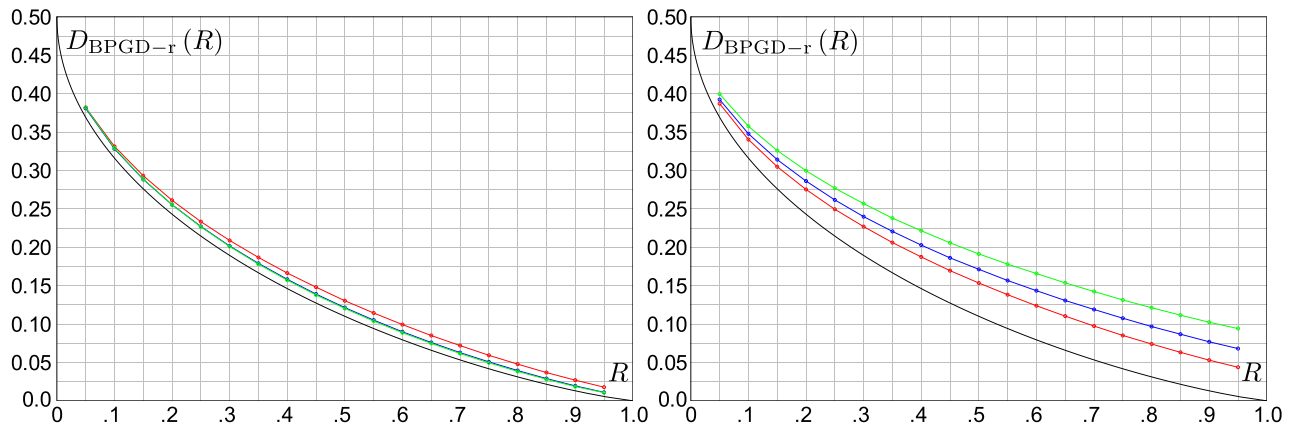

Fig. 4. The BPGD-r algorithmic distortion versus compression rate R compared to the Shannon rate-distortion curve at the bottom. Points are obtained by choosing $\beta = \beta_{sh} = \frac{1}{2} \log(\frac{1 - D_{sh}}{D_{sh}})$ and averaging over 50 instances. Continuous lines are a guide to the eye. Left: spatially coupled LDGM ($l, R, L = 64, w = 3, n = 2000$) ensembles for $l = 3, 4, 5$ (top to bottom). Right: LDGM ($l, R, N = 128000$) ensembles for $l = 3, 4, 5$ (bottom to top).

decision rules) as a function of R, and compares it to the Shannon limit $D_{\text{sh}}(R)$ given by the lowest curve. The distortion is computed for fixed R and for 50 instances, and the empirical average is taken. This average is then optimized over β, giving one dot on the curves (continuous curves are a guide to the eye).

We recall that the design rate of a spatially-coupled ensemble is slightly higher than the rate R of its corresponding underlying ensemble due to the boundary nodes, i.e. $R_{\text{cou}} = R(1 + O(\frac{w-1}{L}))$. The difference between the design rates of both ensembles vanishes as $L \to \infty$. In order to disregard this finite size effect, we reported the algorithmic distortion of the coupled ensembles with respect to the rate R of their corresponding underlying ensembles.

The plots on the right are for the underlying ensembles with $l = 3, 4, 5$ and $n = 128000$. We observe that as the check degree increases the BPGD performance gets worse. But recall from Table I that with increasing degrees the optimal distortion of the ensemble (not shown explicitly on the plots) gets better and approaches the Shannon limit. Thus the situation is similar to the case of LDPC codes where the BP threshold gets worse with increasing degrees, while the MAP threshold approaches Shannon capacity.

The plots on the left show the algorithmic performance for the coupled ensembles with $l = 3, 4, 5$, $n = 2000$, $w = 3$, and $L = 64$ (so again a total length of $N = 128000$). We see that the BPGD performance approaches the Shannon limit as the degrees increase. One obtains a good performance, for a range of rates, without any optimization on the degree sequence of the ensemble, and with simple BPGD schemes.

The simulations, suggest the following. Look at the regime $n \gg L \gg w \gg 1$. When these parameters go to infinity in the specified order *for the coupled ensemble* $D_{\text{BPGD}}(R)$ approaches $D_{\text{opt}}(R)$. In words, the algorithmic distortion approaches the optimal distortion of the ensemble. When furthermore $l \to +\infty$ after the other parameters $D_{\text{BPGD}}(R)$ approaches $D_{\text{sh}}(R)$. At this point it is not possible to assess from the simulations whether these limits are exactly attained.

B. The Choice of the Parameter β

We discuss the empirical observations for the dependence of the curves $D_{\text{BPGD}}(\beta, R)$ on β at fixed rate. We illustrate our results for $R = 1/2$ and with the underlying LDGM($l = 5, R = 0.5, N = 128000$) and coupled LDGM($l = 5, R = 0.5, w = 3, L = 64, n = 2000$) ensembles.
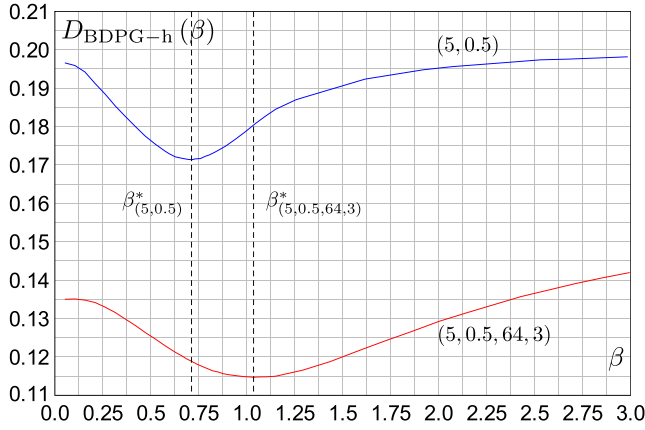
Fig. 5. The BPGD-h algorithmic distortion versus β. Results are obtained for coupled LDGM $(5, 0.5, L = 64, w = 3, n = 2000)$ and LDGM $(5, 0.5, 128000)$ ensemble. Results are averaged over 50 instances. The minimum distortion occurs at $\beta^*_{(5,0.5,64,3)} \approx 1.03 \pm 0.01$ and $\beta^*_{(5,0.5)} \approx 0.71 \pm 0.01$.
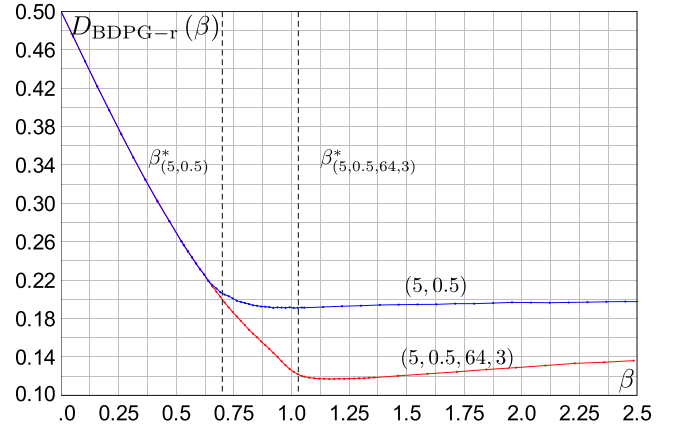

Fig. 6. The BPGD-r algorithmic distortion versus β. Results are obtained for coupled LDGM $(5, 0.5, L = 64, w = 3, n = 2000)$ and LDGM $(5, 0.5, 128000)$ ensemble. Results are averaged over 50 instances. The values β^* of Figure 5 are reported for comparison.
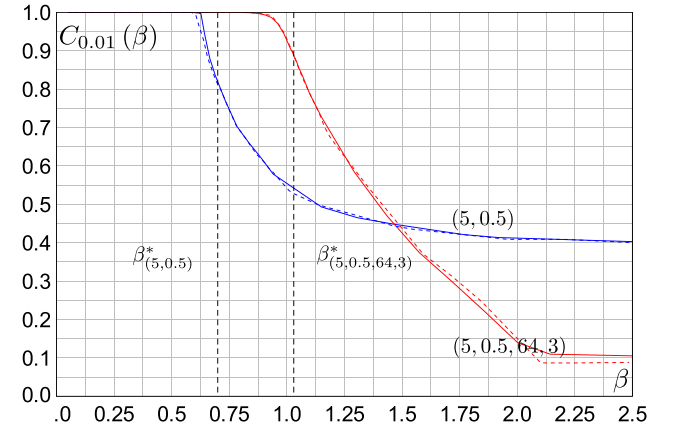

Fig. 7. $C_{0.01}(\beta)$ versus β. Empirical convergence probability for underlying LDGM $(5, 0.5, 128000)$ and coupled LDGM $(5, 0.5, L = 64, w = 3, n = 2000)$ ensembles. Solid (resp. dashed) lines are for the hard (resp. random) decision rule. Results are averaged over 50 instances.

On Fig. 5 we plot the distortion $D_{BPGD-h}(\beta, R = 1/2)$ of the hard decision rule. For *all* values of $0 < \beta < 3$, the algorithmic distortion $D_{BPGD-h}(\beta, R)$ of the coupled ensemble is below the corresponding curve of the underlying ensemble. The most important feature is a clear minimum at a value β^* which is rate dependent. The rate distortion curve for the hard decision rule on Figure 3 is computed at this β^* and is the result of the optimization

$$D_{BPGD-h}(R) = \min_{\beta > 0} D_{BPGD-h}(\beta, R). \tag{28}$$

We observe that the optimal value β^*_{cou} for the coupled ensemble is always larger than β^*_{un} for the underlying ensemble.

An additional observation is the following. As the degree l increases β^*_{un} tends to zero, and β^*_{cou} approaches $\beta_{sh}(R)$ where

$$\beta_{sh}(R) \equiv \frac{1}{2} \ln \left(\frac{1 - D_{sh}(R)}{D_{sh}(R)} \right). \tag{29}$$

This is the information theoretic value corresponding to the half-loglikelihood parameter of a test-BSC with the noise tuned at capacity. This observation is interesting because it shows that for large l, with the coupled ensemble, one does not really need to optimize over β, but it suffices to fix $\beta = \beta_{sh}(R)$. Theoretical motivation for this choice is discussed in Section V.

On Figure 6 we plot the curve $D_{BPGD-r}(\beta, R = 1/2)$ for the randomized algorithm. The behavior of the underlying and coupled ensemble have the same flavor. The curves are first decreasing with respect to β and then flatten. The minimum is reached in the flattened region and as long as β is chosen in the flat region, the optimized distortion is not very sensitive to this choice. We take advantage of this feature, and compute the rate distortion curve of the randomized decision rule at a predetermined value of β. This has the advantage of avoiding optimizing over β. Again, for the coupled case a good choice is to take $\beta_{sh}(R)$ given by Equ. 29. With these considerations the distortion curve on Figure 4 is

$$D_{BPGD-r}(R) = D_{BPGD-r}(\beta_{sh}, R). \tag{30}$$

C. Convergence

We have tested the convergence of the BPGD algorithms for both decision rules. We compute an *empirical probability of convergence* $C_{\epsilon,T}(\beta)$ defined as the fraction of decimation rounds that results from the convergence condition (23). In other words $C_{\epsilon,T}(\beta) = 1$ means that at every round of the decimation process the BP update rules converge in less than T iterations to a fixed point of the BP equations (20) up to a precision ϵ. Figure 7 shows $C_{\epsilon,T}(\beta)$ at $(\epsilon, T) = (0.01, 10)$ for the underlying and coupled ensembles. The hard decision rule is represented by solid lines and the random decision rule by dashed lines. The first observation is that both decision rules have identical behaviors. This is not a priori obvious since the decimation rules are different, and as a result the graph evolves differently for each rule during the decimation process. This suggest that the convergence of the algorithms essentially depends on the convergence of the plain BP algorithm. The second observation is that the values of β where $C_{\epsilon,T}(\beta)$ drops below one are roughly comparable to the values where D_{BPGD-r} flattens and where D_{BPGD-h} attains its minimum.

V. THE PHASE DIAGRAM: PREDICTIONS OF THE CAVITY METHOD

It is natural to expect that the behavior of belief propagation based algorithms should be in a way or another related to the phase diagram of the Gibbs distribution (8). The phase diagram can be derived by using the cavity method. As this is pretty involved, in the present section we provide a high level picture. The cavity equations are presented in section VI. We give a primer on the cavity method in appendix A and the technical derivations for the present problem are given in appendices B, C.

As we vary β the nature of the Gibbs measure and the geometry of the space of its typical configurations changes at special *dynamical* and *condensation* thresholds β_d and β_c. In paragraph V-A we explain what these thresholds are and what is their significance. We discuss how they are affected by spatial coupling in paragraph V-B. Finally in paragraph V-E we discuss some heuristic insights that allow to understand why Shannon's limit is approached with the BPGD algorithm for coupled ensembles with large check degrees.

In this section f and u denote the limits $\lim_{N\to+\infty} f_N$ and $\lim_{N\to+\infty} u_N$.

A. Dynamical and Condensation Thresholds

The cavity method assumes that the random Gibbs distribution (8) can, in the limit of $N \to +\infty$, be decomposed into a convex superposition of "extremal measures"

$$\mu_\beta(\underline{u} \mid \underline{x}) = \sum_{p=1}^{\mathcal{N}} w_p \, \mu_{\beta,p}(\underline{u} \mid \underline{x}) \tag{31}$$

each of which occurs with a weight $w_p = e^{-\beta N(f_p - f)}$, where f_p is a free energy associated to the extremal measure $\mu_{\beta,p}$. Since the weights w_p have to sum to 1, we have

$$e^{-\beta Nf} \approx \sum_{p=1}^{\mathcal{N}} e^{-\beta Nf_p} \approx e^{-\beta N \min_\varphi(\varphi - \beta^{-1}\Sigma(\varphi;\beta))} \tag{32}$$

where $e^{N\Sigma(\varphi;\beta)}$ counts the number of extremal states $\mu_{\beta,p}$ with free energy $f_p \approx \varphi$.

Such convex decompositions of the Gibbs distribution into bona fide extremal measures are under mathematical control for "simple" models such as the (deterministic) Ising model on a square grid [39]. But for spin glass models is it not known how to construct or even precisely define the extremal measures. One important conceptual difference with respect to the Ising model, which has a small number of extremal states, is that for spin glasses one envisions the possibility of having an exponentially large in N number of terms in the decomposition (31).

In the context of sparse graph models it is further assumed that there are "extremal" Bethe measures which are a good proxy for the "extremal measures". The Bethe measures are those measures that have marginals given by BP marginals. When the BP equations have many fixed point solutions there are many possible Bethe measures and one must have a criterion to choose among them. This is provided by the
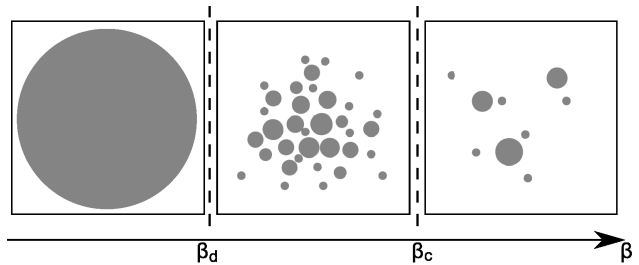

Fig. 8. Pictorial representation of the decomposition of the Gibbs distribution into a convex superposition of extremal states. Balls represent extremal states (their size represents their internal entropy). For $\beta < \beta_d$ there is one extremal state. For $\beta_d < \beta < \beta_c$ there are exponentially many extremal states (with the same internal free energy φ_{int}) that dominate to the convex superposition. For $\beta > \beta_c$ there is a finite number of extremal states that dominate the convex superposition.

Bethe free energy. The Bethe free energy is the functional whose stationary point equations (gradient equal zero) yield the BP equations. Heuristically, the extremal Bethe measures correspond to the solutions of the BP equations that are minima of the Bethe free energy.[4] Similarly, it is assumed that the Bethe free energies corresponding to solutions of the BP equations are good proxy's for the free energies f_p. Moreover one expects that the later concentrate.

Once one chooses to replace f_p by the Bethe free energies, the counting function $\Sigma(\varphi;\beta)$ and the free energy f can be computed through a fairly technical procedure, and a number of remarkable predictions about the decomposition (31) emerge.

The cavity method predicts the existence of two sharply defined thresholds β_d and β_c at which the nature of the convex decomposition (31) changes drastically. Figure 8 gives a pictorial view of the transitions associated with the decomposition (31). For $\beta < \beta_d$ the measure $\mu_\beta(\underline{u} \mid \underline{x})$ is extremal, in the sense that $\mathcal{N} = 1$ in (31). For $\beta_d < \beta < \beta_c$ the measure is a convex superposition of an exponentially large number of extremal states. The exponent $\varphi - \beta^{-1}\Sigma(\varphi;\beta)$ in (32) is minimized at a value $\varphi_{\text{int}}(\beta)$ such that $\Sigma(\varphi_{\text{int}}(\beta);\beta) > 0$. Then

$$\Sigma(\beta) \equiv \Sigma(\varphi_{\text{int}}(\beta);\beta) = \beta(\varphi_{\text{int}}(\beta) - f(\beta)) \tag{33}$$

is strictly positive and gives the growth rate (as $N \to +\infty$) of the number of extremal states that dominate the convex superposition of pure states (31). This quantity is called the complexity. It turns out that the complexity is a decreasing function of β which becomes negative at β_c where it looses its meaning. To summarize, above β_d and below β_c an exponentially large number of extremal states with the same free energy φ_{int} contribute significantly to the Gibbs distribution. For $\beta > \beta_c$ the number of extremal states that dominate the measure is finite. One says that the measure is *condensed* over a small number of extremal states. In fact, there may still be an exponential number of extremal states but they do not contribute significantly to the measure because their weight is exponentially smaller than the dominant ones.

---

[4]Remarkably, it is not very important to be able to precisely select these minima because at low temperatures one expects that they outnumber the other ones.

TABLE II

THE NUMERICAL VALUES OF β_d AND β_c FOR COUPLED POISSON
LDGM (l, $R = 0.5$, L, $w = 3$) ENSEMBLES WITH $l = 3, 4$,
AND 5 AND DIFFERENT VALUES OF L. THE RESULTS
ARE OBTAINED BY POPULATION
DYNAMICS (SEE SECT. VII)

l	β	uncoupled	L		
		coupled	32	64	128
3	β_d	0.883	0.942	0.941	0.941
	β_c	0.940	0.958	0.948	0.946
4	β_d	0.875	1.010	1.010	1.009
	β_c	1.010	1.038	1.023	1.017
5	β_d	0.832	1.032	1.030	1.029
	β_c	1.032	1.067	1.048	1.039

TABLE III

THE NUMERICAL VALUES OF β_d AND β_c FOR COUPLED POISSON
LDGM (5, $R = 0.5$, L, w) ENSEMBLES WITH DIFFERENT
VALUES OF L AND w. THE RESULTS ARE OBTAINED BY
POPULATION DYNAMICS (SEE SECT. VII)

L	β	w		
		2	3	4
128	β_d	1.028	1.029	1.030
	β_c	1.038	1.039	1.043
256	β_d	1.023	1.027	1.029
	β_c	1.035	1.037	1.038

There exist a mathematically more precise definition of β_d and β_c in terms of correlation functions. When these correlation functions are computed within the framework of the cavity method the results for β_d and β_c agree with those given by the complexity curve $\Sigma(\beta)$. Although these definitions nicely complete the perspective, we refrain from giving them here since we will not use them explicitly.

What is the significance of the transitions at β_d and β_c? The condensation threshold is a thermodynamic phase transition point: the free energy $f(\beta)$ and internal energy $u(\beta)$ are not analytic at β_c. At β_d the free and internal energies have no singularities: in particular their *analytical expressions* do not change in the whole range $0 < \beta < \beta_c$. At β_d the (phase) transition is dynamical: Markov chain Monte Carlo algorithms have an equilibration time that diverges when $\beta \uparrow \beta_d$, and are unable to sample the Gibbs distribution for $\beta > \beta_d$. For more details we refer to [30].

B. Complexity and Thresholds of the Underlying and Coupled Ensembles

We have computed the complexity and the thresholds from the cavity theory. These have been computed both from the full cavity equations of Section VI-A and from the simplified ones of Section VI-C. Tables II and III illustrate the results.

Since the free energies of the coupled and underlying ensembles are the same in the limit of infinite length (known from theorem 1) and the condensation threshold is a singularity of the free energy (known from the cavity method), we can conclude on theoretical grounds that

$$\lim_{L \to +\infty} \beta_c(L, w) = \beta_c(w = 1). \qquad (34)$$

Table II shows that the condensation threshold $\beta_c(L, w)$ of the coupled ensemble is higher than $\beta_c(w = 1)$ and decreases as L increases. The finite size effects are still clearly visible at lengths $L = 128$ and are more marked for larger w. This is not surprising since we expect the finite size corrections to be of order $O(w/L)$.

Let us now discuss the behavior of the dynamical threshold. Table III displays the results for the ensembles LDGM($l = 5$, $R = 0.5$) and LDGM($l = 5$, $R = 0.5$, L, w).

The column $w = 1$ gives the dynamical and condensation thresholds of the underlying ensemble, $\beta_d(w = 1)$ and $\beta_c(w = 1)$. We see that for each fixed L the dynamical threshold increases as a function of w. Closer inspection suggests that

$$\lim_{w \to +\infty} \lim_{L \to +\infty} \beta_d(L, w) = \beta_c(w = 1). \qquad (35)$$

Equ. 35 indicates a threshold saturation phenomenon: for the coupled ensemble the phase of non-zero complexity shrinks to zero and the condensation point remains unchanged. This is analogous to the saturation of the BP threshold of LDPC codes towards the MAP threshold [19]. It is also analogous to the saturation of spinodal points in the Curie-Weiss chain [23]. Similar observations have been discussed for constraint satisfaction problems in [24].

C. Comparison of β^* With β_d

We systematically observe that the optimal algorithmic value β^* of the BPGD-h algorithm is always lower, but somewhat close to β_d. For example for the uncoupled case $l = 5$ we have $(\beta^*, \beta_d) \approx (0.71, 0.832)$. For the coupled ensembles with $(L = 64, w = 3)$ we have $(\beta^*, \beta_d) \approx (1.03, 1.038)$. In fact, in the coupled case we observe $\beta^* \approx \beta_d \approx \beta_c$. Thus for the coupled ensemble BPGD-h operates well even close to the condensation threshold.

This is also the case for BPGD-r as we explain in the next paragraph. We use this fact in the next section to explain the good performance of the algorithm for coupled instances.

D. Sampling of the Gibbs Distribution With BPGD-r

Threshold saturation, equation (35), indicates that for L large, the phase of non-zero complexity, occupies a very small portion of the phase diagram close to β_c. This then suggests that for coupled ensembles Markov chain Monte Carlo dynamics, and BPGD-r algorithms are able to correctly sample the Gibbs measure for values of β up to $\approx \beta_c$. Let us discus in more detail this aspect of the BPGD-r algorithm.

By the Bayes rule:

$$\mu_\beta(\underline{u} \mid \underline{x}) = \prod_{i=1}^{m} \mu_\beta(u_i | \underline{x}, u_1, \ldots, u_{i-1}). \qquad (36)$$

Thus we can sample $\underline{u}$ by first sampling u_1 from $\mu_\beta(u_1|\underline{x})$, then u_2 from $\mu_\beta(u_2|\underline{x}, u_1)$ and so on. Then, computing $x_a = \oplus_{i \in \partial a} u_i$ and the resulting average distortion, yields half the internal energy $u(\beta)/2$. With the BPGD-r algorithm the average distortion is computed in the same way except that the sampling is done with the BP marginals. So as long
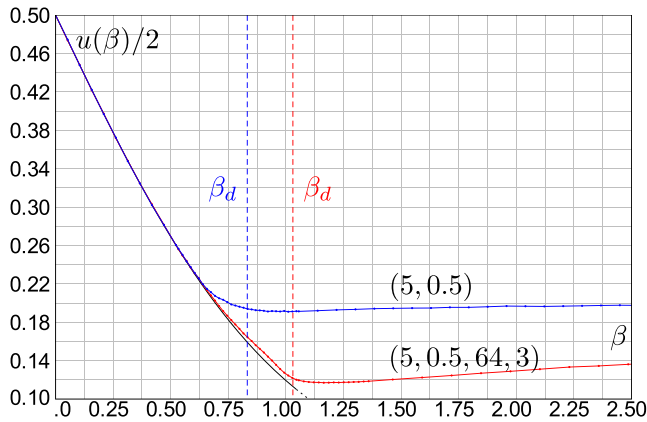
Fig. 9. The performance of the BPGD-r algorithm. The plot shows that the algorithm can approximate average distortion quite precisely for $\beta < \beta' \approx \beta_d$. The black curve shows the average distortion $u(\beta)/2 = (1 - \tanh \beta)/2$ for $\beta < \beta_c$. The results are obtained for the the underlying LDGM $(5, 0.5, 128000)$ and coupled LDGM $(5, 0.5, 64, 3, 2000)$ ensembles. The results are averaged over 50 instances. Numerical values of various thresholds are $\beta_{d,un} = 0.832$, $\beta_{d,cou} = 1.030$, $\beta_c = 1.032$.

as the BP marginals are a good approximation of the true marginals, the average distortion $D_{BPGD-r}(\beta)$ should be close to $u(\beta)/2$. This can be conveniently tested because the cavity method predicts the simple formula[5] $u(\beta)/2 = (1-\tanh\beta)/2$ for $\beta < \beta_c$.

On Fig. 9 we observe $D_{BPGD-r}(\beta) \approx (1 - \tanh\beta)/2$ for $\beta < \beta'$, with a value of β' lower but comparable to β_d. In particular for a coupled ensemble we observe $\beta' \approx \beta_d \approx \beta_c$. So Fig. 9 strongly suggests that BPGD-r correctly samples the Gibbs distribution of coupled instances all the way up to $\approx \beta_c$, and that BP correctly computes marginals for the same range.

E. Large Degree Limit

According to the information theoretic approach to rate-distortion theory, we can view the encoding problem, as a decoding problem for a random linear code on a test-BSC(p) test-channel with noise $p = D_{sh}(R)$. Now, the Gibbs distribution (8) with $\beta = \frac{1}{2}\ln(1 - p)/p$ is a MAP-decoder measure for a channel problem with the noise tuned to the Shannon limit. Moreover, for large degrees the LDGM ensemble is expected to be equivalent to the random linear code ensemble. These two remarks suggest that, since in the case of coupled ensembles with large degrees the BPGD-h encoder with optimal β^*, approaches the rate-distortion limit, we should have

$$\beta^* \approx \frac{1}{2}\ln\frac{1-p}{p} \equiv \frac{1}{2}\ln\frac{1-D_{sh}(R)}{D_{sh}(R)}. \quad (37)$$

In fact this is true. Indeed on the one hand, as explained above, for coupled codes we find $\beta^* \approx \beta_d \approx \beta_c$ (even for finite degrees). On the other hand an analytical *large degree* analysis of the cavity equations in section VI-D allows to compute the complexity and to show the remarkable relation

$$\beta_c \approx \frac{1}{2}\ln\frac{1-D_{sh}(R)}{D_{sh}(R)}, \quad \text{for } l >> 1. \quad (38)$$

---

[5]For $\beta > \beta_c$ the formula is different. Indeed, β_c is a static phase transition point.

These remarks also show that the rate-distortion curve can be interpreted as a line of condensation thresholds for each R.

VI. CAVITY EQUATIONS FOR LDGM COUPLED ENSEMBLES

We display the set of fixed point equations needed to compute the complexity (33) of the coupled ensemble. To get the equations for the underlying ensembles one sets $w = 1$ and drops the positional z dependence in all quantities.

In order to derive the fixed point equations one first writes down the cavity equations for a single instance of the graph and source word. These involve a set of messages on the edges of the graph. These messages are *random probability distributions*. If one assumes independence of messages flowing into a node, it is possible to write down a set of integral fixed point equations - the cavity equations - for the *probability distributions of the messages*. It turns out that the cavity equations are much harder to solve numerically than usual density evolution equations because of "reweighting factors". Fortunately for $\beta < \beta_c$ it is possible to eliminate the reweighting factor, thus obtaining a much simpler set of six integral fixed point equations. This whole derivation is quite complicated and for the benefit of the reader, we choose to present it three stages in appendices A, B and C. The calculations are adapted from the methods of [40] for the K-SAT problem in the SAT phase.

Paragraphs VI-A and VI-B give the set of six integral fixed point equations and the complexity (derived in appendices A, B and C).

We will see that in the present problem for $\beta < \beta_c$, not only one can eliminate the reweighting factors, but there is a further simplification of the cavity equations. With this extra simplification the cavity equations reduce to standard density evolution equations associated to a coupled LDGM code over a test-BSC-channel. This is explained in paragraph VI-C.

A. Fixed Point Equations of the Cavity Method for $\beta \le \beta_c$

Our fixed point equations involve six distributions $q_z(h)$, $\widehat{q}_z(\widehat{h})$, $q_z^\sigma(\eta|h)$ and $\widehat{q}_z^\sigma(\widehat{\eta}|\widehat{h})$ with $\sigma = \pm 1$. The subscript z indicates that the distributions are position dependent, $z = 1, \ldots, L + w - 1$. A hat (resp. no hat) indicates that this is the distribution associated to messages that emanate from a check node (resp. code-bit node). All messages emanating from a node have the *same* distribution. Thus the distributions depend only on the position of the node and not on the direction of the edge.

It is convenient to define two functions g and $\widehat{g}$ (see the BP equations (20))

$$\begin{cases} g(\widehat{h}_1, \ldots \widehat{h}_{r-1}) = \sum_{i=1}^{r-1} \widehat{h}_i \\ \widehat{g}(h_1, \ldots h_{l-1} \mid J) = J\beta^{-1}\tanh^{-1}\left(\tanh\beta \prod_{i=1}^{l-1}\tanh\beta h_i\right) \end{cases}$$

where $J \equiv (-1)^x$ is the random variable representing the source bits. Furthermore we set $P(r) = \frac{(l/R)^r}{r!}e^{-l/R}$ for the Poisson degree distribution of code-bit nodes.

Distributions $q_z(h)$, $\widehat{q}_z(\widehat{h})$ satisfy a set of closed equations[6]

$$q_z(h) = \sum_{r=0}^{\infty} \frac{P(r)}{w^r} \sum_{y_1,\ldots y_k=0}^{w-1} \int \prod_{a=1}^{r} d\widehat{h}_a \widehat{q}_{z-y_a}(\widehat{h}_a)$$
$$\times \delta(h - g(\widehat{h}_1, \ldots, \widehat{h}_r)) \tag{39}$$

and

$$\widehat{q}_z(\widehat{h}) = \frac{1}{w^{l-1}} \sum_{y_1,\ldots,y_{l-1}=0}^{w-1} \int \prod_{i=1}^{l-1} dh_i q_{z+y_i}(h_i)$$
$$\times \frac{1}{2} \sum_{J=\pm 1} \delta(\widehat{h} - \widehat{g}(h_1, \ldots, h_{l-1} \mid J)). \tag{40}$$

Let $\sigma_i = \pm 1$ denote auxiliary "spin" variables. We introduce the conditional measure over $\sigma_1, \ldots, \sigma_{l-1}$,

$$v_1(\sigma_1, \ldots, \sigma_{l-1} \mid J\sigma, h_1, \ldots, h_{l-1})$$
$$= \frac{1 + J\sigma \tanh \beta \prod_{i=1}^{l-1} \sigma_i}{1 + J\sigma \tanh \beta \prod_{i=1}^{l-1} \tanh \beta h_i} \prod_{i=1}^{l-1} \frac{1 + \sigma_i \tanh \beta h_i}{2}. \tag{41}$$

The equations for distributions $q_z^\sigma(\eta \mid h)$ and $\widehat{q}_z^\sigma(\widehat{\eta} \mid \widehat{h})$ are

$$q_z^\sigma(\eta \mid h) q_z(h) = \sum_{r=0}^{\infty} \frac{P(r)}{w^r} \sum_{y_1,\ldots,y_r=0}^{w-1} \times \int \prod_{a=1}^{r} d\widehat{h}_a \widehat{q}_{z-y_a}(\widehat{h}_a)$$
$$\times \delta(h - g(\widehat{h}_1, \ldots, \widehat{h}_r))$$
$$\times \int \prod_{a=1}^{r} d\widehat{\eta}_a \widehat{q}_{z-y_a}^\sigma(\widehat{\eta}_a \mid \widehat{h}_a) \times \delta(\eta - g(\widehat{\eta}_1, \ldots, \widehat{\eta}_r)) \tag{42}$$

and

$$\widehat{q}_z^\sigma(\widehat{\eta} \mid \widehat{h}) \widehat{q}_z(\widehat{h}) = \frac{1}{w^{l-1}} \sum_{y_1,\ldots,y_{l-1}=0}^{w-1} \int \prod_{i=1}^{l-1} dh_i q_{z+y_i}(h_i)$$
$$\times \frac{1}{2} \sum_{J=\pm 1} \sum_{\sigma_1,\ldots,\sigma_{l-1}=\pm 1} v_1(\sigma_1, \ldots, \sigma_{l-1} \mid J\sigma, h_1, \ldots, h_{l-1})$$
$$\times \delta(\widehat{h} - \widehat{g}(h_1, \ldots, h_{l-1} \mid J))$$
$$\times \int \prod_{i=1}^{l-1} d\eta_i q_{z+y_i}^{\sigma_i}(\eta_i \mid h_i) \delta(\widehat{\eta} - \widehat{g}(\eta_1, \ldots, \eta_{l-1} \mid J)). \tag{43}$$

Equations (39), (40), (42), (43) constitutes a closed set of fixed point equations for six probability distributions.

Let us pause for a moment to give some information on these distributions and an interpretation of the equations that relate them.

When there is a proliferation of BP fixed points, usual density evolution does not track correctly the average behavior of the BP messages. In the formalism of the cavity method (see Appendix A) one introduces new messages called *cavity messages* which are random valued distributions over the space of BP fixed points (for a fixed instance). They satisfy "cavity message passing equations" (see equ. (82)). The fixed point equations presented here (39), (40), (42), (43), describe the behavior of the "distributions" of these cavity messages. More precisely the *averages of the cavity messages* - themselves random quantities - satisfy message passing BP equations

(see equ. (97)). The quantities $q_z(h)$ and $\widehat{q}_z(\widehat{h})$ are the distributions of the *averages of the cavity messages* (see equ. (98)) and therefore satisfy the "usual" density evolution equations. The quantities $q_z^\sigma(\eta \mid h)$ and $\widehat{q}_z^\sigma(\widehat{\eta} \mid \widehat{h})$ are conditionnal averages of the random cavity messages (see equ., (100)). The conditioning corresponds to fix the average of the cavity message.

The equations (42) and (43) have an interesting interpretation as a reconstruction problem on a tree (see [41] where the case of coloring is treated in detail and a brief discussion of more general models is presented). Consider a rooted tree of depth $t > 0$ created at random from a stochastic branching process where variable nodes have $r - 1$ descendants with probability $P(r)$ (except for the root node which has r descendants) and check nodes have $l - 1$ descendants. Each check node "broadcasts" the variable σ that is immediately above it, to its $l - 1$ descendants which receive the vector $(\sigma_1, \cdots, \sigma_{l-1})$ with probability $v_1(\sigma_1, \ldots, \sigma_{l-1} \mid J\sigma, h_1, \ldots, h_{l-1})$. This broadcasting process induces a probability distribution on the configurations of the variables at the leaf nodes of the tree. The aim of the reconstruction problem is to infer the value of the root node given the configuration at the leafs at depth t. The analysis of the reconstruction problem on a tree suggests that the equations (42) and (43) possess non-trivial fixed points if and only if the iterations of these equations with the initial condition[7]

$$q_z^{\sigma_i}(\eta \mid h) = \delta_{+\infty}(\sigma_i \eta_i), \tag{44}$$

converges to a non-trivial fixed point. This has the advantage of removing the ambiguity of the initial conditions in order to solve iteratively the fix point equations for $q_z^\sigma(\eta \mid h)$ and $\widehat{q}_z^\sigma(\widehat{\eta} \mid \widehat{h})$.

B. Complexity in Terms of Fixed Point Densities

Let

$$\begin{cases} Z_1(h_1, \ldots, h_l \mid J) = 1 + J(\tanh \beta) \prod_{i=1}^{l} \tanh \beta h_i \\ Z_2(\widehat{h}_1, \ldots, \widehat{h}_r) = \frac{1}{2} \sum_{\sigma=\pm 1} \prod_{i=1}^{r} (1 + \sigma \tanh \beta \widehat{h}_i). \end{cases}$$

We are now ready to give the expression for the complexity in terms of the densities $q_z(h)$, $\widehat{q}_z(\widehat{h})$, $q_z^\sigma(\eta \mid h)$ and $\widehat{q}_z^\sigma(\widehat{\eta} \mid \widehat{h})$. Recall formula (33) which expresses the complexity as $\Sigma(\beta) = \beta(\varphi_{\text{int}}(\beta) - f(\beta))$. In the formulas below it is understood that $n \to +\infty$.

The expression of f is the simplest

$$-\beta f = \ln(1 + e^{-2\beta}) + (R-1) \ln 2$$
$$- \frac{l-1}{L} \sum_{z=1}^{L} \frac{1}{w^l} \sum_{y_1,\ldots,y_l=0}^{w-1} \int \prod_{i=1}^{l} dh_i q_{z+y_i}(h_i)$$
$$\times \frac{1}{2} \sum_{J=\pm 1} \ln Z_1(h_1, \ldots, h_l \mid J)$$
$$+ \frac{R}{L+w-1} \sum_{z=1}^{L+w-1} \sum_{r=0}^{\infty} \frac{P(r)}{w^r}$$
$$\times \sum_{y_1,\ldots,y_r=0}^{w-1} \int \prod_{a=1}^{r} d\widehat{h}_a \widehat{q}_{z-y_a}(\widehat{h}_a) \ln Z_2(\widehat{h}_1, \ldots, \widehat{h}_r).$$

$$\tag{45}$$

---

[6]We use the convention that if z is out of range the corresponding distribution is a unit mass at zero.

[7]Here we adopt the notation $\delta_{+\infty}$ for a unit mass distribution at infinity.

To express φ_{int} we first need to define the conditional measure over $\sigma = \pm 1$

$$\nu_2(\sigma | \widehat{h}_1, \ldots, \widehat{h}_k)$$
$$= \frac{\prod_{a=1}^{k}(1 + \sigma \tanh \beta \widehat{h}_a)}{\prod_{a=1}^{k}(1 + \tanh \beta \widehat{h}_a) + \prod_{a=1}^{k}(1 - \tanh \beta \widehat{h}_a)}.$$

We have

$$-\beta \varphi_{\text{int}} = \ln(1 + e^{-2\beta}) + (R - 1) \ln 2$$
$$- \frac{l-1}{L} \sum_{z=1}^{L} \frac{1}{w^l} \sum_{y_1,\ldots,y_l=0}^{w-1} \int \prod_{i=1}^{l} dh_i q_{z+y_i}(h_i)$$
$$\times \frac{1}{2} \sum_{J=\pm 1} \sum_{\sigma_1,\ldots,\sigma_l=\pm 1} \nu_1(\sigma_1, \ldots, \sigma_l | J, h_1, \ldots, h_l)$$
$$\times \int \prod_{i=1}^{l} d\eta_i q_{z+y_i}^{\sigma_i}(\eta_i | h_i) \ln Z_1(\eta_1, \ldots, \eta_l \mid J)$$
$$+ \frac{R}{L+w-1} \sum_{z=1}^{L+w-1} \sum_{r=0}^{\infty} \frac{P(r)}{w^r}$$
$$\times \sum_{y_1,\ldots,y_r=0}^{w-1} \int \prod_{a=1}^{r} d\widehat{h}_a \widehat{q}_{z-y_a}(\widehat{h}_a) \sum_{\sigma} \nu_2(\sigma | \widehat{h}_1, \ldots, \widehat{h}_r)$$
$$\times \int \prod_{a=1}^{r} d\widehat{\eta}_a \widehat{q}_{z-y_a}^{\sigma}(\widehat{\eta}_a | \widehat{h}_a) \ln Z_2(\widehat{\eta}_1, \ldots, \widehat{\eta}_r). \quad (46)$$

Thanks to (45), (46) the complexity $\Sigma(\beta; L, w)$ of the coupled ensemble is computed, one reads off the dynamical and condensation thresholds $\beta_d(L, w)$ and $\beta_c(L, w)$. The corresponding quantities for the underlying ensemble are obtained by setting $L = w = 1$.

C. Further Simplications of Fixed Point Equations and Complexity

It is immediate to check that $q_z(h) = \delta(h)$ and $\widehat{q}_z(\widehat{h}) = \delta(\widehat{h})$ is a trivial fixed point of (39), (40). When we solve these equations by population dynamics with a uniform initial condition over $[-1, +1]$ for $\widehat{h}$, we find that for fixed degrees and β fixed in a finite range depending on the degrees, the updates converge towards the trivial fixed point. Up to numerical precision, the values of $h, \widehat{h}$ are concentrated on 0. It turns out that the range of β for which this is valid is wider than the interval $[0, \beta_c]$. At first sight this may seem paradoxical, and one would have expected that this range of β is equal to $[0, \beta_c]$. In fact, one must recall that beyond β_c the equations of paragraph VI-A are not valid (see Appendix A), so there is *no* paradox. Theorem 2 in section VIII shows that, for a wide class of initial conditions and given β, for large enough degree l the iterative solution of (39), (40) tends to the trivial point. This theorem, together with the numerical evidence, provides a strong justification for the following simplification.

We assume that for $\beta < \beta_c$, equations (39), (40) have a unique solution $q_z(h) = \delta(h)$ and $\widehat{q}_z(\widehat{h}) = \delta(\widehat{h})$. Note that the initial condition (44) satisfies a symmetry $q^\sigma(\eta|0) = q^{-\sigma}(-\eta|0) = \delta_{+\infty}(\eta\sigma)$ (even for $h \neq 0$). Now for $h = \widehat{h} = 0$ the iterations of (42) and (43) preserve this symmetry. In other

words the solutions of these equations (for $h = \widehat{h} = 0$) found from a symmetric initial condition satisfy $q_z^{\sigma=1}(\eta|0) = q_z^{\sigma=-1}(-\eta|0)$, $\widehat{q}_z^{\sigma=1}(\widehat{\eta}|0) = \widehat{q}_z^{\sigma=-1}(-\widehat{\eta}|0)$.

Therefore we look only for symmetrical solutions, and set

$$q_z^{+}(\eta) = q_z^{\sigma=+1}(\eta|0), \quad \text{and} \quad \widehat{q}_z^{+}(\widehat{\eta}) = \widehat{q}_z^{\sigma=+1}(\widehat{\eta}|0)$$

Then the equations (42), (43) simplify drastically,

$$q_z^{+}(\eta) = \sum_{r=0}^{\infty} \frac{P(r)}{w^r} \sum_{y_1,\ldots,y_r=0}^{w-1}$$
$$\times \int \prod_{a=1}^{r} d\widehat{\eta}_a \widehat{q}_{z-y_a}^{+}(\widehat{\eta}_a) \delta(\eta - g(\widehat{\eta}_1, \ldots, \widehat{\eta}_r)) \quad (47)$$

$$\widehat{q}_z^{+}(\widehat{\eta}) = \frac{1}{w^{l-1}} \sum_{y_1,\ldots,y_{l-1}=0}^{w-1} \int \prod_{i=1}^{l-1} d\eta_i q_{z+y_i}^{+}(\eta_i)$$
$$\times \sum_{J=\pm 1} \frac{1 + J \tanh \beta}{2} \delta(\widehat{\eta} - \widehat{g}(\eta_1, \ldots \eta_{l-1} \mid J)). \quad (48)$$

Remarkably, these are the standard density evolution equations for an LDGM code over a test-BSC-channel with half-log-likelihood parameter equal to β.

The free energy (45) now takes a very simple form

$$-\beta f = \ln(1 + e^{-2\beta}) + (R - 1) \ln 2. \quad (49)$$

At this point let us note that this simple formula has been proven by the *interpolation method* [42], for small enough β. Since it is expected that there is no (static) thermodynamic phase transition for $\beta < \beta_c$, the free energy is expected to be analytic for $\beta < \beta_c$. Thus by analytic continuation, formula (49) should hold for all $\beta < \beta_c$. This also provides justification for the triviality assumption made above for the fixed point. Indeed, a non-trivial fixed point leading to the same free energy would entail miraculous cancellations.

When we compute the complexity, expression (49) cancels with the first line in φ_{int} (see equ. (46)). We find

$$\Sigma(\beta; L, w) = \frac{l-1}{L} \sum_{z=1}^{L} \frac{1}{w} \sum_{y=0}^{w-1} \Sigma_e[q_{z+y}^{+}, \widehat{q}_z^{+}]$$
$$- \frac{l}{L} \sum_{z=1}^{L} \Sigma_v[\widehat{q}_z^{+}] + \frac{R}{L+w-1} \sum_{z=1}^{L+w-1} \Sigma_v[q_z^{+}],$$

where

$$\Sigma_v[q^{+}] = \int d\eta \, q^{+}(\eta) \ln(1 + \tanh \beta \eta)$$
$$\Sigma_e[q^{+}, \widehat{q}^{+}] = \int d\eta d\widehat{\eta} \, q^{+}(\eta) \widehat{q}^{+}(\widehat{\eta}) \ln(1 + \tanh \beta \eta \tanh \beta \widehat{\eta}).$$

For the underlying ensemble ($L = w = 1$) the complexity reduces to

$$\Sigma(\beta) = (l - 1) \Sigma_e[q^{+}, \widehat{q}^{+}] - l \Sigma_v[\widehat{q}^{+}] + R \Sigma_v[q^{+}]. \quad (50)$$

The average distortion or internal energy (see (15), (16)) at temperature β is obtained by differentiating (49), which yields the simple formula $(1 - \tanh \beta)/2$. This is nothing else than the (bottom) curve plotted in Figure 9. It has to be noted that

this expression is only valid for $\beta < \beta_c$. To obtain the optimal distortion of the ensemble D_{opt} (see table I) one needs to recourse to the full cavity formulas in order to take the limit $\beta \to +\infty$.

D. Large Degree Limit

Inspection of the fixed point equations (47) and (48) shows that the distributions

$$q^+(\eta) = \delta_{+\infty}(\eta), \text{ and } \widehat{q}^+(\widehat{\eta}) = \sum_{J=\pm 1} \frac{1 + J \tanh \beta}{2} \delta(\widehat{\eta} - J) \tag{51}$$

are a fixed point solution for the underlying model ($w = 1$) in the limit $l \to +\infty$, R fixed. This is (partially) justified by theorem 3 in section VIII. The fixed point (51) leads to a complexity for the underlying model for $l \to +\infty$,

$$\lim_{l \to +\infty} \Sigma(\beta) = (R - 1) \ln 2$$
$$- \sum_{J=\pm 1} \frac{1 + J \tanh \beta}{2} \ln\left(\frac{1 + J \tanh \beta}{2}\right).$$

On this expression one can read the large degree limit of the dynamical and condensation thresholds for the underlying ensemble. In this limit the complexity is non-zero all the way up to $\beta = 0$ (infinite temperature) so one finds that $\lim_{l \to +\infty} \beta_d = 0$. The condensation threshold on the other hand, $\lim_{l \to +\infty} \beta_c$, is obtained by setting the complexity to zero

$$1 - R = \lim_{l \to +\infty} h_2\left(\frac{1 + \tanh \beta_c}{2}\right), \tag{52}$$

which is equivalent to

$$\lim_{l \to +\infty} \beta_c = \beta_{\text{sh}} \equiv \frac{1}{2} \ln\left(\frac{1 - D_{\text{sh}}(R)}{D_{\text{sh}}(R)}\right). \tag{53}$$

In the large degree limit the condensation threshold is equal to the half-log-likelihood of a BSC test-channel with probability of error $D_{\text{sh}}(R)$, i.e. tuned to capacity.

Notice that since the condensation thresholds for both the underlying and the spatially-coupled ensembles are equal, Equation (53) is also true for coupled ensembles. Moreover the average distortion or internal energy is given for both ensembles by

$$\frac{1}{2} u(\beta) = \begin{cases} \frac{1}{2}(1 - \tanh \beta) & \beta < \beta_{\text{sh}}(R) \\ D_{\text{sh}}(R) & \beta \geq \beta_{\text{sh}}(R) \end{cases} \tag{54}$$

The above equation is a consequence of the monotonicity of $u(\beta)$ and the saturation of the condensation threshold toward the Shannon threshold. We conclude this section with a proof of this fact.

Using (9), (10) and (13), it is not hard to show that the derivative with respect to β of the internal energy for finite size N has a sign opposite to the variance of the distortion

$$\frac{d}{d\beta} u_N(\beta) = -4N \mathbb{E}_{\text{LDGM}, \underline{X}}[\langle d_N(\underline{x}, \widehat{\underline{x}})\rangle^2 - \langle d_N(\underline{x}, \widehat{\underline{x}})^2\rangle]. \tag{55}$$

This proves that for every N the internal energy $u_N(\beta)$ is a non-increasing function with respect to $\beta \in [0, \infty[$. It also proves, thanks to Equation (16), that the free energy $-\beta f_N$ is a convex function with respect to $\beta \in [0, \infty]$. The cavity method predicts that in the thermodynamic limit $N \to +\infty$ the quantity βf_N converges to (49) for $\beta \leq \beta_c$. This prediction combined with the fact that $-\beta f_N$ is convex implies that the internal energy $u(\beta) = \lim_{N \to +\infty} u_N(\beta)$ converges[8] to

$$u(\beta) = \frac{d}{d\beta}(\beta f) = (1 - \tanh \beta), \tag{56}$$

for $\beta \leq \beta_c$. Thus in the limit of large degree the internal energy becomes equal to twice the Shannon distortion at the condensation threshold

$$\lim_{l \to +\infty} u(\beta_c) = u(\beta_{\text{sh}}) = 2D_{\text{sh}}(R). \tag{57}$$

But since $2D_{\text{sh}}(R)$ is a lower bound for $\liminf_{\beta \to +\infty} u(\beta)$ (thanks to the rate-distortion theorem) and $u(\beta)$ is a non-increasing function, we conclude that $u(\beta) \equiv 2D_{\text{sh}}(R)$ for $\beta \geq \beta_{\text{sh}}$.

VII. POPULATION DYNAMICS COMPUTATION OF THE COMPLEXITY

In this section, we describe the population dynamics solutions of the various fixed point equations.

Let us first discuss the solution of (39), (40), (42) and (43). To represent the densities $q_z(h)$, $q_z^\pm(\eta|h)$, $\widehat{q}_z(\widehat{h})$, and $\widehat{q}_z^\pm(\widehat{\eta}|\widehat{h})$ we use two populations: a code-bit population and a check population. The code-bit population is constituted of $L + w - 1$ sets labeled by $z \in [1, L + w - 1]$. Each set, say z, has a population of size n, constituted of triples: $(h_{(z,i)}, \eta_{(z,i)}^+, \eta_{(z,i)}^-)$, $1 \leq i \leq n$. The total size of the code-bit population is $(L + w - 1)n$. Similarly, we have a population of triples with size Ln for check nodes, i.e. $(\widehat{h}_{(z,a)}, \widehat{\eta}_{(z,a)}^+, \widehat{\eta}_{(z,a)}^-)$, $z = 1, \ldots, L, a = 1, \ldots, n$. As inputs, they require the population size n, the maximum number of iterations $t_{\max}$, and the specifications of the coupled LDGM ensemble l, r, L, w. First we solve the two equations (39) and (40) with Algorithm 2.

Then we solve (42) and (43) with the Algorithm[9] 3.

From the final populations obtained after $t_{\max}$ iterations it is easy to compute the complexity and the thresholds β_d, β_c.

It is much simpler to solve the simplified fixed point equations (47), (48). The population dynamics algorithm is almost the same than in Table 2. The only difference is that J is generated according to the p.d.f $(1 + J \tanh \beta)/2$ instead of Ber(1/2). The big advantage is that there is no need to generate the 2^{r-1} configurations $\sigma_1, \ldots, \sigma_{r-1}$ which reduces the complexity of each iteration.

As expected the complexity obtained in either way is the same up to numerical precision. Numerical values of the dynamical and condensation thresholds are presented in tables II and III. Results are obtained with population sizes

---

[8] See for instance [43, p. 203] to understand why convexity enables us to exchange the thermodynamical limit and the derivative.

[9] In the next to last line marked (*) the chosen index is not in a valid range. In an instance of a coupled ensemble, this happens at the boundary, in which the corresponding node has smaller degree. In the message passing equation we discard these indices or equivalently assume that their triples are $(0, 0, 0)$.

Algorithm 2 Population Dynamics for (39) and (40)

1 **for** $z = 1$ *to* $L + w - 1$ **do**
2   **for** $i = 1$ *to* n **do**
3   Draw $\widehat{h}_{(z,i)}$ uniformly from $[-1, +1]$;

4 **for** $t \in \{1, \ldots, t_{max}\}$ **do**
5   **for** $z = 1$ *to* $L + w - 1$ **do**
6   **for** $i = 1$ *to* n **do**
7   *Generate a new $h_{(z,i)}$*;
8   Choose $l - 1$ pair indices $a_1, \ldots, a_{l-1}$ uniformly from nw pairs (y, j), $y \in [z - w + 1, z]$ and $j \in \{1, \ldots, n\}$;
9   **if** *for some index k, $a_k = (y, j)$ and $y < 1$* **then**
10   Set $\widehat{h}_{a_k} = 0$;
11   Set $h_{(z,i)} = \sum_{k=1}^{l-1} \widehat{h}_{a_k}$;

12   **for** $z = 1$ *to* L **do**
13   **for** $a = 1$ *to* n **do**
14   *Generate J randomly and generate a new $\widehat{h}_{(z,a)}$*;
15   Choose $r - 1$ indices $i_1, \ldots, i_{r-1}$ uniformly from nw pairs (y, j), $y \in [z, z + w - 1]$ and $j \in \{1, \ldots, n\}$;
16   Compute $\widehat{h}_{(z,a)}$ according to (40);

$n = 30000$ (uncoupled), $n = 500 - 1000$ (coupled), and iteration number $t_{\max} = 3000$.

VIII. TWO THEOREMS AND DISCUSSION OF THRESHOLD SATURATION

Theorem 2 provides theoretical support for the simplifications of the cavity equations discussed in section VI-C.

Theorem 2: Consider the fixed point equations (39) and (40) for the individual Poisson LDGM(l, R) ensemble with a fixed β. Take any initial continuous density $\hat{q}^{(0)}(\hat{h})$ and consider iterations $\hat{q}^{(t)}(\hat{h})$. There exists $l_0 \in \mathbb{N}$ such that for $l > l_0$, $\lim_{t \to \infty} \widehat{h}^{(t)} = 0$ almost surely.

The proof[10] is presented in Appendix D. Note that l_0 depends on β and R. However we expect that as long as $\beta < \beta_c$ the result holds for all $l \geq 3$ and R. This is corroborated by the numerical observations. When we solve equations (39) and (40) by population dynamics with $\hat{q}^{(0)}(\hat{h})$ the uniform distribution, we observe that for a finite range of β depending on (l, R), the densities $q^{(t)}(h), \widehat{q}^{(t)}(\widehat{h})$ tend to a Dirac distribution at the origin. The range of β for which this occurs always contains the interval $[0, \beta_c]$ irrespective of (l, R). These observations also hold for many other initial distributions. We note that these observations break down for β large enough.

Theorem 3 partially justifies (51) which is the basis for the computation of the complexity in the large degree limit in section VI-D.

---

[10]It can be extended to other irregular degree distributions.

Algorithm 3 Population Dynamics for (42) and (43)

1 **for** $z = 1$ *to* L **do**
2   **for** $i = 1$ *to* n **do**
3   Set $\eta^{\pm}_{(z,i)} = \pm\infty$ and draw $h_{(z,i)}$ from $q_z(h)$;

4 **for** $t \in \{1, \ldots, t_{max}\}$ **do**
5   **for** $z = 1$ *to* L **do**
6   **for** $a = 1$ *to* n **do**
7   *Generate J randomly and generate a new triple $(\widehat{h}_{(z,a)}, \widehat{\eta}^{+}_{(z,a)}, \widehat{\eta}^{-}_{(z,a)})$:*
8   Choose $r - 1$ indices $i_1, \ldots, i_{r-1}$ uniformly from nw pairs (y, j), $y \in [z, z + w - 1]$ and $j \in \{1, \ldots, n\}$;
9   Compute $\widehat{h}_{(z,a)}$ according to (40);
10   Generate a configuration $\sigma_1, \ldots, \sigma_{r-1}$ from $\nu_1(\ldots | + J, h_{i_1}, \ldots, h_{i_{r-1}})$ in (41);
11   Compute $\widehat{\eta}^{+}_{(z,a)}$ by plugging $\eta^{\sigma_1}_{i_1}, \ldots, \eta^{\sigma_{r-1}}_{i_{r-1}}$ in (43);
12   Generate a configuration $\sigma_1, \ldots, \sigma_{r-1}$ from $\nu_1(\ldots | - J, h_{i_1}, \ldots, h_{i_{r-1}})$ in (41);
13   Compute $\widehat{\eta}^{-}_{(z,a)}$ by plugging $\eta^{\sigma_1}_{i_1}, \ldots, \eta^{\sigma_{r-1}}_{i_{r-1}}$ in (43);

14   **for** $z = 1$ *to* $L + w - 1$ **do**
15   **for** $i = 1$ *to* n **do**
16   *Generate a new triple $(h_{(z,i)}, \eta^{+}_{(z,i)}, \eta^{-}_{(z,i)})$:*
17   Choose $l - 1$ pair indices $a_1, \ldots, a_{l-1}$ uniformly from nw pairs (y, j), $y \in [z - w + 1, z]$ and $j \in \{1, \ldots, n\}$;
18   **if** *for some index k, $a_k = (y, j)$ and $y < 1$* **then**
19   Set $(\widehat{h}_{a_k}, \widehat{\eta}^{+}_{a_k}, \widehat{\eta}^{-}_{a_k}) = (0, 0, 0)$;(*)
20   Set $h_{(z,i)} = \sum_{k=1}^{l-1} \widehat{h}_{a_k}$ and $\eta^{\pm}_{(z,i)} = \sum_{k=1}^{l-1} \widehat{\eta}^{\pm}_{a_k}$;

Theorem 3: Consider the fixed point equations (47) and (48) associated to the individual Poisson LDGM(l, R) ensemble for some l, R and β ($w = 1$ in the equations). Let $\widehat{\eta}^{(t)}$ be a random variable distributed according to $\widehat{q}^{+(t)}(\widehat{\eta})$ at iteration t. Assume that the initial density is

$$\widehat{q}^{+(0)}(\hat{\eta}) = \sum_{J = \pm 1} \frac{1 + J \tanh(\beta)}{2} \delta(\hat{\eta} - J).$$

Then,

- *i) For all t,*

$$\widehat{q}^{+(t)}(-\widehat{\eta}) = e^{-2\beta \widehat{\eta}} \widehat{q}^{+(t)}(\widehat{\eta}), \qquad (58)$$
$$q^{+(t)}(-\eta) = e^{-2\beta \eta} q^{+(t)}(\eta). \qquad (59)$$

- *ii) For any $\delta > 0$, $\epsilon > 0$ and $B > 0$, there exits l_1 such that for $l > l_1$ and all t.*

$$\mathbb{P}\left\{1 - \epsilon \leq \widehat{\eta}^{(t)} \leq 1\right\} > \frac{e^{2\beta}}{1 + e^{2\beta}}(1 - \delta), \quad (60)$$

$$\mathbb{P}\left\{-1 \leq \widehat{\eta}^{(t)} \leq -1 + \epsilon\right\} > \frac{1}{1 + e^{2\beta}}(1 - \delta). \quad (61)$$

The proof is presented in Appendix E.

We now wish to briefly discuss the issue of threshold saturation. One of the main observations of this work is the saturation of the dynamical inverse temperature threshold towards the condensation threshold: $\lim_{w \to +\infty} \lim_{L \to +\infty} \beta_d(L, w) = \beta_c(w = 1)$. This is analogous to threshold saturation in coding theory where the Belief Propagation threshold of the coupled code ensemble saturates towards the MAP threshold. In this latter case we have proofs of this phenomenon for the rather general case of irregular LDPC codes (with bounded degrees) and binary-input memoryless-output symmetric channels [19], [21], [44]. The proof in [44] is based on the analysis of a potential function given by the replica-symmetric formula (an average form of the Bethe free energy) for the (infinite length) conditional input-output entropy of the code ensemble. We expect that, for the present problem, a proof of threshold saturation could be based on a potential function given by the complexity functional introduced in Section VI. Theorem 2 hints that the only solutions (for β in the range of interest) of equations (39) and (40) is trivial. Then the complexity functional reduces to a simplified form as explained in Section VI-C. It is possible to check by explicit functional differentiation that the stationary point equations for this functional are precisely the fixed point equations (47), (48), and as already pointed out these are the density evolution relations for an LDGM code over a test-BSC-channel with half-log-likelihood parameter β. A proof of threshold saturation could eventually be achieved along these lines, using the techniques of the recent paper [44], which also addresses LDGM codes.

IX. CONCLUSION

Let us briefly summarize the main points of this paper. We have investigated a simple spatially coupled LDGM code ensemble for lossy source coding. No optimization on the degree distribution is required: the check degree is regular and the code-bit degree is Poisson. We have shown that the algorithmic rate-distortion curve of a low complexity encoder based on BPGD allows to approach the ultimate Shannon rate-distortion curve, for all compression rates, when the check degree grows large. The inverse temperature parameter (or equivalently test-channel parameter) of the encoder may be optimized. However we have observed numerically, and have argued based on large degree calculations, that a good universal choice is $\beta_{\mathrm{sh}}(R)$, given by tuning the test channel to capacity. We recall that for the underlying (uncoupled) ensemble the same encoder does not perform well, indeed as the degree grows large, the difference between the algorithmic rate-distortion and Shannon rate-distortion curves grows. Insight into the excellent performance of the BPGD algorithm for spatially coupled ensemble is gained by studying the phase diagram of the Gibbs measure on which the BPGD encoder is based. We have found, by applying the cavity method to the spatially coupled ensemble, that the dynamical (inverse temperature) threshold β_d saturates towards the condensation (inverse temperature) threshold β_c. For this reason the BPGD encoder can operate close to the condensation threshold β_c, which itself tends in the large degree limit to $\beta_{\mathrm{sh}}(R)$,

the test channel parameter tuned at capacity. For the underlying (uncoupled) ensemble the dynamical threshold moves in the opposite direction in the large degree limit so that the BPGD algorithm cannot operate close to the Shannon limit.

We mention some open questions that are left out by the present study and which would deserve more investigations.

For fixed degrees the best value of the inverse temperature β_* of the BPGD algorithm is close to, but systematically lower, than the dynamical temperature β_d. While the value of β_d can be calculated by the cavity theory, here we determine β_* by purely empirical means and it is not clear what are the theoretical principles that allow to determine its value. As the graph is decimated the degree distribution changes and the effective dynamical temperature of the decimated graphs should evolve to slightly different values. It is tempting to conjecture that β_* is the limit of such a sequence of dynamical temperatures. A related phenomenon has been observed for the dynamical threshold with respect to clause density for random constraint satisfaction problems in their SAT phase [45].

The decimation process used in this paper is hard to analyze rigorously because it is not clear how to keep track of the statistics of the decimated graph. As a consequence it is also not clear how to compute the optimal value of the inverse temperature along the decimation process (we fix this value once for all). Progress on this problem could maybe be achieved by redesigning the decimation process, however how to achieve this is at the moment not clear. We would like to point out that a related process has been investigated in recent works [46] for the K-SAT problem in the large K limit up to the dynamical threshold(in the SAT phase). These methods could be of use also in the present case.

In this contribution we have investigated a linear decoding rule. Source coding with non-linear rules are of interest and have been studied in [32]. It is an open question to look at the algorithmic performance of such codes in the framework of spatial coupling.

Finally, while a rigorous control of the full cavity method is, in general, beyond present mathematical technology, there are sub-problems for which progress can presumably be made. For example in the present case we have observed that the cavity equations reduce (in the dynamical phase $\beta_d < \beta < \beta_c$) to density evolution equations for an LDGM code on a BSC. The saturation of the dynamical temperature β_d to the condensation temperature β_c appears to be very similar to the threshold saturation phenomenon of channel coding theory. We have by now a host of mathematical methods pertaining to this effect for LDPC on general binary memoryless channels [19], [21]. We think that these methods could be adapted to prove the saturation of β_d towards β_c. One extra difficulty faced in the present problem is that the "trivial" fixed point of density evolution equations of LDPC codes is not always present in the LDGM case.

APPENDIX A
A PRIMER ON THE CAVITY METHOD

We give a brief introduction to the cavity method for general spin systems on sparse graphs. As explained in Sect. V-A,

turning this formalism into a rigorous mathematical theory is a long standing open problem. However, it allows to compute many quantities of interest. In appendices B and C we specialize to the probability distribution (8).

The treatment given here applies to *single instances*. Let $\Gamma = (V, C, E)$ a factor graph which is assumed to be locally tree like. We attach spins σ_j, $j \in V$ to variable nodes, and constraint functions $\psi_a (\{\sigma_i, i \in \partial a\})$, $a \in C$ to check nodes. We sometimes use the notation $\underline{\sigma}_{\partial a} = \{\sigma_i, i \in \partial a\}$ as a shorthand. The formalism developed in this appendix is valid for general spin variables belonging to a finite alphabet $\sigma_j \in \mathcal{X}$. The constraint functions depend only on the set of spins connected to a. We are interested in the thermodynamic limit where $|V| = N$ and $|C| = M$ tend to infinity and the ratio M/N is kept fixed. We consider the general class of Gibbs distributions of the form

$$\mu (\underline{\sigma}) = \frac{1}{Z} \prod_{a \in C} \psi_a (\{\sigma_i, i \in \partial a\}), \qquad (62)$$

where Z is the partition function. The free energy of an instance is defined as usual

$$\phi (\beta) = -\frac{1}{N\beta} \ln Z (\beta). \qquad (63)$$

One of the goals of the cavity method is to compute this free energy in the limit $N \to +\infty$.

Let us first outline the general strategy. For locally tree like graphs, one can compute the marginals for a given node by restricting the measure to a tree. In the absence of long range correlations[11] the marginal does not depend on the boundary conditions at the leaf nodes, and the BP equations have one relevant solution. The BP marginals then constitute a good description of the measure (62). In particular, the true free energy is well approximated by replacing this solution in the Bethe free energy functional. As the control parameters vary long range correlations may appear. In such a situation the marginals computed on a tree will depend on the boundary conditions at the leaf nodes, and the BP equations will have many relevant solutions yielding nearly the same Bethe free energy. The cavity method assumes that the measure (62) is then described by a convex superposition of "extremal measures". There may be a large number of extremal measures. A good proxy for the extremal measures is given by the BP marginals. The convex superposition of extremal measures yields a new statistical model on the same factor graph, the so-called *level-one model*. Assuming that the level one model does not display long range correlations, one can solve it using BP equations and the Bethe free energy. Otherwise, the cavity method iterates the previous considerations and constructs a level-two model. However, this usually becomes bewildering and one stops at the first level. In the following paragraphs we give a concrete implementation of these ideas.

The BP equations are a set of fixed point equations satisfied by messages $\{\nu_{i \to a}, \widehat{\nu}_{a \to i}\} = (\underline{\nu}, \underline{\widehat{\nu}})$,

$$\widehat{\nu}_{a \to i} = \widehat{g}_{\mathrm{BP}} \left(\{\nu_{j \to a}\}_{j \in \partial a \setminus i} \right), \ \nu_{i \to a} = g_{\mathrm{BP}} (\{\widehat{\nu}_{b \to i}\}_{b \in \partial i \setminus a}), \qquad (64)$$

[11]More precisely point-to-set correlations [30].

where

$$\widehat{g}_{\mathrm{BP}} \left(\{\nu_{j \to a}\}_{j \in \partial a \setminus i} \right) = \frac{\sum_{\underline{\sigma}_{\partial a \setminus i}} \psi_a (\underline{\sigma}_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{j \to a} (\sigma_j)}{\sum_{\underline{\sigma}_{\partial a}} \psi_a (\underline{\sigma}_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{j \to a} (\sigma_j)}$$

$$g_{\mathrm{BP}} (\{\widehat{\nu}_{b \to i}\}_{b \in \partial i \setminus a}) = \frac{\prod_{b \in \partial i \setminus a} \widehat{\nu}_{b \to i} (\sigma_i)}{\sum_{\sigma_i} \prod_{b \in \partial i \setminus a} \widehat{\nu}_{b \to i} (\sigma_i)}.$$

When there is only one relevant solution, the BP marginal for σ_j is $\nu(\sigma_j) = \prod_{a \in \partial j} \nu_{a \to j}(\sigma_j)$. The set of messages is a proxy for the measure (62) in the sense that in principle one can "reconstruct" the measure from this set. The Bethe free energy functional which approximates $\phi(\beta)$ is given by

$$\phi^{\mathrm{Bethe}} (\underline{\nu}, \underline{\widehat{\nu}}) = \frac{1}{N} \left\{ \sum_{i \in V} \phi_i + \sum_{a \in C} \phi_a - \sum_{(i,a) \in E} \phi_{ai} \right\} \qquad (65)$$

where

$$\phi_i (\{\widehat{\nu}_{b \to i}\}_{b \in \partial i}) = -\frac{1}{\beta} \ln \sum_{\sigma_i} \prod_{b \in \partial i} \widehat{\nu}_{b \to i} (\sigma_i)$$

$$\phi_a \left(\{\nu_{j \to a}\}_{j \in \partial a} \right) = -\frac{1}{\beta} \ln \sum_{\underline{\sigma}_{\partial a}} \psi_a (\underline{\sigma}_{\partial a}) \prod_{j \in \partial a} \nu_{j \to a} (\sigma_j)$$

$$\phi_{ai} (\nu_{i \to a}, \widehat{\nu}_{a \to i}) = -\frac{1}{\beta} \ln \sum_{\sigma_i} \nu_{i \to a} (\sigma_i) \widehat{\nu}_{a \to i} (\sigma_i).$$

As explained before, in the presence of long range correlations this formalism is too simplistic. The cavity method assumes that: (i) the Gibbs distribution (62) is a convex sum of extremal measures; (ii) to leading exponential order, the number of solutions of the BP equations is equal to the number of extremal measures; (iii) the free energy of an extremal measure is well approximated by the Bethe free energy of the BP fixed point. These assumptions suggest that the Gibbs distribution (62) is well approximated by the following convex superposition

$$\mu (\underline{\sigma}) \approx \frac{1}{Z} \sum_{(\underline{\nu}, \underline{\widehat{\nu}}) \in \mathrm{BP}} e^{-\beta N \phi^{\mathrm{Bethe}} (\underline{\nu}, \underline{\widehat{\nu}})} \mu_{(\underline{\nu}, \underline{\widehat{\nu}})} (\underline{\sigma}) \qquad (66)$$

The measures $\mu_{(\underline{\nu}, \underline{\widehat{\nu}})}$ are the ones whose marginals are given by the BP marginals computed from $(\underline{\nu}, \underline{\widehat{\nu}})$. They play the role of the "extremal measures". The sum is over solutions of the BP equations. In principle one should sum only over stable solutions, i.e. local minima of the Bethe free energy. However at low temperatures these are expected to be exponentially more numerous than the other critical points and it is assumed to be a good approximation to sum over all BP solutions. The normalization factor yields the partition function

$$Z \approx \sum_{(\underline{\nu}, \underline{\widehat{\nu}}) \in \mathrm{BP}} e^{-\beta N \phi^{\mathrm{Bethe}} (\underline{\nu}, \underline{\widehat{\nu}})}. \qquad (67)$$

In order to compute this partition function and uncover the properties of the convex decomposition (66) we introduce the level-one statistical mechanical model. The dynamical variables of this model are the BP messages $(\underline{\nu}, \underline{\widehat{\nu}})$. According to (66), (67) the probability distribution over $(\underline{\nu}, \underline{\widehat{\nu}})$ is

$$\mu_{\mathrm{level}-1} (\underline{\nu}, \underline{\widehat{\nu}}) = \frac{e^{-\beta N \phi^{\mathrm{Bethe}} (\underline{\nu}, \underline{\widehat{\nu}})}}{Z_{\mathrm{level}-1}} \mathbb{I} ((\underline{\nu}, \underline{\widehat{\nu}}) \in \mathrm{BP}), \qquad (68)$$

and

$$Z_{\text{level}-1} = \sum_{(\underline{v},\widehat{\underline{v}}) \in \text{BP}} e^{-\beta N \phi^{\text{Bethe}}(\underline{v},\widehat{\underline{v}})}, \tag{69}$$

The level-one free energy is defined as usual,

$$\phi_{\text{level}-1}(\beta) = -\frac{1}{\beta N} \ln Z_{\text{level}-1}. \tag{70}$$

From (67) it should be clear that $\phi(\beta) \approx \phi_{\text{level}-1}(\beta)$. The average Bethe free energy, or level-one *internal energy*, is given by

$$\varphi_{\text{int}}(\beta) = \frac{1}{N} \langle \phi^{\text{Bethe}}[\underline{v},\widehat{\underline{v}}] \rangle_{\text{level}-1} \tag{71}$$

Here the bracket denotes the average with respect to (68).

One also needs to compute the Shannon-Gibbs entropy $\Sigma(\beta)$ of $\mu_{\text{level}-1}$. An important "trick" is to replace the *explicit* β dependence in (68), (69), (70) by βx where x is for the moment an arbitrary parameter.[12] This parameter turns out to play a crucial role and is called the Parisi parameter. This gives us an x-dependent level-one auxiliary model

$$\mu_{\text{level}-1}\left(\underline{v},\widehat{\underline{v}}; x\right) = \frac{e^{-\beta x N \phi^{\text{Bethe}}(\underline{v},\widehat{\underline{v}})}}{Z_{\text{level}-1(x)}} \mathbb{I}\left((\underline{v},\widehat{\underline{v}}) \in \text{BP}\right), \quad (72)$$

and

$$Z_{\text{level}-1}(x) = \sum_{(\underline{v},\widehat{\underline{v}}) \in \text{BP}} e^{-\beta x N \phi^{\text{Bethe}}(\underline{v},\widehat{\underline{v}})}, \tag{73}$$

and also

$$\phi_{\text{level}-1}(\beta; x) = -\frac{1}{\beta x N} \ln Z_{\text{level}-1}(x). \tag{74}$$

It is then a matter of simple algebra to check that the Shannon-Gibbs entropy $\Sigma(\beta)$ is given by

$$\Sigma(\beta) = \Sigma(\beta; x) \equiv \beta x^2 \frac{\partial}{\partial x} \phi_{\text{level}-1}(\beta; x)|_{x=1}, \tag{75}$$

and that

$$\Sigma(\beta) = \beta(\varphi_{\text{int}}(\beta) - \phi_{\text{level}-1}(\beta)). \tag{76}$$

Considering formulas (69), (71) and (76), it is not hard to argue that $e^{N\Sigma(\beta)}$ is (to leading exponential order) the number of BP solutions with free energy $\varphi_{\text{int}}(\beta)$ contributing to the sum (69). The quantity $\Sigma(\beta)$ (a kind of entropy) is called the complexity. It is the growth rate of the number of extremal measures dominating the convex decomposition (66).

We explain later on how to concretely compute $\phi_{\text{level}-1}(\beta)$, $\varphi_{\text{int}}(\beta)$ and $\Sigma(\beta)$. Let us immediately describe how $\Sigma(\beta)$ informs us about the convex decomposition of the Gibbs distribution. For a large class of problems one finds that $\Sigma(\beta) = 0$ for $\beta < \beta_d$, which signals that only one extremal measure contributes to the Gibbs distribution. At β_d the complexity jumps to a non-zero value and then decreases as a function of β till β_c after which it takes negative values. In the range $\beta_d < \beta < \beta_c$ where $\Sigma(\beta) > 0$ an exponentially large (with respect to N) number of extremal measures with the same internal free energy $\varphi_{\text{int}}(\beta)$ contribute to the
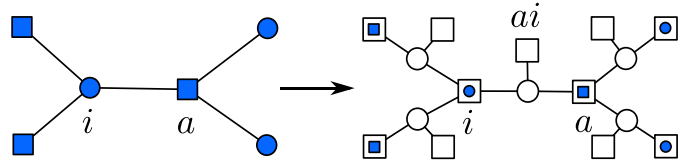
[12] Note that there is also an *implicit* β dependence in $\phi^{\text{Bethe}}[\underline{v},\widehat{\underline{v}}]$.

Fig. 10. On the left, an example of an original graph Γ. On the right its corresponding graph Γ_1 for the level-one model.

Gibbs distribution. Beyond β_c one finds a negative complexity: this is inconsistent with the fact that it is an entropy. In order to enforce this constraint correctly one is forced to take the Parisi parameter $0 < x < 1$ in (75). More precisely, one sets x to the largest possible value (less than 1) such that $\Sigma(\beta) = 0$. With this prescription[13] for the correct value of x when $\beta > \beta_c$, one computes the internal free energy and the free energy and the complexity from the x-dependent level-one model. The complexity is zero by construction which means that there exist at most a sublinear (believed to be finite) number of extremal measures contributing to the Gibbs distribution. This phenomenon is called condensation.

The nature of the thresholds β_d and β_c has been discussed in Sect. (V-A) and we do not come back to this issue here.

We now show how the (x-dependent) level-one model is solved in practice. The main idea is to apply again the BP and Bethe equations for this model. The first step is to recognize that, if $\Gamma = (V, C, E)$ is the original factor graph, then the level-one model has the factor graph $\Gamma_1 = (V_1, C_1, E_1)$ described on Fig. 10.

A variable node $i \in V$, becomes a function node $i \in C_1$, with the function

$$\psi_i^{(1)} = \prod_{a \in \partial i} \mathbb{I}\left(v_{i \to a} = g_{\text{BP}}\right) e^{-x\beta\phi_i}. \tag{77}$$

A function node $a \in C$ remains a function node $a \in C_1$ with factor

$$\psi_a^{(1)} = \prod_{i \in \partial a} \mathbb{I}\left(\widehat{v}_{a \to i} = \widehat{g}_{\text{BP}}\right) e^{-x\beta\phi_a}. \tag{78}$$

An edge $(a, i) \in E$, becomes a variable node $(a, i) \in V_1$. The dynamical variables are now couples of distributions $(v_{a \to i}, \widehat{v}_{a \to i})$. There is also an extra function node attached to each variable node of the new graph, or equivalently attached to each edge of the old graph. The corresponding function is

$$\psi_{ai}^{(1)} = e^{x\beta\phi_{ai}}. \tag{79}$$

With these definitions, Equ. (72) can be written as

$$\mu_{\text{level}-1}(\underline{v},\widehat{\underline{v}}; x) = \frac{1}{Z_{\text{level}-1}(x)} \prod_{i \in V} \psi_i^{(1)} \prod_{a \in C} \psi_i^{(1)} \prod_{ai \in E} \psi_{ai}^{(1)}. \tag{80}$$

For the distributions $(\underline{v},\widehat{\underline{v}})$ that satisfy the BP equations (64), some algebra leads to the useful formulas

$$\begin{cases} e^{-x\beta(\phi_a - \phi_{ai})} = \widehat{z}_{a \to i}^x \\ e^{-x\beta(\phi_i - \phi_{ai})} = z_{i \to a}^x \end{cases}$$

[13] One can argue that the Parisi parameter is a kind of "Lagrange multiplier" that enforces the non-negativity of the complexity in the level-one model.
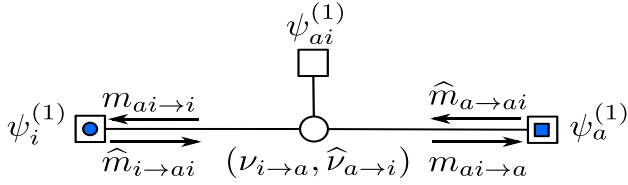
Fig. 11. Messages are labeled by m if they are outgoing from a variable node in V_1 and by $\widehat{m}$ if they are outgoing from a function node in C_1.

where

$$
\begin{cases}
z_{i \to a} = \sum_{\sigma_i} \prod_{b \in \partial i \setminus a} \widehat{v}_{b \to i}(\sigma_i) \\
\widehat{z}_{a \to i} = \sum_{\sigma_{\partial a}} \psi_a(\sigma_{\partial a}) \prod_{\partial j \in a \setminus i} \widehat{v}_{j \to a}(\sigma_i)
\end{cases}
$$

The BP equations for (80) involve four kind of messages as shown on figure 11.

Messages from a (new) function node to a (new) variable node satisfy

$$
\widehat{m}_{a \to ai} \simeq \sum_{(\underline{v}, \underline{\widehat{v}}) \setminus (v_{i \to a}, \widehat{v}_{a \to i})} \psi_a^{(1)} \prod_{aj \in \partial a \setminus ai} m_{aj \to a}
$$

$$
= e^{-x\beta\phi_{ai}} \sum_{\underline{v} \setminus v_{i \to a}} \mathbb{I}(\widehat{v}_{a \to i} = \widehat{g}_{\mathrm{BP}})(\widehat{z}_{a \to i})^x \prod_{aj \in \partial a \setminus ai} m_{aj \to a}
$$

and

$$
\widehat{m}_{i \to ai} \simeq \sum_{(\underline{v}, \underline{\widehat{v}}) \setminus (v_{i \to a}, \widehat{v}_{a \to i})} \psi_i^{(1)} \prod_{bi \in \partial i \setminus a} \widehat{m}_{bi \to i}
$$

$$
= e^{-x\beta\phi_{ai}} \sum_{\underline{\widehat{v}} \setminus \widehat{v}_{a \to i}} \mathbb{I}(v_{i \to a} = g_{\mathrm{BP}})(z_{i \to a})^x \prod_{bi \in \partial i \setminus a} \widehat{m}_{bi \to i}.
$$

where the symbol $\simeq$ means equal up to a normalization factor. Messages from a (new) function node to a (new) variable node satisfy

$$
\begin{cases}
m_{ai \to i} \simeq e^{x\beta\phi_{ai}} \widehat{m}_{a \to ai} \\
m_{ai \to a} \simeq e^{x\beta\phi_{ai}} \widehat{m}_{i \to ai}.
\end{cases}
\tag{81}
$$

Notice that $m_{ai \to i}$ (resp. $m_{ai \to a}$) becomes independent of $\widehat{v}_{a \to i}$ (resp. $v_{i \to a}$). This allows us to make a simplification by defining the following distributions

$$
\begin{cases}
Q_{i \to a}(v_{i \to a}) = m_{ai \to a}(v_{i \to a}, \widehat{v}_{a \to i}) \\
\widehat{Q}_{a \to i}(\widehat{v}_{a \to i}) = m_{ai \to i}(v_{i \to a}, \widehat{v}_{a \to i}).
\end{cases}
$$

Distributions Q and $\widehat{Q}$ are called cavity messages, and live on the edges of the original factor graph $\Gamma = (V, C, E)$. From now on we can forget about the factor graph $\Gamma_1 = (V_1, C_1, E_1)$. The cavity messages satisfy

$$
\widehat{Q}_{a \to i}(\widehat{v}_{a \to i}) \simeq \sum_{\underline{v}} \mathbb{I}(\widehat{v}_{a \to i} = \widehat{g}_{\mathrm{BP}}) \widehat{z}_{a \to i}^x \prod_{j \in \partial a \setminus i} Q_{j \to a}(v_{j \to a})
$$

$$
Q_{i \to a}(v_{i \to a}) \simeq \sum_{\underline{\widehat{v}}} \mathbb{I}(v_{i \to a} = g_{\mathrm{BP}}) z_{i \to a}^x \prod_{b \in \partial i \setminus a} \widehat{Q}_{b \to i}(\widehat{v}_{b \to i}).
$$

$$
\tag{82}
$$

The Bethe free energy functional of the level-one model can be expressed as a functional of the cavity messages (one way to determine this functional is to write down the functional whose critical points are given by Equ. (82)). This is an

approximation for the true free energy (70) of the level-one model.

$$
\phi_{\mathrm{level}-1}^{\mathrm{Bethe}}(\underline{Q}, \underline{\widehat{Q}}; x) := \frac{1}{N} \left\{ \sum_{i \in V} \mathcal{F}_i \right.
$$

$$
\left. + \sum_{a \in C} \mathcal{F}_a^{\mathrm{Bethe}} - \sum_{(i,a) \in E} \mathcal{F}_{ai} \right\}
\tag{83}
$$

where

$$
\mathcal{F}_i\left(\{\widehat{Q}_{b \to i}\}_{b \in \partial i}\right) = -\frac{1}{x\beta} \ln \sum_{\underline{\widehat{v}}} e^{-x\beta\phi_i} \prod_{b \in \partial i} \widehat{Q}_{b \to i}
$$

$$
\mathcal{F}_a\left(\{Q_{j \to a}\}_{j \in \partial a}\right) = -\frac{1}{x\beta} \ln \sum_{\underline{v}} e^{-x\beta\phi_a} \prod_{j \in \partial a} Q_{j \to a}
$$

$$
\mathcal{F}_{ai}(Q_{i \to a}, \widehat{Q}_{a \to i}) = -\frac{1}{x\beta} \ln \sum_{v, \widehat{v}} e^{-x\beta\phi_{ai}} Q_{i \to a} \widehat{Q}_{a \to i}. \tag{84}
$$

In principle one has to solve the cavity equations (82) for $0 < x \le 1$, and compute the x-dependent free energy $\phi_{\mathrm{level}-1}^{\mathrm{Bethe}}$. From this free energy we obtain the complexity by computing the derivative in equation (75). This allows to determine the thresholds β_d and β_c. For $\beta < \beta_c$ the free energy is given by $\phi_{\mathrm{level}-1}^{\mathrm{Bethe}}|_{x=1}$. This function has no singularities, which means that there are no static (thermodynamic) phase transitions for $\beta < \beta_c$. In this phase one has $\Sigma(\beta; x = 1) \ge 0$. For $\beta > \beta_c$ one enforces a zero complexity by setting the Parisi parameter to a value $0 < x_* < 1$ s.t. $\Sigma(\beta; x_*) = 0$. The free energy is not analytic at β_c, due to the change of x parameter. This a static phase transition threshold.

In practice, as long as we are interested only in the range $\beta < \beta_c$ we can set $x = 1$. It is then possible to simplify the cavity equations (82) and the level-1 free energy (83). In the next appendix we perform these simplifications for the case at hand.

APPENDIX B
APPLICATION OF THE CAVITY EQUATIONS
TO THE LOSSY SOURCE CODING

We apply the formalism of appendix A to the measure $\mu_\beta(\underline{u}|\underline{x})$ (see Equ.(8)). Instead of working with the alphabet $\{0, 1\}$, we find it convenient to use the mapping $\sigma_i = (-1)^{u_i}$ and $J_a = (-1)^{x_a}$ to the alphabet $\{-1, +1\}$. The measure (8) is of the form (62) with

$$
\psi_a(\{\sigma_i, i \in \partial a\}) = e^{-\beta(1 - J_a \prod_{i \in \partial a} \sigma_i)}. \tag{85}
$$

The probability distributions $v_{i \to a}(\sigma_i)$ and $\widehat{v}_{a \to i}(\sigma_i)$ are entirely characterized by their means, $\tanh\beta\eta_{i \to a}$ and $\tanh\beta\widehat{\eta}_{a \to i}$, as follows (we drop the subscripts)

$$
v(\sigma) = \frac{1 + \sigma \tanh\beta\eta}{2}. \tag{86}
$$

With this parameterization, the BP equations (64) for the model (8) become

$$
\begin{cases}
\widehat{\eta}_{a \to i} = \widehat{g}_{\mathrm{BP}}(\{\eta_{j \to a}\}_{j \in \partial a \setminus i} \mid J_a) \\
\eta_{i \to a} = g_{\mathrm{BP}}(\{\widehat{\eta}_{b \to i}\}_{b \in \partial i \setminus a}),
\end{cases}
\tag{87}
$$

where

$$\widehat{g}_{BP}\left(\{\eta_{j\to a}\}_{j\in\partial a\setminus i}\mid J_a\right) = \frac{J_a}{\beta}\operatorname{atanh}(\tanh\beta$$
$$\times \prod_{j\in\partial a\setminus i}\tanh\beta\eta_{j\to a}) \quad (88)$$

and

$$g_{BP}(\{\widehat{\eta}_{b\to i}\}_{b\in\partial i\setminus a}) = \sum_{b\in\partial i\setminus a}\widehat{\eta}_{bi}. \quad (89)$$

The Bethe free energy per variables (65) reads

$$\phi^{Bethe}\left(\underline{\eta},\underline{\widehat{\eta}}\right) = -\beta^{-1}(\ln\left(1+e^{-2\beta}\right) + (R-1)\ln 2)$$
$$- \frac{1}{\beta N}\sum_{a\in C}(1-|\partial a|)\ln Z_1(\{\eta_{j\to a}\}_{j\in\partial a}\mid J_a)$$
$$- \frac{R}{\beta M}\sum_{i\in V}\ln Z_2(\{\widehat{\eta}_{b\to i}\}_{b\in\partial i}), \quad (90)$$

where

$$\begin{cases} Z_1(\{\eta_{j\to a}\}_{j\in\partial a}\mid J_a) = 1 + J_a(\tanh\beta)\prod_{i\in\partial a}\tanh\beta\eta_{i\to a} \\ Z_2(\{\widehat{\eta}_{a\to i}\}_{a\in\partial i}) = \frac{1}{2}\sum_{s\in\{-1,1\}}\prod_{a\in\partial i}(1+s\tanh\beta\,\widehat{\eta}_{a\to i}). \end{cases}$$

Since we have parameterized the BP messages by real numbers, the cavity messages $Q_{i\to a}$, $\widehat{Q}_{a\to i}$ become distributions on $\eta_{i\to a}$, $\widehat{\eta}_{a\to i}$. The cavity equations (82) reduce to

$$Q_{i\to a}(\eta_{i\to a}) \simeq \int\prod_{b\in\partial i\setminus a}d\widehat{\eta}_{b\to i}\widehat{Q}_{b\to i}(\widehat{\eta}_{b\to i})$$
$$\times Z_2^x(\{\widehat{\eta}_{b\to i}\}_{b\in\partial i\setminus a})\delta\left(\eta_{i\to a} - g_{BP}(\{\widehat{\eta}_{b\to i}\}_{b\in\partial i\setminus a})\right) \quad (91)$$

and

$$\widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i}) \simeq \int\prod_{j\in\partial a\setminus i}d\eta_{j\to a}Q_{j\to a}(\eta_{j\to a})$$
$$\times Z_1^x(\{\eta_{j\to a}\}_{b\in\partial i\setminus a})\delta\left(\widehat{\eta}_{a\to i} - \widehat{g}_{BP}(\{\eta_{j\to a}\}_{j\in\partial a\setminus i}\mid J_a)\right). \quad (92)$$

For the Bethe free energy of the level-one model one finds

$$\phi_{level-1}^{Bethe}\left(\underline{\eta},\underline{\widehat{\eta}};x\right) = -\beta^{-1}(\ln(1+e^{-2\beta}) + (R-1)\ln 2)$$
$$- \frac{1}{\beta x N}\sum_{a\in C}(1-|\partial a|)\ln\left\{\int\prod_{i\in\partial a}d\eta_{i\to a}Q_{i\to a}(\eta_{i\to a})\right.$$
$$\left.\times Z_1^x(\{\eta_{i\to a}\}_{i\in\partial a}\mid J_a)\right\}$$
$$- \frac{R}{\beta x M}\sum_{i\in V}\ln\left\{\int\prod_{a\in\partial i}d\widehat{\eta}_{a\to i}\widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i})\right.$$
$$\left.\times Z_2^x(\{\widehat{\eta}_{a\to i}\}_{a\in\partial i})\right\}. \quad (93)$$

We are interested in the range $\beta < \beta_c$ for which the Parisi parameter is set to $x = 1$. In this case the above equations greatly simplify. We first define average cavity messages

$$\begin{cases} h_{i\to a} = Av[Q_{i\to a}] \\ \widehat{h}_{a\to i} = Av[\widehat{Q}_{a\to i}], \end{cases} \quad (94)$$

where the functional $Av[P]$ is

$$Av[P] = \frac{1}{\beta}\operatorname{atanh}\left\{\int d\eta P(\eta)\tanh\beta\eta\right\}. \quad (95)$$

Thus $\tanh\beta h_{i\to a}$ and $\tanh\beta\widehat{h}_{a\to i}$ are real valued messages and are averages of $\tanh\beta\eta_{i\to a}$ and $\tanh\beta\widehat{\eta}_{a\to i}$ with respect to the cavity distributions $Q_{i\to a}(\eta_{i\to a})$ and $\widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i})$ respectively. The free energy of the level-one model for $x = 1$ can be expressed in terms of these real valued messages, and one finds

$$\phi_{level-1}^{Bethe}\left(\underline{h},\underline{\widehat{h}}\right) = -\beta^{-1}(\ln(1+e^{-2\beta}) + (R-1)\ln 2)$$
$$- \frac{1}{\beta N}\sum_{a\in C}(1-|\partial a|)\ln Z_1(\{h_{j\to a}\}_{j\in\partial a}\mid J_a)$$
$$- \frac{R}{\beta M}\sum_{i\in V}\ln Z_2(\{\widehat{h}_{b\to i}\}_{b\in\partial i}). \quad (96)$$

Remarkably, is the same than the original Bethe free energy functional $\phi^{Bethe}\left(\underline{\eta},\underline{\widehat{\eta}}\right)$ defined in (90), but now evaluated for the average fields $h_{i\to a}$ and $\widehat{h}_{a\to i}$. From the cavity equations (91)-(92) for $x = 1$, one can deduce that the average fields $h_{i\to a}$ and $\widehat{h}_{a\to i}$ satisfy

$$\begin{cases} \widehat{h}_{a\to i} = \widehat{g}_{BP}(\{h_{j\to a}\}_{j\in\partial a\setminus i}\mid J_a) \\ h_{i\to a} = g_{BP}(\{\widehat{h}_{b\to i}\}_{b\in\partial i\setminus a}). \end{cases} \quad (97)$$

Thus the average fields satisfy the BP equations (87).

To summarize, when $x = 1$, $\phi_{level-1}^{Bethe}$ equals ϕ^{Bethe} computed at a certain appropriate BP fixed point. This fixed point corresponds to messages $\tanh\beta h_{i\to a}$, $\tanh\beta\widehat{h}_{a\to i}$ which are an average of the BP solutions $\tanh\beta\eta_{i\to a}$, $\tanh\beta\widehat{\eta}_{a\to i}$ over the cavity distributions $Q_{i\to a}(\eta_{i\to a})$ and $\widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i})$. The messages $\tanh\beta\eta_{i\to a}$, $\tanh\beta\widehat{\eta}_{a\to i}$ describe the "extremal states" whereas the messages $\tanh\beta h_{i\to a}$, $\tanh\beta\widehat{h}_{a\to i}$ describe their convex superposition.

APPENDIX C
DENSITY EVOLUTION FOR THE CAVITY EQUATIONS OF LOSSY SOURCE CODING

The discussion in appendices A and B is valid for a single instance. It is expected that the free energy, internal free energy and complexity concentrate on their ensemble average, and in practice one computes their ensemble average. The ensemble average is performed over the graph ensemble and the Bernoulli source. In the present context this leads to the complicated set of fixed point equations (39)-(43) that links six densities.

To perform the ensemble average we assume that the cavity messages $Q_{i\to a}(\eta_{i\to a})$ and $\widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i})$ can be considered as i.i.d. realizations of random variables $Q_z(\eta)$ and $\widehat{Q}_z(\widehat{\eta})$. The random variables depend only on the *position* z along the spatial dimension and not on the *direction* of the edges $i\to a$ and $a\to i$. The distributions of these random variables are denoted $\mathcal{Q}_z$ and $\widehat{\mathcal{Q}}_z$. Note that the cavity messages are already distributions over real numbers, so that $\mathcal{Q}_z$ and $\widehat{\mathcal{Q}}_z$ are distributions of distributions. From the cavity equations (91), (92) it is easy to formally write down

the set of integral equations that these distributions of distributions satisfy.

We can write down probability distributions for the average fields $h_{i \to a}$ and $h_{a \to i}$,

$$\begin{cases} q_z(h) = \int \mathcal{D}\mathcal{Q}_z[Q]\delta(h - Av[Q]) \\ \widehat{q}_z(\widehat{h}) = \int \mathcal{D}\widehat{\mathcal{Q}}_z[\widehat{Q}]\delta(\widehat{h} - Av[\widehat{Q}]). \end{cases} \quad (98)$$

With the independence assumption on the cavity messages, relations (97) imply that these distributions satisfy (39) and (40). Furthermore from (96) we deduce formula (45) for the average level-one free energy.

We define the conditional distributions $q_z(\eta|h)$ and $\widehat{q}_z(\widehat{\eta}|\widehat{h})$

$$\begin{cases} q_z(\eta|h)q_z(h) = \int \mathcal{D}\mathcal{Q}_z[Q]Q(\eta)\delta(h - Av[Q]) \\ \widehat{q}_z(\widehat{\eta}|\widehat{h})\widehat{q}_z(\widehat{h}) = \int \mathcal{D}\widehat{\mathcal{Q}}_z[\widehat{Q}]\widehat{Q}(\widehat{\eta})\delta(\widehat{h} - Av[\widehat{Q}]), \end{cases} \quad (99)$$

and for $\sigma = \pm 1$,

$$\begin{cases} q_z^{\sigma}(\eta|h) = \dfrac{1 + \sigma \tanh \beta \eta}{1 + \sigma \tanh \beta h} q_z(\eta|h) \\ \widehat{q}_z^{\sigma}(\widehat{\eta}|\widehat{h}) = \dfrac{1 + \sigma \tanh \beta \widehat{\eta}}{1 + \sigma \tanh \beta \widehat{h}} q_z(\widehat{\eta}|\widehat{h}). \end{cases} \quad (100)$$

These distributions satisfy (42)-(43).

With the six distributions $q_z(h)$, $\widehat{q}_z(\widehat{h})$, $q_z^{\sigma=\pm 1}(\eta|h)$ and $\widehat{q}_z^{\sigma=\pm 1}(\widehat{\eta}|\widehat{h})$ we can compute the complexity. We use (see (76))

$$\Sigma(\beta) = \beta(\varphi_{\text{int}}^{\text{Bethe}}(\beta) - \phi_{\text{level}-1}^{\text{Bethe}}(\beta)). \quad (101)$$

Since we already know that $\phi_{\text{level}-1}^{\text{Bethe}}(\beta)$ is given by (45), it remains to compute the internal free energy in the Bethe approximation. For this purpose we use

$$\varphi_{\text{int}}^{\text{Bethe}}(\beta) = \frac{\partial}{\partial x}(x\phi_{\text{level}-1}^{\text{Bethe}}(\beta; x))|_{x=1}. \quad (102)$$

We compute the x-derivative on (93), and average over the cavity distributions, the graph ensemble and the Bernoulli source. After some algebra one finds that $\varphi_{\text{int}}^{\text{Bethe}}(\beta)$ is given by (46).

APPENDIX D
PROOF OF THEOREM 2

We first state two useful lemmas

Lemma 1: Let the random variable X is distributed according to a Poisson distribution with mean λ.

$$\mathbb{P}(X < \frac{\lambda t}{2}) < \exp(-\frac{\lambda t}{10}), \quad t \leq 1,$$

$$\mathbb{P}(X > \frac{3\lambda t}{2}) < \exp(-\frac{\lambda t}{10}), \quad t \geq 1.$$

Proof: Use the Chernoff bound. ∎

Lemma 2: Let

$$\epsilon_1 = \beta \frac{3l}{2R}(\tanh \epsilon_0)^{(lR^3)^{1/4}}$$

$$\delta_1 = \exp(-\frac{l}{10R}) + \frac{l}{R}\exp(-\frac{\epsilon_0\sqrt{Rl}}{\beta\sqrt{3\pi}}).$$

with $\epsilon_0 = \min(1/2, \beta/2)$. Consider the recursions for $t \geq 1$

$$\epsilon_{t+1} = (t+1)\beta\frac{3l}{2R}(\tanh \epsilon_t)^{(lR^3)^{1/4}},$$

$$\delta_{t+1} = \exp(-\frac{l}{10R}(t+1)) + \frac{l}{R}(2\sqrt{\delta_t})^{l-1}.$$

There exist an integer l_0 (depending only on R and β) such that for $l \geq l_0$,

- i) $\epsilon_t \leq \frac{1}{2^{t+1}}$ for $t \geq 0$.
- ii) $\delta_t < 2\exp(-\frac{l}{5R}t)$ for $t \geq 2$.

Proof: Consider (i). At $t = 0$, $\epsilon_0 \leq 1/2$. Assume that $\epsilon_{t-1} \leq \frac{1}{2^t}$ for $t \geq 1$, then

$$\epsilon_t = t\beta\frac{3l}{2R}(\tanh \epsilon_{t-1})^{(lR^3)^{1/4}} \leq t\beta\frac{3l}{2R}(\epsilon_{t-1})^{(lR^3)^{1/4}}$$

$$\leq t\beta\frac{3l}{2R}(\frac{1}{2^t})^{(lR^3)^{1/4}} = \frac{t\beta\frac{3l}{R}}{2^{t((lR^3)^{1/4}-1)}} \times \frac{1}{2^{t+1}}.$$

The proof is complete if $t\beta\frac{3l}{R} < 2^{t(\sqrt[4]{lR^3}-1)}$ for $t \geq 1$. It is clear that this is true for l large enough. ∎

Now consider (ii). Clearly for l large enough such that

$$\delta_2 = \exp\left(-\frac{l}{5R}\right) + \frac{l}{R}(2\sqrt{\delta_1})^{l-1} \leq 2\exp\left(-\frac{l}{5R}\right).$$

To complete the proof by induction, we remark that $\delta_t < 2\exp\left(-\frac{l}{5R}t\right)) < 1$ implies

$$\frac{l}{R}(2\sqrt{\delta_t})^{l-1} < \exp\left(-\frac{l}{5R}(t+1)\right)$$

for l large enough independent of t.

We now turn to the proof of Theorem 2. It is organized in three steps:

- 1) We first show that for any small δ_1 and ϵ_1, one can find an integer l_1 such that for $l \geq l_1$

$$p_1 \equiv \mathbb{P}\left\{|h^{(1)}| \leq \frac{\epsilon_1}{\beta}\right\} \geq 1 - \delta_1.$$

- 2) We then show by induction on $t \geq 1$ that

$$p_t \equiv \mathbb{P}\left\{|h^{(t)}| < \frac{\epsilon_t}{\beta}\right\} \geq 1 - \delta_t.$$

- 3) Finally using Lemma 2 we deduce that $h^{(t)} \to 0$ almost surely as $t \to +\infty$.

Proof: [Proof of Theorem 2] ∎

We begin by noting that regardless of the initial distribution, $\widehat{q}^{(t)}(\widehat{h})$ has a symmetric density due to the symmetric distribution of J. Moreover, $|\widehat{h}^{(t)}| \leq 1$ from (40). Thus, $\mathbb{E}_{\widehat{q}^{(t)}}(\widehat{h}^{(t)}) = 0$ and $\text{Var}(\widehat{h}^{(t)}) = \mathbb{E}_{\widehat{q}^{(t)}}(\widehat{h}^2) \leq 1$.

Step 1: We set $P(r) = e^{-\lambda}\frac{\lambda^r}{r!}$ and $\lambda = l/R$. Let $h^{(r,t)} = \sum_{a=1}^{r} \widehat{h}_a^{(t)}$ where $\widehat{h}_a^{(t)}$ are i.i.d random variables with probability density $\widehat{q}^{(t)}(\widehat{h})$. Let $\sigma_0^2 = \mathbb{E}((\widehat{h}_a^{(0)})^2) \leq 1$. According to [47, Th. 3.5.3] we have

$$\lim_{r \to \infty} \sqrt{r}\mathbb{P}\left\{|h^{(r,0)}| < \frac{\epsilon_0}{\beta}\right\} = \frac{2\epsilon_0}{\beta\sqrt{2\pi\sigma_0^2}},$$

for any $\epsilon_0 > 0$. Thus, there exists $r'(\epsilon_0, \beta) \in \mathbb{N}$ such that for $r > r'$,

$$\mathbb{P}\left\{|h^{(r,0)}| < \frac{\epsilon_0}{\beta}\right\} \geq \frac{\epsilon_0}{\beta\sqrt{2\pi r}}.$$

Take l such that $\lambda = l/R \geq l'/R = 2r'$, then

$$p_0 = \mathbb{P}\left\{|h^{(0)}| < \frac{\epsilon_0}{\beta}\right\}$$

$$= \sum_{r=0}^{\infty} P(r)\mathbb{P}\left\{|h^{(r,0)}| < \frac{\epsilon_0}{\beta}\right\}$$

$$\geq \sum_{r=\lambda/2}^{3\lambda/2} P(r)\mathbb{P}\left\{|h^{(r,0)}| < \frac{\epsilon_0}{\beta}\right\}$$

$$\geq \frac{\epsilon_0}{\beta\sqrt{3\pi\lambda}} \sum_{r=\lambda/2}^{3\lambda/2} P(r)$$

$$> \frac{\epsilon_0}{\beta\sqrt{3\pi\lambda}}(1 - 2e^{-\frac{\lambda}{10}}).$$

The last inequality follows from lemma 1. Thus for l large enough

$$p_0 = \mathbb{P}\left\{|h^{(0)}| < \frac{\epsilon_0}{\beta}\right\} > \frac{\epsilon_0}{2\beta\sqrt{3\pi\lambda}} \equiv 1 - \delta_0. \quad (103)$$

Recall $\widehat{h}^{(t+1)} = \frac{1}{\beta}\tanh^{-1}\left(J\tanh\beta \prod_{i=1}^{l-1}\tanh\beta h_i^{(t)}\right)$. From $\tanh^{-1}(a\tanh\beta) \leq a\beta$ for $0 < a < 1$, we have

$$\left|\widehat{h}^{(t+1)}\right| \leq \prod_{i=1}^{l-1}\tanh\left|\beta h_i^{(t)}\right|.$$

Define

$$Z_l^{(t)} \equiv \ln\left(\prod_{i=1}^{l-1}\tanh\left|\beta h_i^{(t)}\right|\right) = \sum_{i=1}^{l-1}\ln\left(\tanh\left|\beta h_i^{(t)}\right|\right).$$

Note that $Z_l^{(t)}$ is always negative and if one of $h_i^{(t)}$ tends to zero, it diverges to $-\infty$. Consider $t = 0$. We will show that $Z_l^{(0)}$ has a large negative value with high probability. Define

$$u_i \equiv \begin{cases} u_{i-1}, & \text{if } \left|h_{i-1}^{(0)}\right| > \frac{\epsilon_0}{\beta}, \\ u_{i-1} + \ln\tanh\epsilon_0, & \text{otherwise}, \end{cases}$$

with $u_0 = 0$. One can check for later use that $Z_l^{(0)} \leq u_l$. Moreover, because of (103) one can consider u_l as a random walk (with negative jumps),

$$u_i = \begin{cases} u_{i-1}, & \text{with prob. } 1 - p_0 \\ u_{i-1} + \ln\tanh\epsilon_0, & \text{with prob. } p_0. \end{cases}$$

Let $s = \ln(\tanh(\epsilon_0))$. Using the Chernoff's theorem [48, p. 151],

$$\mathbb{P}\left\{\frac{1}{l-1}\frac{u_l}{s} < \lambda^{-3/4}\right\} < \exp\left(-(l-1)D(\lambda^{-3/4}\|p_0)\right),$$

where $D(x\|y) = x\ln(\frac{x}{y}) + (1-x)\ln(\frac{1-x}{1-y})$. Now, since

$$x\ln\left(\frac{x}{p_0}\right) > x\ln(x),$$

$$(1-x)\ln\left(\frac{1-x}{1-p_0}\right) > (1-x)\ln\left(\frac{1-x}{\delta_0}\right),$$

we have

$$D(\lambda^{-3/4}\|p_0) > -H_2(\lambda^{-3/4})\ln(2) - (1-\lambda^{-3/4})\ln(\delta_0), \quad (104)$$

for δ_0 defined in (103). By a large λ expansion of the right hand side of (104):

$$-H_2(\lambda^{-3/4})\ln 2 - (1-\lambda^{-3/4})\ln\delta_0$$
$$= \frac{\epsilon_0}{2\beta\sqrt{3\pi\lambda}} + o\left(\frac{1}{\sqrt{\lambda}}\right).$$

Thus, there exists $l'' \in \mathbb{N}$ depending on R, β and ϵ_0 such that for $l > l''$,

$$\mathbb{P}\left\{\frac{1}{l-1}\frac{u_l}{s} < \lambda^{-3/4}\right\} < \exp\left(-\frac{\epsilon_0(l-1)}{4\beta\sqrt{3\pi\lambda}}\right). \quad (105)$$

By replacing $s = \ln\tanh\epsilon_0$ and $\lambda = \frac{l}{R} \approx \frac{l-1}{R}$ for large degrees,

$$\mathbb{P}\left\{u_l > (lR^3)^{1/4}\ln\tanh\epsilon_0\right\} < \exp\left(-\frac{\epsilon_0\sqrt{Rl}}{4\beta\sqrt{3\pi}}\right),$$

Note that the inequality in $\mathbb{P}(\dots)$ is reversed since $s < 0$. Now recall $Z_l^{(0)} \leq u_l$. Therefore,

$$\mathbb{P}\left\{Z_l^{(0)} \leq (lR^3)^{1/4}\ln\tanh\epsilon_0\right\}$$

$$\geq \mathbb{P}\left\{u_l \leq (lR^3)^{1/4}\ln\tanh\epsilon_0\right\} \geq 1 - \exp\left(-\frac{\epsilon_0\sqrt{Rl}}{4\beta\sqrt{3\pi}}\right).$$

Consequently,

$$\mathbb{P}\left\{\left|\widehat{h}^{(1)}\right| \leq (\tanh\epsilon_0)^{(lR^3)^{1/4}}\right\}$$

$$\geq \mathbb{P}\left\{Z_l^{(0)} \leq (lR^3)^{1/4}\ln\tanh\epsilon_0\right\} \geq 1 - \exp\left(-\frac{\epsilon_0\sqrt{Rl}}{4\beta\sqrt{3\pi}}\right).$$

From r, $|h^{(r,1)}| = \left|\sum_{a=1}^{r}\widehat{h}_a^{(1)}\right| \leq \sum_{a=1}^{r}\left|\widehat{h}_a^{(1)}\right|$. we deduce

$$\mathbb{P}\left\{\left|h^{(r,1)}\right| \leq r(\tanh\epsilon_0)^{(lR^3)^{1/4}}\right\}$$

$$\geq \mathbb{P}\left\{\left|\widehat{h}^{(1)}\right| \leq (\tanh\epsilon_0)^{(lR^3)^{1/4}}\right\}^r$$

$$\geq \left\{1 - \exp\left(-\frac{\epsilon_0\sqrt{Rl}}{4\beta\sqrt{3\pi}}\right)\right\}^r$$

$$\geq 1 - r\exp\left(-\frac{\epsilon_0\sqrt{Rl}}{4\beta\sqrt{3\pi}}\right).$$

for l large enough. Therefore we have,

$$\mathbb{P}\left\{\left|h^{(1)}\right| \leq \frac{3}{2}\lambda(\tanh\epsilon_0)^{(lR^3)^{1/4}}\right\}$$

$$= \sum_{r=0}^{\infty} P(r)\mathbb{P}\left\{\left|h^{(r,1)}\right| \leq \frac{3}{2}\lambda(\tanh\epsilon_0)^{(lR^3)^{1/4}}\right\}$$

$$\geq \sum_{r=0}^{3\lambda/2} P(r)\mathbb{P}\left\{\left|h^{(r,1)}\right| \leq \frac{3}{2}\lambda(\tanh\epsilon_0)^{(lR^3)^{1/4}}\right\}$$

$$\geq \sum_{r=0}^{3\lambda/2} P(r)\mathbb{P}\left\{\left|h^{(r,1)}\right| \leq r(\tanh\epsilon_0)^{(lR^3)^{1/4}}\right\}$$

$$\geq \sum_{r=0}^{3\lambda/2} P(r)\left(1 - r\exp\left(-\frac{\epsilon_0\sqrt{Rl}}{4\beta\sqrt{3\pi}}\right)\right)$$

$$\geq 1 - \exp(-0.1\lambda) - \lambda\exp\left(-\frac{\epsilon_0\sqrt{Rl}}{4\beta\sqrt{3\pi}}\right).$$

To summarize, we have obtained

$$p_1 = \mathbb{P}\left\{ \left| h^{(1)} \right| \leq \frac{\epsilon_1}{\beta} \right\} \geq 1 - \delta_1. \tag{106}$$

This completes step 1.

Step 2: The proof is by induction. Assume that

$$p_t = \mathbb{P}\left\{ |h^{(t)}| \leq \frac{\epsilon_t}{\beta} \right\} \geq 1 - \delta_t.$$

We prove that this holds also for $t + 1$. This mainly consists in repeating the derivations (103) to (106) for p_t, ϵ_t and δ_t. We briefly repeat them here:

$$\mathbb{P}\left\{ \left| \widehat{h}^{(t+1)} \right| \leq (\tanh \epsilon_t)^{(lR^3)^{1/4}} \right\}$$
$$\geq \mathbb{P}\left\{ Z_l^{(t)} \leq (lR^3)^{1/4} \ln (\tanh \epsilon_t) \right\}$$
$$\geq 1 - \exp\left(-(l-1)D(\lambda^{-3/4} \| p_t) \right).$$

Assume that $\delta_t \ll 1$. From (104),

$$D(\lambda^{-3/4} \| p_t) > -H_2(\lambda^{-3/4}) \ln(2) - (1 - \lambda^{-3/4}) \ln (\delta_t).$$

If $\lambda^{-3/4} < \frac{1}{2}$ (equivalently, $l > 2^{4/3}R$),

$$D(\lambda^{-3/4} \| p_t) > -\ln 2 - \frac{1}{2} \ln \delta_t.$$

Thus,

$$\mathbb{P}\left\{ \left| \widehat{h}^{(t+1)} \right| \leq (\tanh \epsilon_t)^{(lR^3)^{1/4}} \right\} \geq 1 - (2\sqrt{\delta_t})^{l-1},$$

and finally,

$$\mathbb{P}\left\{ \left| \widehat{h}^{(t+1)} \right| \leq (t+1)\frac{3}{2}\lambda(\tanh \epsilon_t)^{(lR^3)^{1/4}} \right\}$$
$$\geq \sum_{r=0}^{3(t+1)\lambda/2} P(r)\mathbb{P}\left\{ \left| h^{(r,t+1)} \right| \leq (t+1)\frac{3}{2}\lambda(\tanh \epsilon_t)^{(lR^3)^{1/4}} \right\}$$
$$\geq \sum_{r=0}^{3(t+1)\lambda/2} P(r)\mathbb{P}\left\{ \left| h^{(r,t+1)} \right| \leq r(\tanh \epsilon_t)^{(lR^3)^{1/4}} \right\}$$
$$\geq \sum_{r=0}^{3(t+1)\lambda/2} P(r) \left(1 - r(2\sqrt{\delta_t})^{l-1} \right)$$
$$\geq 1 - \exp(-(t+1)\frac{\lambda}{10}) - \lambda(2\sqrt{\delta_t})^{l-1}.$$

Or equivalently,

$$p_{t+1} = \mathbb{P}\left\{ |h^{(t+1)}| < \frac{\epsilon_{t+1}}{\beta} \right\} \geq 1 - \delta_{t+1}.$$

This completes step 2.

Step 3: Using lemma 2, for l large enough (depending on β and R, but independent of t)

$$\mathbb{P}\left\{ \left| h^{(t)} \right| > \frac{1}{\beta 2^{(t+1)}} \right\} \leq \delta_t \leq 2 \exp\left(-\frac{l}{5R}t \right).$$

The Borel-Cantelli lemma [47, Th. 2.3.1] states that, $h^{(t)} \to 0$ almost surely if for all $\alpha > 0$,

$$\sum_{t=1}^{\infty} \mathbb{P}\left\{ \left| h^{(t)} \right| > \alpha \right\} < +\infty.$$

Let us verify that $h^{(t)}$ has this property. For any α, there is τ such that $1/2^{\tau+1} < \beta\alpha$. Therefore, for $t \geq \tau$,

$$\mathbb{P}\left\{ \left| h^{(t)} \right| > \epsilon \right\} \leq \mathbb{P}\left\{ \left| h^{(t)} \right| > \frac{1}{2^{(t+1)}\beta} \right\} < \delta_t$$

and hence,

$$\sum_{t=1}^{\infty} \mathbb{P}\left\{ \left| h^{(t)} \right| > \epsilon \right\} \leq \tau + \sum_{t=\tau}^{\infty} \mathbb{P}\left\{ \left| h^{(t)} \right| > \epsilon \right\}$$
$$< \tau + \sum_{t=\tau}^{\infty} \delta_t$$
$$< \tau + \sum_{t=\tau}^{\infty} 2 \exp\left(-\frac{l}{10R}t \right) < +\infty.$$

This completes step 3.

APPENDIX E
PROOF OF THEOREM 3

Proof: We first show the property (i). Note that it is satisfied by $\widehat{q}^{+(0)}$ and $q^{+(0)}$. The equations (47) and (48) are density evolution equations an LDGM ensemble on the BSC. In [49], It is known that (i) is preserved under density evolution recursions (see [49] for similar properties in the case of LDPC codes). ∎

Let us turn to the proof of (ii). First note that (60) implies (61). Indeed

$$\mathbb{P}\{\widehat{\eta}^{(t)} < -1 + \epsilon\} = \int_{-1}^{-1+\epsilon} \widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta}$$
$$= \int_{1-\epsilon}^{1} e^{-2\beta\widehat{\eta}}\widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta}$$
$$\geq e^{-2\beta}\mathbb{P}\{\widehat{\eta}^{(t)} > 1 - \epsilon\}$$
$$\geq \frac{1}{1 + e^{2\beta}}(1 - \delta).$$

So we only have to prove (60). We will use induction. The induction hypothesis is (60) for some $\delta > 0$ and $\epsilon > 0$ at iteration t. It is obviously true at $t = 0$.

Let us first show that

$$\mathbb{E}(\eta^{(t)}) = \lambda\mathbb{E}(\widehat{\eta}^{(t)}) \geq 2\lambda s. \tag{107}$$

for $s = \frac{1}{2}(1 - \delta)(1 - \epsilon)(1 - e^{-2\beta(1-\epsilon)})/(1 + e^{-2\beta})$. We have

$$\mathbb{E}(\widehat{\eta}^{(t)}) = \int_{-1}^{1} \widehat{\eta}\widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta}$$
$$= \int_{-1}^{0} \widehat{\eta}\widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta} + \int_{0}^{1} \widehat{\eta}\widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta}$$
$$= -\int_{0}^{1} \widehat{\eta}e^{-2\beta\widehat{\eta}}\widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta} + \int_{0}^{1} \widehat{\eta}\widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta}$$
$$= \int_{0}^{1} \widehat{\eta}(1 - e^{-2\beta\widehat{\eta}})\widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta}$$
$$\geq \int_{1-\epsilon}^{1} \widehat{\eta}(1 - e^{-2\beta\widehat{\eta}})\widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta}$$
$$\geq (1 - e^{-2\beta(1-\epsilon)})(1 - \epsilon)\int_{1-\epsilon}^{1} \widehat{q}^{+(t)}(\widehat{\eta})d\widehat{\eta}$$
$$> (1 - \delta)(1 - \epsilon)\frac{1 - e^{-2\beta(1-\epsilon)}}{1 + e^{-2\beta}}.$$

This proves (107).

By applying Hoeffding's inequality [48] for $\lambda/2 < r < 3\lambda/2$,

$$\mathbb{P}\left\{\sum_{a=1}^{r}\widehat{\eta}_a^{(t)} < \lambda\frac{s}{2}\right\}$$

$$= \mathbb{P}\left\{\sum_{a=1}^{r}(\widehat{\eta}_a^{(t)} - \mathbb{E}(\widehat{\eta}^{(t)})) < \lambda\frac{s}{2} - r\mathbb{E}(\widehat{\eta}^{(t)})\right\}$$

$$\leq \mathbb{P}\left\{\sum_{a=1}^{r}(\widehat{\eta}_a^{(t)} - \mathbb{E}(\widehat{\eta}^{(t)})) < \lambda\frac{s}{2} - 2rs\right\}$$

$$\leq \mathbb{P}\left\{\sum_{a=1}^{r}(\widehat{\eta}_a^{(t)} - \mathbb{E}(\widehat{\eta}^{(t)})) < -\lambda\frac{s}{2}\right\}$$

$$< \exp(-\frac{\lambda^2 s^2}{8r})$$

$$< \exp(-\lambda\frac{s^2}{12}).$$

From

$$\mathbb{P}\left\{\eta^{(t)} < \lambda\frac{s}{2}\right\} = \sum_{r=0}^{\infty} P(r)\mathbb{P}\left\{\sum_{a=1}^{r}\widehat{\eta}_a^{(t)} < \lambda\frac{s}{2}\right\}$$

$$\leq \sum_{r=0}^{\lambda/2} P(r) + \sum_{r=\lambda/2}^{3\lambda/2} P(r)\mathbb{P}\left\{\sum_{a=1}^{r}\widehat{\eta}_a^{(t)} < \lambda\frac{s}{2}\right\}$$

$$+ \sum_{r=3\lambda/2}^{\infty} P(r).$$

and Lemma 1, we get

$$\mathbb{P}\left\{\eta^{(t)} > \lambda\frac{s}{2}\right\} > 1 - 2\exp\left(-\frac{\lambda}{10}\right) - \exp(-\lambda\frac{s^2}{12}). \quad (108)$$

Now consider the density evolution equation (48). We have

$$\mathbb{P}\left\{\widehat{\eta}^{(t+1)} > \frac{1}{\beta}\mathrm{atanh}\left(\tanh(\beta)\left[\tanh(\beta\lambda\frac{s}{2})\right]^{l-1}\right)\right\}$$

$$\geq \mathbb{P}\left\{J = 1, \eta_1^{(t)} > \frac{\lambda s}{2}, \ldots, \eta_{l-1}^{(t)} > \frac{\lambda s}{2}\right\}$$

$$= \frac{1 + \tanh(\beta)}{2}\left(\mathbb{P}\left\{\eta^{(t)} > \frac{\lambda s}{2}\right\}\right)^{l-1}$$

$$\geq \frac{1 + \tanh(\beta)}{2}\left(1 - 2\exp\left(-\frac{\lambda}{10}\right) - \exp(-\lambda\frac{s^2}{12})\right)^{l-1}$$

$$\geq \frac{e^{2\beta}}{1 + e^{2\beta}}\left(1 - (l-1)\left(2\exp(-\frac{l}{10R}) + \exp(-\frac{ls^2}{12R})\right)\right)$$

Let

$$1 - \varepsilon(l, R, \beta) = \frac{1}{\beta}\mathrm{atanh}\left(\tanh(\beta)\left[\tanh(\beta s\frac{l}{2R})\right]^{l-1}\right),$$

$$\Delta(l, R) = (l-1)\left(2\exp(-0.1\frac{l}{R}) + \exp(-\frac{ls^2}{12R})\right).$$

Inequality (60) holds at $t+1$, if $\varepsilon(l, R, \beta) \leq \epsilon$ and $\Delta(l, R) \leq \delta$. This is true for $l > l_1$ large enough since $\varepsilon(l, R, \beta)$ and $\Delta(l, R)$ are decreasing functions of l (for large values of l).

REFERENCES

[1] T. J. Goblick, "Coding for a discrete information source with a distortion measure," Ph.D. dissertation, Dept. Elect. Eng., MIT, Cambridge, MA, USA, 1963.

[2] A. J. Viterbi and J. Omura, "Trellis encoding of memoryless discrete-time sources with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 325–332, May 1974.

[3] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

[4] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[5] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.

[6] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[7] E. Martinian and J. S. Yedidia, "Iterative quantization using codes on graph," in *Proc. 41st Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2003.

[8] T. Murayama, "Thouless–Anderson–Palmer approach for lossy compression," *Phys. Rev. E*, vol. 69, p. 035105, Mar. 2004. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.69.035105

[9] S. Ciliberti and M. Mézard, "The theoretical capacity of the parity source coder," *J. Statist. Mech., Theory Experim.*, vol. 2005, no. 10, p. P10003, Oct. 2005. [Online]. Available: http://stacks.iop.org/1742-5468/2005/i=10/a=P10003

[10] M. J. Wainwright, E. Maneva, and E. Martinian, "Lossy source compression using low-density generator matrix codes: Analysis and algorithms," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1351–1368, Mar. 2010.

[11] T. Filler and J. Fridrich, "Binary quantization using belief propagation with decimation over factor graphs of LDGM codes," in *Proc. 45th Allerton Conf. Coding, Commun., Control*, Monticello, IL, USA, Sep. 2007.

[12] D. Castanheira and A. Gameiro, "Lossy source coding using belief propagation and soft-decimation over LDGM codes," in *Proc. IEEE 21st Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2010, pp. 431–436.

[13] A. J. Felstrom and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2181–2191, Sep. 1999.

[14] M. Lentmaier, A. Sridharan, D. J. Costello, Jr., and K. S. Zigangirov, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5274–5289, Oct. 2010.

[15] M. Lentmaier, A. Sridharan, K. S. Zigangirov, and D. J. Costello, Jr., "Terminated LDPC convolutional codes with thresholds close to capacity," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Sep. 2005, pp. 1372–1376.

[16] M. Lentmaier, D. G. M. Mitchell, G. P. Fettweis, and D. J. Costello, Jr., "Asymptotically regular LDPC codes with linear distance growth and thresholds close to capacity," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Jan./Feb. 2010, pp. 1–8.

[17] V. Aref and R. L. Urbanke, "Universal rateless codes from coupled LT codes," in *Proc. Inf. Theory Workshop (ITW)*, Oct. 2011, pp. 277–281.

[18] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[19] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec. 2013.

[20] A. Yedla, Y.-Y. Jian, P. S. Nguyen, and H. D. Pfister, "A simple proof of threshold saturation for coupled scalar recursions," in *Proc. 7th Int. Symp. Turbo Codes Iterative Inf. Process. (ISTC)*, Aug. 2012, pp. 51–55.

[21] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "A proof of threshold saturation for spatially-coupled LDPC codes on BMS channels," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2012, pp. 176–184.

[22] S. H. Hassani, N. Macris, and R. L. Urbanke, "Coupled graphical models and their thresholds," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Aug./Sep. 2010, pp. 1–5.

[23] S. H. Hassani, N. Macris, and R. L. Urbanke, "Chains of mean-field models," *J. Statist. Mech., Theory Experim.*, vol. 2012, no. 2, p. P02011, Feb. 2012. [Online]. Available: http://stacks.iop.org/1742-5468/2012/i=02/a=P02011

[24] S. H. Hassani, N. Macris, and R. L. Urbanke, "Threshold saturation in spatially coupled constraint satisfaction problems," *J. Statist. Phys.*, vol. 150, no. 5, pp. 807–850, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10955-012-0664-x

[25] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep./Oct. 2010, pp. 347–353.

[26] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, p. 021005, May 2012. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevX.2.021005

[27] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.

[28] V. Aref, N. Macris, R. L. Urbanke, and M. Vuffray, "Lossy source coding via spatially coupled LDGM ensembles," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 373–377.

[29] V. Aref, N. Macris, and M. Vuffray, "Approaching the rate-distortion limit by spatial coupling with belief propagation and decimation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1177–1181.

[30] M. Mézard and A. Montanari, *Information, Physics, and Computation*. London, U.K.: Oxford Univ. Press, 2009.

[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.

[32] S. Ciliberti, M. Mézard, and R. Zecchina, "Message-passing algorithms for non-linear nodes and data compression," *Complexus*, vol. 3, nos. 1–3, pp. 58–65, 2006.

[33] A. Giurgiu, N. Macris, and R. L. Urbanke, "And now to something completely different: Spatial coupling as a proof technique," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 2443–2447.

[34] M. Mézard and G. Parisi, "The Bethe lattice spin glass revisited," *Eur. Phys. J. B, Condens. Matter Complex Syst.*, vol. 20, no. 2, pp. 217–233, 2001.

[35] S. Franz and M. Leone, "Replica bounds for optimization problems and diluted spin systems," *J. Statist. Phys.*, vol. 111, nos. 3–4, pp. 535–564, 2003.

[36] F. Guerra and F. L. Toninelli, "The thermodynamic limit in mean field spin glass models," *Commun. Math. Phys.*, vol. 230, no. 1, pp. 71–79, 2002.

[37] A. R. Iyengar, P. H. Siegel, R. L. Urbanke, and J. K. Wolf, "Windowed decoding of spatially coupled codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2277–2292, Apr. 2013.

[38] N. Ul Hassan, A. E. Pusane, M. Lentmaier, G. P. Fettweis, and D. J. Costello, Jr., "Reduced complexity window decoding schedules for coupled LDPC codes," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2012, pp. 20–24.

[39] H.-O. Georgii, *Gibbs Measures and Phase Transitions*, vol. 9. Berlin, Germany: Walter de Gruyter, 2011.

[40] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, "Clusters of solutions and replica symmetry breaking in random *k*-satisfiability," *J. Statist. Mech., Theory Experim.*, vol. 2008, no. 4, p. P04004, Apr. 2008. [Online]. Available: http://stacks.iop.org/1742-5468/2008/i=04/a=P04004

[41] M. Mézard and A. Montanari, "Reconstruction on trees and spin glass transition," *J. Statist. Phys.*, vol. 124, no. 6, pp. 1317–1350, 2006. [Online]. Available: http://dx.doi.org/10.1007/s10955-006-9162-3

[42] F. Guerra and F. L. Toninelli, "The high temperature region of the Viana–Bray diluted spin glass model," *J. Statist. Phys.*, vol. 115, nos. 1–2, pp. 531–555, 2004. [Online]. Available: http://dx.doi.org/10.1023/B%3AJOSS.0000019815.11115.54

[43] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Berlin, Germany: Springer-Verlag, 2001.

[44] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "Threshold saturation for spatially coupled LDPC and LDGM codes on BMS channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7389–7415, Dec. 2014.

[45] F. Ricci-Tersenghi and G. Semerjian, "On the cavity method for decimated random constraint satisfaction problems and the analysis of belief propagation guided decimation algorithms," *J. Statist. Mech., Theory Experim.*, vol. 2009, p. P09001, Sep. 2009.

[46] A. Coja-Oghlan, "On belief propagation guided decimation for random *k*-SAT," in *Proc. 22nd Annu. ACM-SIAM Symp. Discrete Algorithms*, 2011, pp. 957–966.

[47] R. Durrett, *Probability: Theory and Examples*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[48] P. Billingsley, *Probability and Measure*. New York, NY, USA: Wiley, 1995.

[49] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

Vahid Aref earned his B.Sc. and M.Sc. in electrical engineering from Sharif University of Technology, Iran, in 2006 and 2008. He received his PhD degree in computer and communication sciences from École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in March 2014. He was a research assistant in the communication theory laboratory (LTHC), EPFL, from 2010 till 2014. Until January 2015, he conducted postdoctoral research at University of Stuttgart in collaboration with Bell Laboratories, Alcatel-lucent, Stuttgart, Germany. Since then, he is a researcher at Bell Laboratories, Alcatel-lucent, Stuttgart, Germany. His fields of interest include coding and information theory, optical communications theory, and the theory and applications of graphical models.

Nicolas Macris (M'04) received a PhD degree in physics from the Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland. He spent two years with the Department of Mathematics in Rutgers University, NJ, and then the Institute for Theoretical Physics in EPFL. During that period, he held frequent visiting appointments in the Ecole Normale Superieure de Lyon, the Centre de Physique Theorique in Luminy Marseille and the University College Dublin. He spent a year in ETH Zuerich, the Universities of Fribourg and Neuchatel and then joined the School of Computer and Communication Science of EPFL. His research in mathematical physics has centered upon Coulomb systems, quantum spin systems, random magnetic Schroedinger operators and the quantum Hall effect. Currently he is working on graphical models at large with emphasis on coding theory, constraint satisfaction models and their connections with statistical mechanics of disordered systems.

Marc Vuffray received his M.Sc. in physics in 2008 from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He was a research assistant between 2008 and 2009 in the department of applied mathematics of EPFL and a PhD student between 2009 and 2014 in the communication theory laboratory of EPFL. He received his Ph.D. in computer and communication sciences in 2014. Since 2014 he conducts post-doctoral research in the theoretical division and center for nonlinear studies at the Los Alamos National Laboratory in New Mexico, USA. His research in applied mathematics led him to study non-linear and non-local Schrödinger operators. During his Ph.D. he worked on graphical models in coding theory and their connections with statistical physics of spin glasses. He his currently working on optimization problems that arise in natural gas networks and smart grids.