Approaching the Rate-Distortion Limit by Spatial Coupling with Belief Propagation and Decimation

Vahid Aref, Nicolas Macris and Marc Vuffray
LTHC, IC, EPFL, Lausanne, Switzerland

***Abstract*—We investigate an encoding scheme for lossy compression based on spatially coupled Low-Density Generator-Matrix codes. The degree distributions are regular, or are Poisson on the code-bit side and check-regular which allows use for any compression rate. The performance of a low complexity Belief Propagation Guided Decimation algorithm is excellent, and for large check degrees it gets close to Shannon's rate-distortion limit. We investigate links between the algorithmic performance and the phase diagram of a relevant random Gibbs measure. The associated dynamical and condensation thresholds are computed within the framework of the cavity method. We observe that: (i) the dynamical threshold of the spatially coupled construction saturates towards the condensation threshold; (ii) for large degrees the condensation threshold approaches the information theoretic test-channel parameter of rate-distortion theory. This provides heuristic insight into the excellent performance of the BPGD algorithm.**

***Index Terms*—Lossy source coding, spatial coupling, LDGM, belief propagation guided decimation, rate distortion bound.**

I. Introduction

Lossy source coding is one of the fundamental problems in communications. The use of linear codes for encoding is almost as old as the problem itself, and it has been established long ago that the rate-distortion limit for binary sources can be achieved using linear codes [1]. All considerations in this paper are in this same framework: the source is binary Bernoulli, the distortion is measured by Hamming distance and we use linear codes. Designing a low (linear) complexity encoder that achieves the Shannon limit has been an open problem for a long time. Polar Codes solve the problem with a low complexity encoder which has the great advantage of being mathematically tractable in terms of analysis [2].

Another interesting direction uses Low-Density Generator-Matrix (LDGM) codes, for which Belief Propagation (BP) and/or Survey Propagation (SP) equipped with a decimation process yield low complexity encoding schemes. For naive choices (say regular, or check-regular) of degree distributions the rate-distortion limit *is not* approached. However it has been observed that *it is* approached (but not achieved) for degree distributions that have been optimized for channel LDPC coding [3–5]. 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 distributions.

Note however, that it is proven by second moment methods that check-regular LDGM based codes achieve the rate-distortion limit under optimal encoding [3] in the limit of large degrees. This has also been observed by using cavity and/or replica methods of statistical physics [6], [7].

Spatially coupled codes were first introduced in the context of channel coding in the form of convolutional LDPC codes [8]. It has been recognized and proved recently that *simple* spatially coupled graphical constructions are able to universally achieve capacity under BP decoding (see [9] and references therein). As the width of the coupling window increases, the BP threshold *saturates* towards the maximum posterior MAP threshold, the later being close to capacity for large degrees.

In this contribution we investigate a *simple* spatially coupled LDGM construction: the degree distributions that we consider are either regular or Poisson on the code-bit side, and are regular on the check side. We explore a low complexity encoding Belief Propagation Guided Decimation (BPGD) algorithm, that takes advantage of spatial coupling, and approaches the rate-distortion limit for large check degrees and any compression rate. In [10] we already provided preliminary evidence of this good performance for regular degree constructions. Besides testing a rate-universal construction, we provide links between the algorithmic performance and the phase diagram of the Gibbs measure underlying the BPGD algorithm. It is known from work on constraint satisfaction problems [11] that the performance of BPGD algorithms is affected by so-called dynamical and condensation thresholds predicted by the cavity theory. We show here that the dynamical threshold *saturates* towards the condensation threshold. This is analogous to threshold saturation in channel coding and provides insight into the good performance of BPGD on coupled constructions for lossy source coding.

II. Spatially Coupled LDGM Ensembles for Lossy Compression

Let $\underline{X} = \{X_1, X_2, \ldots, X_N\}$ be a source of length N, where $X_a \in \{0,1\}$, $a = 1, \ldots, N$ are i.i.d Bernoulli$(1/2)$. 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. The decoding operation maps the stored word $\underline{u}$ to a reconstructed sequence $\widehat{\underline{x}}(\underline{u}) \in \{0,1\}^N$. We measure the distortion by the relative Hamming distance $d_N(\underline{x}, \widehat{\underline{x}}) = \frac{1}{N}\sum_{a=1}^{N} |x_a - \widehat{x}_a|$. The quality of reconstruction is defined as the average distortion $\mathbb{E}_{\underline{X}}[d_N(\underline{x}, \widehat{\underline{x}})]$. For any encoding-decoding scheme, the (limsup) average distortion is lower bounded by Shannon's rate-distortion curve $D_{\mathrm{sh}}(R) = h_2^{-1}(1-R)$ where $h_2 : [0, \frac{1}{2}] \to [0,1]$ is the binary entropy function.

Let us introduce the LDGM graphical constructions used here. The *underlying Poisson LDGM*(l, R) *ensemble* contains bipartite graphs $\Gamma(C, V; E)$ with C a set of n check nodes of constant degree l, V a set of m code-bit nodes of variable degree, and E the set of edges connecting C and V. The graphs are generated by connecting each edge emanating from a check node uniformly at random to a code-bit node. In the large block length limit, with $m/n = R$, the code-bit node degrees are i.i.d Poisson distributed with an average degree l/R.

To construct the *Spatially Coupled LDGM*(l, R, L, w, n) *ensemble* we first lay out a spatial dimension with positions indexed by $z \in \mathbb{Z}$. Each position $z \in \{1, \ldots, L\}$ contains n check and m code-bit nodes, and there are also m extra code-bit nodes at each position $z \in \{L+1, \ldots, L+w-1\}$. All checks at $z \in \{1, \ldots, L\}$ have l emanating edges that are connected uniformly at random to code-bit nodes within the range $\{z, \ldots, z+w-1\}$. It is easy to see that for $z \in \{w, \ldots, L\}$ the code-bit nodes have (asymptotically) Poisson degrees with average l/R, while for the boundary positions $z \in \{1, \ldots, w-1\}$ the average is $l/R \times z/w$ and similarly for $z \in \{L+1, \ldots, L+w-1\}$. See [12] for further details.

We will also use *regular ensembles* with check node degree l and code-bit node degree r. The underlying ensemble LDGM(l, r) is given by the usual construction. For the detailed construction of the coupled ensemble LDGM(l, r, L, w, n) we refer to [9].

We use letters a, b, c for check nodes and i, j, k for code-bit nodes; we use ∂a for the set of all code-bit nodes connected to a; ∂i for the set of all check nodes connected to i.

We can now describe the *decoding rule* and the *optimal encoder*. To each code-bit node $i \in V$ we "attach" a code bit u_i, and to each check node $a \in C$ we "attach" the reconstructed bit $\widehat{x}_a$ and the source bit x_a. By definition the source sequence has length N. Thus, we must take $n = N$ 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 $m/n = R$ for the underlying ensemble, and is $R_{\text{true}} = m(L+w-1)/nL = R(1+\frac{w-1}{L})$ for the coupled ensemble. The compression rate of the coupled ensemble is slightly larger than R, but in the asymptotic regime $n, m >> L >> w$ this difference vanishes.

Decoding Rule: The reconstruction mapping is given by the linear operation $\widehat{x}_a(\underline{u}) = \oplus_{i \in \partial a} u_i$.

Optimal Encoding: Given source word $\underline{x}$, the optimal encoder seeks to minimize the Hamming distortion, and so searches among all $\underline{u} \in \{0,1\}^{NR}$ a configuration $\underline{u}^*$ such that $\underline{u}^* = \text{argmin}_{\underline{u}} d_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))$. The resulting distortion is $d_{N,\min} = d_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))$. The "optimal distortion" of the ensemble is $D_{N,\text{opt}} = \mathbb{E}_{\text{LDGM},\underline{X}}[d_{N,\min}]$ where $\mathbb{E}_{\text{LDGM},\underline{X}}$ is an expectation over the graphical ensemble and the symmetric Bernoulli source $\underline{X}$.

In this paper we use methods and concepts from *statistical mechanics*. We equip the configuration space $\{0,1\}^{NR}$ with the probability measure,

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

where $\beta > 0$ is a real number, and $Z_\beta(\underline{x})$ is a normalizing factor or partition function. Equation (1) defines the Gibbs measure of a random spin system with random Hamiltonian $2Nd(\underline{x}, \widehat{\underline{x}}(\underline{u}))$ at inverse temperature β. Finding $\underline{u}^*$ amounts to finding the ground state configuration with ground state energy (per node) $2d_{N,\min}$. It is easy to check the identity $2d_{N,\min} = -N^{-1} \lim_{\beta\to\infty} \beta^{-1} \ln Z_\beta(\underline{x})$. As this identity shows, a fundamental role is played by the average *free energy*, $f_N(\beta) = -\beta^{-1} N^{-1} \mathbb{E}_{\underline{X},\Gamma}[\ln Z_\beta(\underline{x})]$, the knowledge of which allows to obtain the average optimal distortion $D_{N,\text{opt}} = \lim_{\beta\to+\infty} f_N(\beta)/2$. We also use another useful relationship between average distortion and free energy. Consider the *internal energy* $u_N(\beta) = \mathbb{E}_{\text{LDGM},\underline{X}}[\langle 2d_N(\underline{x}, \widehat{\underline{x}}(\underline{u}))\rangle]$ where the bracket $\langle - \rangle$ denotes the average over $\underline{u}$ with respect to (1). It is straightforward to check that the internal energy can be computed from the free energy $u_N(\beta) = \frac{\partial}{\partial\beta}(\beta f_N(\beta))$; and that in the zero temperature limit it reduces to the average ground state energy or optimal distortion $D_{N,\text{opt}} = \lim_{\beta\to+\infty} u_N(\beta)/2$.

What is the relation between the quantities $f_N(\beta)$, $u_N(\beta)$, and $D_{N,\text{opt}}$ for the underlying and coupled ensembles? Consider the infinite block length limit defined as $\lim_{N\to+\infty} = \lim_{n\to+\infty}$ for the underlying ensemble, and as $\lim_{N\to+\infty} = \lim_{L\to+\infty} \lim_{n\to+\infty}$ for the coupled ensemble, with m/n fixed. In [13] it is proved that for Poisson ensembles with even l, $\lim_{N\to+\infty} f_N(\beta)$, $\lim_{N\to+\infty} u_N(\beta)$ and $\lim_{N\to+\infty} D_{N,\text{opt}}$ exist and are the same for the two ensembles. We expect this to be also valid for general degree distributions. From now on we drop the subscript N when we refer to asymptotic quantities.

It is conjectured that the "one-step-replica-symmetry-breaking-formulas" (1RSB) obtained from the cavity theory [11] yield exact values for the $N \to +\infty$ limit of the free, internal and ground state energies. As an illustrative example, for $(l, r) = (3, 6)$; $(5, 10)$, the 1RSB formulas yield $D_{\text{opt}} = 0.1139$; 0.1105 [6]. The gap to the Shannon limit $D_{\text{sh}}(1/2) = 0.1100$ decreases to zero as one increases the degrees.

III. Belief Propagation Guided Decimation

We describe a low complexity encoding rule based on ***BP estimates*** of marginals of the Gibbs measure (1).

The BP equations are a set of fixed point equations involving $2|E|$ real valued messages $\eta_{i\to a}$ and $\widehat{\eta}_{a\to i}$ attached to the edges $(i, a) \in E$ of the graph,

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

A BP estimate of the marginal is entirely characterized by a single bias $\eta_i = \sum_{a\in\partial i} \widehat{\eta}_{a\to i}$. Unfortunately, the number of solutions of the BP equations which lead to roughly the same minimal distortion grows exponentially large in terms of N, and it is not possible to find the relevant fixed point by a

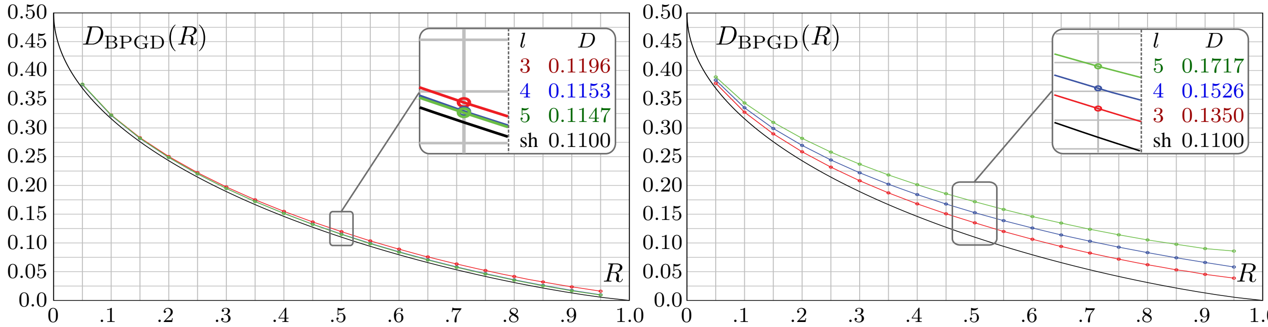

Fig. 1. BPGD distortion versus rate R compared to the Shannon rate-distortion limit (lower curve). Left: spatially coupled $(l, R, L = 64, w = 3, n = 2000)$ ensembles for $l = 3$, $l = 4$, $l = 5$. Right: $(l, R, n = 128000)$ ensembles for $l = 3$, $l = 4$, $l = 5$. When plotted in terms of $R_{\mathrm{true}} = R(1 + \frac{w-1}{L})$ these curves are increased by an amount $\frac{w-1}{L} R |\frac{\partial D(R)}{\partial R}|$ to leading order. At $R = 1/2$ and $L = 64$, $w = 3$ this amount is approximately 0.005.

plain iterative method. To get around this problem, the BP iterations are used in conjunction with a heuristic decimation process that we describe below.

Let Γ, $\underline{x}$ be a graph and source instance. Fix an initial set of messages $\eta^0_{i\to a}$ at time $t = 0$. Iterating the BP equations we get a set of messages $\eta^t_{i\to a}$, $\widehat{\eta}^t_{a\to i}$ at time $t \geq 0$. Let $\epsilon > 0$ be some "small" positive number and T some "large" time. We define a *decimation instant* t_{dec} as follows: (i) *ϵ- convergence*: if the total variation of messages does not change significantly in two successive iterations, $\frac{1}{|E|}\sum_{(i,a)\in E} |\widehat{\eta}^t_{a\to i} - \widehat{\eta}^{t-1}_{a\to i}| < \epsilon$ for some $t \leq T$, then set $t_{\mathrm{dec}} = t$; (ii) if ϵ-convergence does not occur for all $t \leq T$ then set $t_{\mathrm{dec}} = T$.

At time t_{dec} each code-bit has a BP marginal with a bias $\eta_i^{t_{\mathrm{dec}}}$. We *select* and *fix one particular code-bit* $u_{i_{\mathrm{dec}}}$ according to a decision rule $u_{i_{\mathrm{dec}}} \leftarrow \mathcal{D}(\underline{\eta}^{t_{\mathrm{dec}}}) \in \{0, 1\}$ described later. At this point we update $x_a \leftarrow x_a \oplus u_{i_{\mathrm{dec}}}$ for all $a \in \partial i_{\mathrm{dec}}$ and decimate the graph $\Gamma \leftarrow \Gamma \setminus i$. This defines a new graph, a new 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} in the previous round.

To choose i_{dec} we first evaluate the *maximum bias* $B_{t_{\mathrm{dec}}} = \max_{i\in V} |\eta_i^{t_{\mathrm{dec}}}|$. When $B_{t_{\mathrm{dec}}} > 0$, we consider the set of maximizers and choose i_{dec} u.a.r among them. When $B_{t_{\mathrm{dec}}} = 0$, in the case of the *underlying ensemble*, we choose i_{dec} u.a.r from $\{1, \ldots m\}$. When $B_{t_{\mathrm{dec}}} = 0$, in the case of the *coupled ensemble*, we choose i_{dec} u.a.r from the w left-most positions of the current graph.

We adopt a *hard decision rule*

$$\mathcal{D}(\underline{\eta}^{t_{\mathrm{dec}}}) = \begin{cases} \theta(\eta_{i_{\mathrm{dec}}}^{t_{\mathrm{dec}}}), & \text{if } B_{t_{\mathrm{dec}}} > 0 \\ \mathrm{Bernoulli}(\frac{1}{2}), & \text{if } B_{t_{\mathrm{dec}}} = 0 \end{cases} \tag{2}$$

where $\theta(.)$ is the Heaviside step function. A related randomized decision rule is discussed in section VI.

We have checked that if we choose i_{dec} at random from the whole chain (for coupled graphs) the performance is not improved by coupling. It is important to adopt a schedule that seeds the encoding process at the left boundary where the code-bit nodes are less constrained. In [10] we adopted periodic boundary condition so the seeding region was 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. The pseudo-code for the present BPGD algorithm is similar to the one given in [10].

There are two parameters ϵ and T which in practice are set to $\epsilon = 0.01$, and $T = 10$. The simulation results do not change significantly when we take ϵ smaller and T larger. They do depend on the choice of β which enters in the BP equations, and one may optimize on β. We do not have a first principle theory for the optimal choice β^*. We will see in section V that for spatially coupled ensembles one can take β^* close to the condensation threshold of (1). In the *large degree limit* this rule coincides with the information theoretic guess related to the log-likelihood of the flip probability of a BSC test-channel, namely $\beta^* \approx \frac{1}{2}\ln(1 - D_{\mathrm{sh}}(R))/D_{\mathrm{sh}}(R)$.

It is not difficult to see that the complexity of the plain BPGD algorithm is $O(N^2T)$, in other words $O(n^2T)$ for underlying and $O(n^2L^2T)$ for coupled ensembles. By updating BP messages only in a sliding window of width $O(w)$ (as in window decoding [14, 15] for spatial LDPC codes) we reduce the complexity to $O(n^2LT)$. This can be further reduced to $O(nLT/\delta)$ by decimating δn code-bits at each step for some small δ. For $\delta \approx 0.02$ we do not observe a significant loss in performance.

IV. BPGD Performance: Simulations

Fig. 1 displays the average distortion $D_{\mathrm{BPGD}}(R)$ obtained by the BPGD algorithm as a function of R for the Poisson ensembles, and compares it to the Shannon limit $D_{\mathrm{sh}}(R)$ given by the lowest (black) 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). The plot on the right is for the underlying ensembles with $l = 3, 4, 5$ and $n = 128000$. We observe that as the degree increases the BPGD performance gets worse. But recall that, with increasing degrees, the optimal distortion of the ensemble (not shown explicitly) gets better and approaches the Shannon limit [3]. Thus the situation is similar as for the BP threshold of LDPC codes which gets worse with increasing degrees, while the MAP threshold approaches Shannon capacity. The plot on the left shows 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 the simplest possible BPGD scheme.

Theoretical arguments, and simulations, suggest the following. Look at the regime $n >> L >> w >> 1$. When these parameters go to infinity in the specified order *for the coupled ensemble* $D_{\rm BPGD}(R)$ approaches $D_{\rm opt}(R)$. When $l \to \infty$ after the other parameters $D_{\rm BPGD}(R)$ approaches $D_{\rm sh}(R)$.

Our empirical observations show that the optimal β^* is related to the convergence properties of the BP iterations along the decimation path. We run BPGD over 50 instances and count the fraction of decimation steps that results from the ϵ-convergence condition being satisfied in less than $T = 10$ BP updates. This fraction defines an *empirical probability of convergence* $C_{\epsilon,T}(\beta)$. We observe that, for $\beta < \beta^*$ the empirical probability of convergence is close to 1, while it drops markedly for $\beta > \beta^*$. This behavior is more marked for the coupled ensemble.

We come back to other interesting observations about β^* in the next section.

V. Predictions of the Cavity Method

It is natural to expect that the behavior of BP based algorithms should in a way or another be related to the phase diagram of the Gibbs measure (1). As we vary β, the geometry of the space of "typical configurations" displays transitions at special *dynamical* and *condensation* thresholds β_d and β_c. Let us briefly explain what these thresholds are. Assume that the random Gibbs measure can, in the limit of $N \to +\infty$, be decomposed into a convex superposition of "pure states" (or extremal measures)

$$\mu_\beta(\underline{u} \mid \underline{x}) = \sum_{p=1}^{\mathcal{N}} w_p \, \mu_{\beta,p}(\underline{u} \mid \underline{x}) \tag{3}$$

each of which occurs with a weight $w_p = e^{-\beta N(f_p - f)}$, where f_p is the free energy of pure state p and f the total free energy (normalized by $1/N$). It is furthermore assumed that for $N \to +\infty$ these free energies concentrate with respect to the graph and source instance. In the decomposition (3) the number of pure states that effectively contribute to the measure depends on a balance between their free energies and their numbers. A standard argument shows that $f = \min_\varphi(\varphi - \beta^{-1}\Sigma(\varphi))$ where $\exp(N\Sigma(\varphi))$ is the number of pure states with $f_p \approx \varphi$. The exponential growth rate $\Sigma(\varphi)$ is called the *complexity*. By inverting the Legendre transform one obtains the complexity $\Sigma(\beta)$ as a function of β. The cavity method makes a number of remarkable predictions. For $\beta < \beta_d$ the Gibbs measure is pure in the sense that $\mathcal{N} = 1$ and $\Sigma(\beta) = 0$. For $\beta_d < \beta < \beta_c$ an exponentially large number of pure states contribute to the measure: $\mathcal{N} \approx e^{N\Sigma(\beta)}$, and $\Sigma(\beta) > 0$. For $\beta > \beta_c$ only a finite number of pure states contribute to the measure and $\Sigma(\beta) = 0$: the measure is *condensed* over a small number of pure states. In practice we compute the complexity and read from it the values of 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. Moreover it is known that BP marginals are not correct above the condensation threshold β_c. At β_d the free and internal energies have no singularities. This is a dynamical transition threshold: Markov chain Monte Carlo dynamics have an equilibration time that diverges when $\beta \uparrow \beta_d$, and are unable to sample the Gibbs measure in a finite time for $\beta > \beta_d$. for more details we refer to [11].

We have computed the complexity and β_d, β_c from the cavity theory. The details of this theory are much to complicated to be presented here, and we refer to [16] for the version of the cavity method followed here. Let us just say that for $\beta < \beta_c$ the equations of the cavity method adapted to the present problem simplify drastically and become very similar to density evolution equations. These are then solved numerically. The table displays results for the regular ensembles LDGM$(5, 10)$ and LDGM$(5, 10, L, w)$. The column

L	β	w				
		1	2	3	4	5
64	β_d	0.828	1.034	1.038	1.039	1.040
	β_c	1.041	1.051	1.059	1.066	1.074
128	β_d	0.828	1.033	1.037	1.038	1.039
	β_c	1.041	1.046	1.051	1.053	1.057
256	β_d	0.828	1.031	1.035	1.037	1.038
	β_c	1.041	1.043	1.045	1.047	1.048

$w = 1$ gives the dynamical and condensation thresholds of the underlying *uncoupled* ensemble, $\beta_d(w = 1)$ and $\beta_c(w = 1)$. We observe that the condensation threshold $\beta_c(L, w)$ of the coupled ensemble is higher than $\beta_c(w = 1)$. This is due to a finite length effect that disappears as L increases. Since the free energies of the coupled and underlying ensembles are the same in the limit of infinite length and the condensation threshold is a singularity of the free energy [13],

$$\lim_{L\to+\infty} \beta_c(L, w) = \beta_c(w = 1). \tag{4}$$

Concerning the dynamical threshold, we see that for each fixed L it increases as a function of w. Closer inspection suggests that

$$\lim_{w\to+\infty} \lim_{L\to+\infty} \beta_d(L, w) = \beta_c(w = 1). \tag{5}$$

This relation can also be partially justified on theoretical grounds, by an analysis of the (simplified) cavity equations by potential methods similar to [17]. Equ. 5 indicates a threshold saturation phenomenon: for the coupled ensemble the phase of non-zero complexity shrinks to zero (and the condensed phase remains unchanged, at least threshold-wise). Similar observations have been discussed for constraint satisfaction problems in [13].

We systematically observe that the optimal algorithmic value β^* is *always lower but somewhat close* to β_d. For example for the uncoupled case $(l, r) = (5, 10)$ we have

$(\beta^*, \beta_d) \approx (0.71, 0.828)$. 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 $\beta^* \approx \beta_d \approx \beta_c$. Thus for the coupled ensemble BPGD operates well even close to the condensation threshold. We use this fact in the next section to explain the good performance of the algorithm for coupled instances.

VI. Discussion

According to the information theoretic approach to rate-distortion theory, we can view the optimal encoding problem, as a decoding problem for a random linear code on a BSC(p) test-channel with $p = D_{\rm sh}(R)$. Now, the Gibbs measure (1) 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 encoder with an optimal β^*, approaches the rate-distortion limit, we should have

$$\beta^* \approx \frac{1}{2}\ln\frac{1-p}{p} \equiv \frac{1}{2}\ln\frac{1-D_{\rm sh}(R)}{D_{\rm sh}(R)}. \tag{6}$$

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

$$\beta_c \approx \frac{1}{2}\ln\frac{1-D_{\rm sh}(R)}{D_{\rm sh}(R)}, \quad \text{for } l >> 1. \tag{7}$$

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

It is interesting to investigate the range of β for which BP correctly computes marginals of (1) for coupled instances. For this purpose, we run BPGD with a *randomized decision rule* (called BPGD-r)

$$\mathcal{D}(\underline{\eta}^{t_{\rm dec}}) = \begin{cases} 0, & \text{with prob } \frac{1}{2}(1-\tanh\beta\eta_{i_{\rm dec}}^{t_{\rm dec}}) \\ 1, & \text{with prob } \frac{1}{2}(1+\tanh\beta\eta_{i_{\rm dec}}^{t_{\rm dec}}). \end{cases} \tag{8}$$

If BP is effective at computing correct marginals, then BPGD-r should correctly sample the Gibbs measure; and in that case the average distortion $D_{\rm BPGD-r}(\beta)$ should be equal to $u(\beta)/2$. On the other hand, the cavity method predicts the simple expression $u(\beta)/2 = (1-\tanh\beta)/2$ for $\beta < \beta_c$. In Fig. 2 we observe $D_{\rm BPGD-r}(\beta) \approx (1-\tanh\beta)/2$ 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. 2 suggests that for coupled instances BPGD-r correctly samples the Gibbs measure of coupled instances all the way up to $\approx \beta_c$, and that BP correctly computes marginals for the same range.

These properties of BPGD-r suggest to use it as an algorithm for encoding. It turns out that with spatial coupling the resulting performance curves are similar to the ones in Fig. 1 of BPGD with hard decision rule. With $\beta^* \approx \beta_c$ there is not much difference between the hard and randomized decision rules.

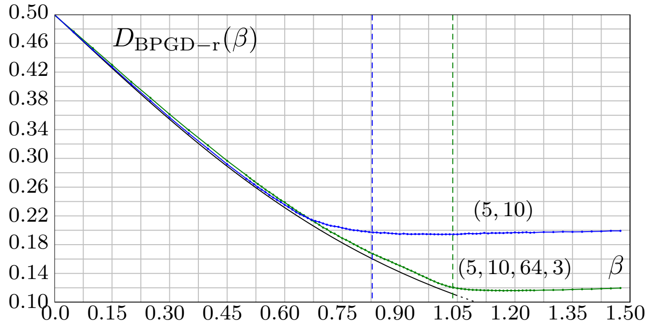

Fig. 2. Internal energy computed by BPGD with randomized rounding for LDGM$(5, 10, 64, 3, 2000)$ and LDGM$(5, 10, 128000)$ ensembles. The lower (black) curve is the internal energy $(1-\tanh\beta)/2$ for $\beta < \beta_c = 1.041$. Vertical dotted lines are $\beta_d = 0.828$ (underlying) and $\beta_d = 1.038$ (coupled).

Acknowledgment. We thank R. Urbanke for many useful discussions and comments on the work presented here. V.A. was supported by grant No. 200021-125347, and M.V. by grant No. 200020-140388 of the Swiss National Science Foundation.

References

[1] T. J. Goblick, "Coding for discrete information source with a distortion measure," Ph.D. dissertation, MIT, 1963.

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

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

[4] T. Filler and J. Fridrich, "Binary quantization using belief propagation with decimation over factor graphs of ldgm codes," in *Proc. Allerton Conf.*, 2007.

[5] D. Castanheira and A. Gameiro, "Lossy source coding using belief propagation and soft-decimation over ldgm codes," in *PIMRC*, 2010, pp. 431 –436.

[6] S. Ciliberti and M. Mézard, "The theoretical capacity of the parity source coder," *JSTAT*, p. P10003, 2005.

[7] T. Murayama, "Thouless-anderson-palmer approach for lossy compression," *Phys. Rev. E*, vol. 69, no. 3, p. 035105, 2004.

[8] A. J. Felstrom and K. S. Zigangirov, "Time-varying periodic convolutional codes with low density parity check matrix," *IEEE Trans. Inf. Th.*, vol. 45, no. 5, pp. 2181–2190, 1999.

[9] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially Coupled Ensembles Universally Achieve Capacity under Belief Propagation," *ArXiv*, Jan. 2012.

[10] V. Aref, N. Macris, R. Urbanke, and M. Vuffray, "Lossy source coding via spatially coupled ldgm ensembles," in *ISIT*, 2012, pp. 373–377.

[11] M. Mézard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[12] V. Aref and R. L. Urbanke, "Universal rateless codes from coupled lt codes," in *ITW*, oct. 2011, pp. 277 –281.

[13] S. H. Hassani, N. Macris, and R. L. Urbanke, "Threshold saturation in spatially coupled constraint satisfaction problems," *J. Stat. Phys*, pp. 1–44, 2012.

[14] A. Iyengar, P. Siegel, R. Urbanke, and J. Wolf, "Windowed decoding of spatially coupled codes," in *ISIT*, 2011, pp. 2552–2556.

[15] N. ul Hassan, A. E. Pusane, M. Lentmaier, G. P. Fettweis, and D. J. Costello, "Reduced complexity window decoding schedules for coupled ldpc codes," in *ITW*, 2012, pp. 20 –24.

[16] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, "Clusters of solutions and replica symmetry breaking in random k-satisfiability," *JSTAT*, p. P04004, 2008.

[17] S. Kumar, A. Young, N. Macris, and H. Pfister, "A proof of threshold saturation for irregular ldpc codes on bms channels," *Proc. Allerton Conf.*, 2012.