Dynamic phase transition for decoding algorithms

Silvio Franz*
International Center for Theoretical Physics, P.O. Box 586, I-34100 Trieste, Italy

Michele Leone†
International Center for Theoretical Physics, INFM and SISSA, via Beirut 8, I-34100 Trieste, Italy

Andrea Montanari‡
Laboratoire de Physique Théorique de l'Ecole Normale Supérieure,§ 24, rue Lhomond, 75231 Paris Cedex 05, France

Federico Ricci-Tersenghi‖
Dipartimento di Fisica and SMC and UdR1 of INFM, Università di Roma "La Sapienza," Piazzale Aldo Moro 2, I-00185 Roma, Italy

(Received 3 May 2002; revised manuscript received 5 August 2002; published 16 October 2002)

The state-of-the-art error correcting codes are based on large random constructions (random graphs, random permutations, etc.) and are decoded by linear-time iterative algorithms. Because of these features, they are remarkable examples of diluted mean-field spin glasses, both from the static and dynamic points of view. We analyze the behavior of decoding algorithms by mapping them onto statistical-physics models. This allows us to understand the intrinsic (i.e., algorithm independent) features of this behavior.

DOI: 10.1103/PhysRevE.66.046120 PACS number(s): 89.70.+c, 75.50.Lk

I. INTRODUCTION

Recently there has been some interest in studying complexity phase transitions, i.e., abrupt changes in the computational complexity of hard combinatorial problems as some control parameter is varied [1]. These phenomena are thought to be somehow related to the physics of glassy systems, where the physical dynamics experiences a dramatic slowing down as the temperature is lowered [2].

Complexity is a central issue also in coding theory [3,4]. Coding theory [5–7] deals with the problem of communicating information reliably through an unreliable channel of communication. This task is accomplished by making use of *error correcting codes*. In 1948 Shannon [8] proved that almost any error correcting code allows to communicate without errors, as long as the rate of transmitted information is kept below the *capacity* of the channel. However decoding is an intractable problem for almost any code. Coding theory is therefore a rich source of interesting computational problems.

On the other hand it is known that error correcting codes can be mapped onto disordered spin models [9–13]. Remarkably there has recently been a revolution in coding theory which has lead to the invention of new and very powerful codes based on random constructions: turbo codes [14], low density parity check codes (LDPCC) [15,16], repetition accumulated codes [17], etc. As a matter of fact, equivalent spin models have been intensively studied in the last few years. These are diluted spin glasses, i.e., spin glasses on random (hyper)graphs [18–21].

The new codes are decoded by using approximate iterative algorithms, which are closely related to the cavity approach to mean-field spin glasses [22,23]. We think therefore that a close investigation of these systems from a statistical physics point of view, having in mind complexity (i.e., dynamical) issues, can be of great theoretical interest.[1]

Let us briefly recall the general setting of coding theory [5] in order to fix a few notations (cf. Fig. 1 for a pictorial description). A source of information produces a stream of symbols. Let us assume, for instance, that the source produces unbiased random bits. The stream is partitioned into *blocks* of length N_{block}. Each of the possible $2^{N_{\mathrm{block}}}$ blocks is mapped to a *code word* (i.e., a sequence of bits) of length $N > N_{\mathrm{block}}$ by the *encoder* and transmitted through the channel. An error correcting code is therefore defined either as a mapping $\{0,1\}^{N_{\mathrm{block}}} \to \{0,1\}^N$, or as a list of $2^{N_{\mathrm{block}}}$ code words. The *rate* of the code is defined as $R = N_{\mathrm{block}}/N$.

Let us denote[2] the transmitted code word by $\underline{\mathbf{x}}^{\mathrm{in}} = [\mathbf{x}_1^{\mathrm{in}}, \ldots, \mathbf{x}_N^{\mathrm{in}}]^T$. Due to the noise, a different sequence of symbols $\underline{\mathbf{x}}^{\mathrm{out}} = [\mathbf{x}_1^{\mathrm{out}}, \ldots, \mathbf{x}_N^{\mathrm{out}}]^T$ is received. The decoding problem is to infer $\underline{\mathbf{x}}^{\mathrm{in}}$, given $\underline{\mathbf{x}}^{\mathrm{out}}$, the definition of the code, and the properties of the noisy channel.

It is useful to summarize the general picture which

*Email address: franz@ictp.trieste.it

†Email address: micleone@ictp.trieste.it

‡Email address: Andrea.Montanari@lpt.ens.fr

§UMR 8549, Unité Mixte de Recherche du Centre National de la Recherche Scientifique et de l' Ecole Normale Supérieure.

‖Email address: Federico.Ricci@roma1.infn.it

[1] The reader is urged to consult Refs. [24–33] for a statistical mechanics analysis of the optimal decoding (i.e., of static issues).

[2] We shall denote transmitted and received symbols by typographic characters, with the exception of symbols in $\{+1,-1\}$. In this case we use the physicists notation and denote such symbols by σ. When considering binary symbols we will often pass from the $\mathbf{x}$ notation to the σ notation, the correspondence $\sigma = (-1)^{\mathbf{x}}$ being understood. Finally vectors of length N will be always denoted by underlined characters: e.g., $\underline{\mathbf{x}}$ or $\underline{\sigma}$.

1063-651X/2002/66(4)/046120(17)/$20.00

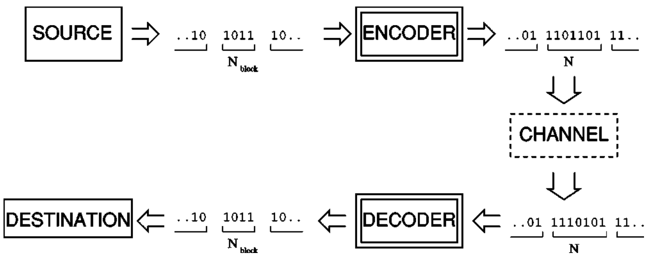

FIG. 1. A schematic description of how error correcting codes work.

emerges from our work. We shall focus on Gallager codes (both *regular* and *irregular*). The optimal decoding strategy (maximum-likelihood decoding) is able to recover the transmitted message below some noise threshold: $p<p_c$. Iterative, linear-time algorithms get stuck (in general) at a lower noise level, and are successful only for $p<p_d^{(\mathrm{alg.})}$, with $p_d^{(\mathrm{alg.})}\leqslant p_c$. In general, the "dynamical" threshold $p_d^{(\mathrm{alg.})}$ depends upon the details of the algorithm. However, it seems to be always smaller than some universal (although code-dependent) value p_d. Moreover, some "optimal" linear-time algorithms are successful up to p_d [i.e., $p_d^{(\mathrm{alg.})}=p_d$]. The universal threshold p_d coincides with the dynamical transition [2] of the corresponding spin model.

The plan of the paper is the following. In Sec. II we introduce LDPCC, focusing on Gallager's *ensembles*, and we describe *message-passing* decoding algorithms. We briefly recall the connection between these algorithms and the cavity equations for mean-field spin glasses. In Sec. III we define a spin model which describes the decoding problem, and introduce the replica formalism. In Sec. IV we analyze this model for a particular choice of the noisy channel (the *binary erasure channel*). In this case calculations can be fully explicit and the results are particularly clear. Then, in Sec. V, we address the general case. Finally we draw our conclusions in Sec. VI. The Appendixes collect some details of our computations.

II. ERROR CORRECTING CODES, DECODING ALGORITHMS, AND CAVITY EQUATIONS

This section introduces the reader to some basic terminology in coding theory. In the first part we define some *ensembles* of codes, namely, *regular* and *irregular* LDPCC. In the second one we describe a class of iterative decoding algorithms. These algorithms have a very clear physical interpretation, which we briefly recall. Finally we explain how these algorithms are analyzed in the coding theory community. This section does not contain any original result. The interested reader may consult Refs. [7,15,23,34] for further details.

A. Encoding

Low density parity check codes are defined by assigning a binary $N\times M$ matrix $\mathbb{H}=\{H_{ij}\}$, with $H_{ij}\in\{0,1\}$. All the code words are required to satisfy the constraint

$$\mathbb{H}\underline{\mathbf{x}}=0(\mathrm{mod}2). \tag{2.1}$$

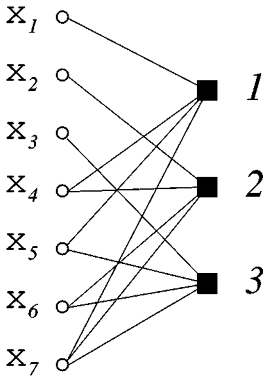

FIG. 2. The Tanner graph for a simple code with $N=7$, $M=3$. The code words satisfy the three parity check equations $\mathbf{x}_1+\mathbf{x}_4+\mathbf{x}_5+\mathbf{x}_7=0$, $\mathbf{x}_2+\mathbf{x}_4+\mathbf{x}_6+\mathbf{x}_7=0$, $\mathbf{x}_3+\mathbf{x}_5+\mathbf{x}_6+\mathbf{x}_7=0$ (mod2).

The matrix $\mathbb{H}$ is called the *parity check matrix* and the M equations summarized in Eq. (2.1) are the *parity check equations* (or, for short, *parity checks*). If the matrix $\mathbb{H}$ has rank M (this is usually the case), the rate is $R=1-M/N$.

There exists a nice graphic representation of Eq. (2.1) which is often used in the coding theory community: the *Tanner graph* representation [35,36]. One constructs a bipartite graph by associating a left-hand node with each one of the N variables, and a right-hand node with each one of the M parity checks. An edge is drawn between the *variable* node i and the parity check node α if and only if the variable $\mathbf{x}_i$ appears with a nonzero coefficient in the parity check equation α. We refer to Fig. 2 for a simple example.

In general, one considers ensembles of codes by defining a random construction of the parity check matrix. One of the simplest ensembles is given by regular (k,l) Gallager codes. In this case one chooses the matrix $\mathbb{H}$ randomly among all the $N\times M$ matrices having k nonzero entries per row, and l per column.

Amazingly good codes [37–39] were obtained by slightly more sophisticated irregular constructions. In this case one assigns the distributions of the degrees of parity check nodes and variable nodes in the Tanner graph. We shall denote by $\{c_k\}$ the degree distribution of the check nodes and by $\{v_l\}$ the degree distribution of the variable nodes. This means that there are Nv_l bits of the code word belonging to l parity checks and Nc_k parity checks involving k bits for each k and l. We shall always assume $c_k=0$ for $k<3$ and $v_l=0$ for $l<2$

It is useful to define the generating polynomials

$$c(x)\equiv\sum_{k=3}^{\infty}c_kx^k,\quad v(x)\equiv\sum_{l=2}^{\infty}v_lx^l, \tag{2.2}$$

which satisfy the normalization condition $c(1)=v(1)=1$. Moreover, we define the average variable and check degrees $\overline{l}=v'(1)$ and $\overline{k}=c'(1)$. Particular examples of this formalism are the regular codes whose generating polynomials are $c(x)=x^k$, $v(x)=x^l$.

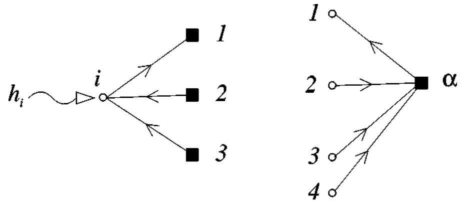

FIG. 3. A graphic representation of the operations executed in a message-passing algorithm. At the variable node i (on the left): $x^{(t+1)}_{i\to 1}=F(y^{(t)}_{2\to i},y^{(t)}_{3\to i};h_i)$. At the check node α (on the right): $y^{(t+1)}_{\alpha\to 1}=G(x^{(t)}_{2\to\alpha},x^{(t)}_{3\to\alpha},x^{(t)}_{4\to\alpha})$.

B. Decoding

The code words are transmitted through a noisy channel. We assume antipodal signalling: one sends $\sigma^{\text{in}}\in\{+1,-1\}$ signals instead of $x^{\text{in}}\in\{0,1\}$ through the channel [the correspondence being given by $\sigma=(-1)^{\mathbf{x}}$]. At the end of the channel, a corrupted version of this signals is received. This means that if $\sigma^{\text{in}}\in\{+1,-1\}$ is transmitted, the value $\mathbf{x}^{\text{out}}$ is received with probability density $Q(\mathbf{x}^{\text{out}}|\sigma^{\text{in}})$. The information conveyed by the received signal $\mathbf{x}^{\text{out}}$ is conveniently described by the log likelihood:[3]

$$h(\mathbf{x}^{\text{out}})=\frac{1}{2}\ln\frac{Q(\mathbf{x}^{\text{out}}|+1)}{Q(\mathbf{x}^{\text{out}}|-1)}. \tag{2.3}$$

We can represent this information by wavy lines in the Tanner graph.

The decoding problem is to compute the probability for each transmitted bit σ_i^{in} to take the value σ_i, given the structure of the code and the received message $\underline{\mathbf{x}}^{\text{out}}=[\mathbf{x}_1^{\text{out}},\ldots,\mathbf{x}_N^{\text{out}}]^T$. This is, in general, an intractable problem [3,4]. Recently there has been a great interest in dealing with this problem using approximate message-passing algorithms.

Message-passing algorithms are iterative: at each step t one keeps track of $M\overline{k}$ messages from the variable nodes to the check nodes $\{y^{(t)}_{\alpha\to i}\}$ and vice versa $\{x^{(t)}_{i\to\alpha}\}$. Messages can be thought to travel along the edges and computations to be executed at the nodes. A node computes the message to be sent along each one of the edges, using the messages received from the other edges at the previous iteration [the variable nodes also make use of the log likelihoods $h(\mathbf{x}_i^{\text{out}})$], cf. Fig. 3. At some point the iteration is stopped (there exists no general stopping criterion), and a choice for the bit σ_i is taken using all the incoming messages [plus the log-likelihood $h(\mathbf{x}_i^{\text{out}})$].

The functions that define the "new" messages in terms of the "old" ones can be chosen to optimize the decoder performances. A particularly interesting family is the following:

$$x^{(t+1)}_{i\to\alpha}=h_i+\sum_{\alpha'\ni i:\alpha'\neq\alpha} y^{(t)}_{\alpha'\to i}, \tag{2.4}$$

$$y^{(t+1)}_{\alpha\to i}=\frac{1}{\zeta}\text{arctanh}\left[\prod_{j\in\alpha:j\neq i}\tanh\zeta x^{(t)}_{j\to\alpha}\right], \tag{2.5}$$

where we used the notation $i\in\alpha$ whenever the bit i belongs to the parity check α. The messages $\{x^{(\cdot)}_{i\to\alpha}\}$ and $\{y^{(\cdot)}_{\alpha\to i}\}$ can be rescaled in such a way to eliminate the parameter ζ everywhere except in front of h_i. Therefore ζ allows to tune the importance given to the information contained in the received message.

After the convergence of the above iteration one computes the *a posteriori* log likelihoods as follows:

$$H_i=h_i+\sum_{\alpha\ni i} y^{(\infty)}_{\alpha\to i}. \tag{2.6}$$

The meaning of $\{H_i\}$ is analogous to that of $\{h_i\}$ (but for the fact that the H_i incorporate the information coming from the structure of the code): the best guess for the bit i is $\sigma_i=+1$ or $\sigma_i=-1$, depending whether $H_i>0$ or $H_i<0$.

The most popular choice for the free parameter ζ is $\zeta=1$: this algorithm has been invented separately by Gallager [15] in the coding theory context (and named the *sum-product* algorithm) and by Pearl [40] in the artificial intelligence context (and named the *belief propagation* algorithm). Also $\zeta=\infty$ is sometimes used (the *max-product* algorithm).

The alert reader will notice that Eqs. (2.4) and (2.5) are nothing but the cavity equations at inverse temperature ζ for a properly constructed spin model. This remark is the object of Refs. [22,41].

In the analysis of the above algorithm it is convenient to assume that $\sigma_i^{\text{in}}=+1$ for $i=1,\ldots,N$. This assumption can be made without loss of generality if the channel is symmetric [i.e., if $Q(x|+1)=Q(-x|-1)$]. With this assumption, the h_i are i.i.d. random variables with density

$$p(h)\equiv Q(\mathbf{x}(h)|+1)|\mathbf{x}'(h)|, \tag{2.7}$$

where $\mathbf{x}(h)$ is the function which inverts Eq. (2.3). In the following we shall consider two particular examples of noisy channels, the generalization being straightforward.

(1) The binary erasure channel (BEC). In this case a bit can either be received correctly or erased.[4] There are therefore three possible outputs: $\{+1,-1,0\}$. The transition probability is

$$Q(\mathbf{x}^{\text{out}}|+1)=\begin{cases}(1-p) & \text{if } \mathbf{x}^{\text{out}}=+1,\\ p & \text{if } \mathbf{x}^{\text{out}}=0,\\ 0 & \text{if } \mathbf{x}^{\text{out}}=-1,\end{cases}$$

$$Q(\mathbf{x}^{\text{out}}|-1)=\begin{cases}0 & \text{if } \mathbf{x}^{\text{out}}=+1,\\ p & \text{if } \mathbf{x}^{\text{out}}=0,\\ (1-p) & \text{if } \mathbf{x}^{\text{out}}=-1.\end{cases} \tag{2.8}$$

[3] Notice the unconventional normalization: the factor 1/2 is inserted to make it consistent with the statistical mechanics formulation.

[4] This is what happens, for instance, to packets in the Internet traffic.

We get therefore the following distribution for the log likelihoods: $p(h)=(1-p)\delta_\infty(h)+p\,\delta(h)$ (where δ_∞ is a Dirac delta function centered at $+\infty$). Let us recall that the capacity of the BEC is given by $C_{BEC}=1-p$: this means that a rate-R code cannot assure error correction if $p>1-R$. (2) The binary symmetric channel (BSC). The channel flips each bit independently with probability p. Namely,

$$Q(\mathbf{x}^{\text{out}}|+1)=\begin{cases}(1-p) & \text{if } \mathbf{x}^{\text{out}}=+1,\\ p & \text{if } \mathbf{x}^{\text{out}}=-1,\end{cases}$$

$$Q(\mathbf{x}^{\text{out}}|-1)=\begin{cases}p & \text{if } \mathbf{x}^{\text{out}}=+1,\\ (1-p) & \text{if } \mathbf{x}^{\text{out}}=-1.\end{cases} \tag{2.9}$$

The corresponding log-likelihood distribution is $p(h)=(1-p)\delta(h-h_0)+p\,\delta(h+h_0)$, with $h_0=\text{arctanh}(1-2p)$. The capacity of the BSC is[5] $C_{BSC}=1-h(p)$: a rate-R code cannot correct errors if $p>\delta_{GV}(R)$.

It is quite easy [34,42] to write a recursive equations for the probability distributions of the messages $\pi_t(x)$ and $\hat{\pi}_t(y)$:

$$\pi_{t+1}(x)=\frac{1}{\bar{l}}\sum_{l=2}^{\infty} v_l l\int\prod_{i=1}^{l-1} dy_i\,\hat{\pi}_t(y_i)\times\int dh\,p(h)\,\delta\left(x-h-\sum_{i=1}^{l-1} y_i\right), \tag{2.10}$$

$$\hat{\pi}_{t+1}(y)=\frac{1}{\bar{k}}\sum_{k=3}^{\infty} c_k k\int\prod_{i=1}^{k-1} dx_i\,\pi_t(x_i)\times\delta\left(y-\frac{1}{\zeta}\,\text{arctanh}\left[\prod_{i=1}^{k-1}\tanh\zeta x_i\right]\right). \tag{2.11}$$

These equations (usually called the *density evolution* equations) are correct for times $t\ll\ln N$ due to the fact that the Tanner graph is locally treelike. They allow us therefore to predict whether, for a given ensemble of codes and noise level [recall that the noise level is hidden in $p(h)$] the algorithm is able to recover the transmitted code word (for large N). If this is the case, the distributions $\pi_t(x)$ and $\hat{\pi}_t(y)$ will concentrate on $x=y=+\infty$ as $t\to\infty$. In the opposite case the above iteration will converge to some distribution supported on finite values of x and y. In Table I we report the threshold noise levels for several regular codes, obtained using the density evolution method, together with the thresholds for the optimal decoding strategy, see Ref. [32].

Finally let us notice that the fixed point of the iteration, Eqs. (2.10) and (2.11), is the replica symmetric order parameter for the equivalent spin model.

TABLE I. The statical and dynamical points for several regular codes and decoding algorithms, cf. Eqs. (2.4) and (2.5).

	BEC		BSC			
(k,l)	p_c	p_d	p_c	$p_d(\zeta=1)$	$p_d(\zeta=2)$	$p_d(\zeta=\infty)$
(6,3)	0.4882	0.4294	0.100	0.084	0.078	0.072
(10,5)	0.4995	0.3416	0.109	0.070	0.056	0.046
(14,7)	0.5000	0.2798	0.109	0.056	0.039	0.029
(6,5)	0.8333	0.5510	0.264	0.139	0.102	0.078

III. STATISTICAL MECHANICS FORMULATION AND THE REPLICA APPROACH

We want to define a statistical mechanics model which describes the decoding problem. The probability distribution for the input code word $\underline{\sigma}=(\sigma_1,\ldots,\sigma_N)$ conditional to the received message takes the form

$$P(\underline{\sigma})=\frac{1}{Z}\,\delta_{\mathbb{H}}[\underline{\sigma}]\exp\left\{\sum_{i=1}^{N} h_i\sigma_i\right\}, \tag{3.1}$$

where $\delta_{\mathbb{H}}[\underline{\sigma}]=1$ if $\underline{\sigma}$ satisfies the parity checks encoded by the matrix $\mathbb{H}$, cf. Eq. (2.1), and $\delta_{\mathbb{H}}[\sigma]=0$ otherwise. Since we assume the input code word to be $\underline{\sigma}^{\text{in}}=(+1,+1,\ldots,+1)$, the h_i are i.i.d. with distribution $p(h)$.

We modify the probability distribution (3.1) in two ways.

(1) We multiply the fields h_i by a weight $\hat{\zeta}$. This allows us to tune the importance of the received message, analogously to Eqs. (2.4) and (2.5). This modification was already considered in Ref. [32]. Particularly important cases are $\hat{\zeta}=1$ and $\hat{\zeta}=0$.

(2) We relax the constraints implied by the characteristic function $\delta_{\mathbb{H}}[\underline{\sigma}]$. More precisely, let us denote each parity check by the unordered set of bit positions $(i_1,\ldots,i_k)$ which appear in it. For instance the three parity checks in Fig. (2) are (1,4,5,7), (2,4,6,7), (3,5,6,7). Moreover, let Ω_k be the set of all parity checks involving k bits (in the irregular ensemble, the size of Ω_k is Nc_k). We can write explicitly the characteristic function $\delta_{\mathbb{H}}[\underline{\sigma}]$ as follows:

$$\delta_{\mathbb{H}}[\underline{\sigma}]=\prod_{k=3}^{\infty}\prod_{(i_1\cdots i_k)\in\Omega_k}\delta(\sigma_{i_1}\cdots\sigma_{i_k},+1), \tag{3.2}$$

where $\delta(\cdot,\cdot)$ is the Kronecker delta function. Now it is very simple to relax the constraints by making the substitution $\delta(\sigma_{i_1}\cdots\sigma_{i_k},+1)\to\exp\{\beta[\sigma_{i_1}\cdots\sigma_{i_k}-1]\}$.

Summarizing the above considerations, we shall consider the statistical mechanics model defined by the Hamiltonian

$$H(\sigma)=-\sum_{k=3}^{\infty}\sum_{(i_1\ldots i_k)\in\Omega_k}(\sigma_{i_1}\cdots\sigma_{i_k}-1)-\frac{\hat{\zeta}}{\beta}\sum_{i=1}^{N} h_i\sigma_i \tag{3.3}$$

at inverse temperature β.

We address this problem by the replica approach [43]. The replicated partition function reads

[5] We denote by $h(p)$ the binary entropy function $h(p)=-p\log_2 p-(1-p)\log_2(1-p)$. It is useful to define its inverse: we denote by $\delta_{GV}(R)$ (the so-called Gilbert-Varshamov distance) the smallest solution of $h(\delta)=1-R$.

$$\langle Z^n\rangle \sim \int \prod_{\vec{\sigma}} d\lambda(\vec{\sigma})d\hat{\lambda}(\vec{\sigma})e^{-NS[\lambda,\hat{\lambda}]}, \tag{3.4}$$

with the action

$$\begin{aligned}S[\lambda,\hat{\lambda}] &= \overline{l}\sum_{\vec{\sigma}} \lambda(\vec{\sigma})\hat{\lambda}(\vec{\sigma}) \\ &\quad -\frac{\overline{l}}{\overline{k}}\sum_{k=3}^{\infty} c_k \sum_{\vec{\sigma}_1\cdots\vec{\sigma}_k} J_\beta(\vec{\sigma}_1,\dots,\vec{\sigma}_k)\lambda(\vec{\sigma}_1)\cdots\lambda(\vec{\sigma}_k) \\ &\quad -\sum_{l=2}^{\infty} v_l \ln\left[\sum_{\vec{\sigma}} \hat{\lambda}(\vec{\sigma})^l\mathcal{H}(\vec{\sigma})\right] - \overline{l} + \frac{\overline{l}}{\overline{k}},\end{aligned} \tag{3.5}$$

where

$$J_\beta(\vec{\sigma}_1,\dots,\vec{\sigma}_k) \equiv \exp\left[\beta\sum_a (\sigma_1\dots\sigma_k - 1)\right],$$

$$\mathcal{H}(\vec{\sigma}) = \left\langle \exp\left(\hat{\zeta}h\sum_a \sigma_a\right)\right\rangle_h, \tag{3.6}$$

$\langle\cdot\rangle_h$ being the average over $p(h)$. The order parameters $\lambda(\vec{\sigma})$ and $\hat{\lambda}(\vec{\sigma})$ are closely related, at least in the replica symmetric approximation, to the distribution of messages in the decoding algorithm [32], cf. Eqs. (2.10) and (2.11).

In the case of the BEC an irrelevant infinite constant must be subtracted from the action (3.5) in order to get finite results. This corresponds to taking

$$\mathcal{H}_{BEC}(\vec{\sigma}) \equiv p + (1-p)\,\delta_{\vec{\sigma},\vec{\sigma}_0}, \tag{3.7}$$

where $\vec{\sigma}_0 = (+1,\dots,+1)$.

IV. BINARY ERASURE CHANNEL: ANALYTICAL AND NUMERICAL RESULTS

The binary erasure channel is simpler than the general case. Intuitively this happens because one cannot receive misleading indications concerning a bit. Nonetheless it is an important case both from the practical [44] and from the theoretical point of view [34,38,45].

A. The decoding algorithm

Belief propagation becomes particularly simple in this context, and can be interpreted as an iterative decimation of the Tanner graph [38]. Since the knowledge about a received bit is completely sure, the log likelihoods $\{h_i\}$, cf. Eq. (2.3), take the values $h_i = +\infty$ (when the bit has been received[6]) or $h_i = 0$ (when it has been erased).

The analysis of this algorithm [34] uses the density evolution equations (2.10) and (2.11) and is greatly simplified because the messages $\{x^{(t)}_{i\to\alpha}\}$ and $\{y^{(t)}_{\alpha\to i}\}$ take only two values. Their distributions have the form

$$\begin{aligned}\pi_t(x) &= \rho_t\delta(x) + (1-\rho_t)\,\delta_\infty(x),\\ \hat{\pi}_t(x) &= \hat{\rho}_t\delta(y) + (1-\hat{\rho}_t)\,\delta_\infty(y),\end{aligned} \tag{4.1}$$

where $\delta_\infty(\cdot)$ is a delta function centered at $+\infty$. The parameters ρ and $\hat{\rho}$ give the fraction of zero messages, respectively, from variables to checks and from checks to variables. Using Eqs. (2.10) and (2.11), we get

$$\rho_{t+1} = p\frac{v'(\hat{\rho}_t)}{v'(1)}, \quad \hat{\rho}_{t+1} = 1 - \frac{c'(1-\rho_t)}{c'(1)}. \tag{4.2}$$

The initial condition $\rho_0 = \hat{\rho}_0 = 1$ converges to the perfect recovery fixed point $\rho = \hat{\rho} = 0$ if $p < p_d$. This corresponds to perfect decoding. For $p > p_d$ the algorithm gets stuck on a nontrivial linear system: $\rho_t \to \rho_*$, $\hat{\rho}_t \to \hat{\rho}_*$, with $0 < \rho_*$, $\hat{\rho}_* < 1$. The two regimes are illustrated in Fig. 4.

B. Statical transition

In the spin model corresponding to the situation described above, we have two types of spins: those corresponding to correctly received bits, which are fixed by an infinite magnetic field $h_i = +\infty$; and those corresponding to erased bits, on which no magnetic field acts: $h_i = 0$. We can therefore consider an effective model for the erased bits once the received ones are fixed to $+1$. This corresponds somehow to what is done by the decoding algorithm: the received bits are set to their values in the very first step of the algorithm and remain unchanged thereafter.

Let us consider the zero-temperature limit. If the system is in equilibrium, its probability distribution will concentrate on zero-energy configurations: the code words. We will have typically $\mathcal{N}_{\text{words}}(p) \sim 2^{Ns_{\text{words}}(p)}$ code words compatible with the received message. Their entropy $s_{\text{words}}(p)$ can be computed within the replica formalism, cf. Appendix A. The result is

$$s_{\text{words}}(\rho,\hat{\rho};p) = \overline{l}\rho(1-\hat{\rho}) + \frac{\overline{l}}{\overline{k}}c(1-\rho) + pv(\hat{\rho}) - \frac{\overline{l}}{\overline{k}}, \tag{4.3}$$

which has to be maximized with respect to the order parameters ρ and $\hat{\rho}$. The saddle point equations have exactly the same form as the fixed point equations corresponding to the dynamics (4.2), namely, $\rho = pv'(\hat{\rho})/v'(1)$ and $\hat{\rho} = 1 - c'(1-\rho)/c'(1)$

The saddle point equations have two stable solutions, i.e., local maxima of the entropy (4.3): (i) a completely ordered solution $\rho = \hat{\rho} = 0$, with entropy $s_{\text{words}}(0,0) = 0$ (in some cases this solution becomes locally unstable above some noise p_{loc}); (ii) (for sufficiently high noise level) a paramagnetic solution $\rho_*, \hat{\rho}_* > 0$. The paramagnetic solution appears at the same value p_d of the noise above which the decoding algorithm gets stuck.

[6] Recall that we are assuming the channel input to be $\sigma_i^{\text{in}} = +1$ for $i = 1,\dots,N$.

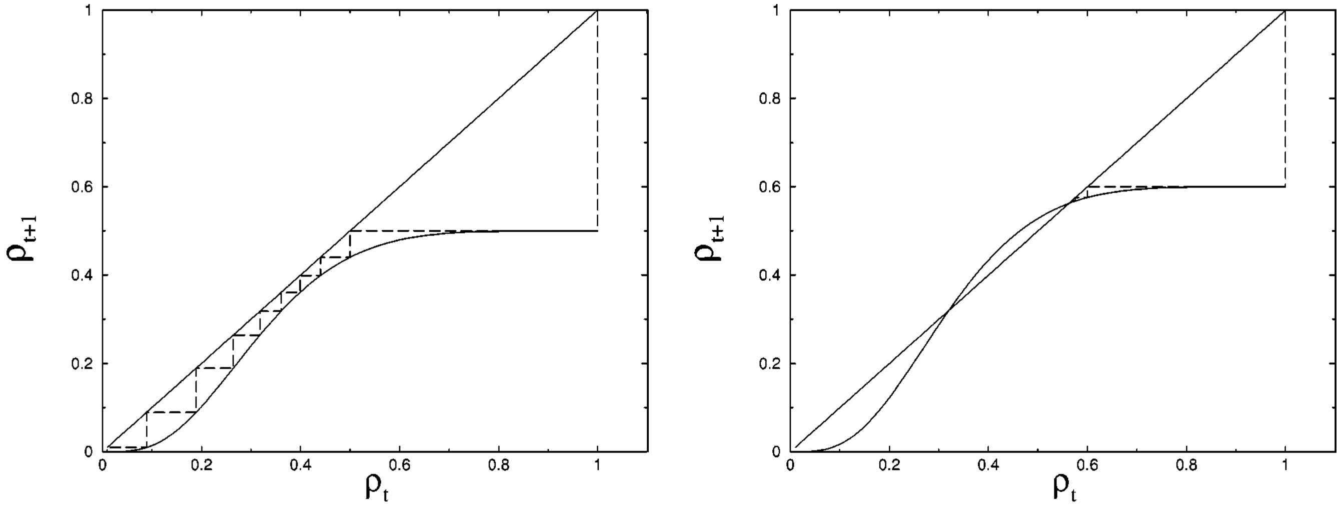

FIG. 4. The evolution of the iterative decoding algorithm on the BEC, cf. Eqs. (4.2). Here we consider the (6,5) code: $\rho_{t+1}=p[1-(1-\rho_t)^5]^4$. On the left $p=0.5<p_d$, on the right $p=0.6>p_d$.

The fixed point to which the dynamics (4.2) converges coincides with the statistical mechanics result for $\rho_*,\hat{\rho}_*$. However the entropy of the paramagnetic solution $s_{\text{words}}(\rho_*,\hat{\rho}_*)$ is negative at p_d and becomes positive only above a certain critical noise p_c. This means that the linear system produced by the algorithm continues to have a unique solution below p_c, although our linear-time algorithm is unable find such a solution.

The "dynamical" critical noise p_d is the solution of the following equation:

$$p\frac{v''(\hat{\rho}_*)c''(1-\rho_*)}{v'(1)c'(1)}=-1, \tag{4.4}$$

where ρ_* and $\hat{\rho}_*$ solve the saddle point equations. The statical noise can be obtained by setting $s_{\text{words}}(\rho_*,\hat{\rho}_*)=0$. Finally the completely ordered solution becomes locally unstable for

$$p_{loc}=\frac{c'(1)v'(1)}{v''(0)c''(1)}. \tag{4.5}$$

As an example, let us consider the one-parameter family of $R=1/2$ codes specified by the following generating polynomials: $c(x)=\alpha x^4+(1-\alpha)x^6$, $v(x)=\alpha x^2+(1-\alpha)x^3$. This is an irregular code which smoothly interpolates between the regular (6,3) and (4,2) codes. The local stability threshold is given by

$$p_{loc}(\alpha)=\frac{(3-\alpha)^2}{6\alpha(5-3\alpha)}. \tag{4.6}$$

The dynamical and critical curves $p_d(\alpha)$ and $p_c(\alpha)$ are reported in Fig. 5. Notice that the α value where $p_d(\alpha)$ reaches its maximum, corresponding to the best code in this family, is neither 0 nor 1. This is a simple example showing that irregular codes $(0<\alpha<1)$ are generally superior to regular ones ($\alpha=0$ or $\alpha=1$ in this example). Notice also that above the tricritical point $\alpha_t\approx 0.793\,014\,12$, $p_t\approx 0.390\,577\,24$, the three curves $p_{loc}(\alpha)$, $p_c(\alpha)$, and $p_d(\alpha)$ coincide. In the following we shall study in some detail the $\alpha=0$ case, which corresponds to a regular (6,3) code, the corresponding critical and dynamical points p_c and p_d are given in Table I.

C. Dynamical transition

The dynamical transition is not properly described within the replica symmetric treatment given above. Indeed, the paramagnetic solution cannot be considered, between p_d and p_c, as a metastable state because it has negative entropy. One cannot therefore give a sensible interpretation of the

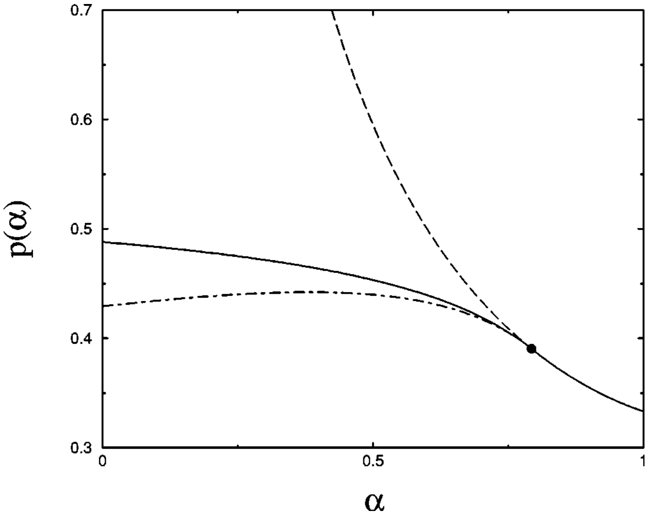

FIG. 5. The phase diagram of the family of codes with generating polynomials $c(x)=\alpha\mathbf{x}^4+(1-\alpha)\mathbf{x}^6$, $v(x)=\alpha\mathbf{x}^2+(1-\alpha)\mathbf{x}^3$. The dashed line gives the local stability threshold for the completely ordered ferromagnetic phase. The continuous and dot-dashed lines refer (respectively) to the static and dynamic critical points $p_c(\alpha)$ and $p_d(\alpha)$.

coincidence between the critical noise for the decoding algorithm and the appearance of the paramagnetic solution.

Before embarking into the one-step replica symmetry-breaking (1RSB) calculation, let us review some well-known facts [46,47]. Let us call $m\phi(\beta,m)$ the free energy of m weakly coupled "real" replicas times β. This quantity can be computed in the 1RSB calculation. In the limit $\beta\to\infty$, with $m\beta=\mu$ fixed, we have $m\phi(\beta,m)\to\mu\phi(\mu)$. The number of metastable states with a given energy density ϵ is

$$\mathcal{N}_{MS}(\epsilon)\sim e^{N\Sigma(\epsilon)}, \tag{4.7}$$

where the complexity $\Sigma(\epsilon)$ is the Legendre transform of the m replicas free energy:

$$\Sigma(\epsilon)=\mu\epsilon-\mu\phi(\mu)\big|_{\epsilon=\partial[\mu\phi(\mu)]}. \tag{4.8}$$

The (zero-temperature) dynamic energy ϵ_d and the static energy ϵ_s are,[7] respectively, the maximum and the minimum energy such that $\Sigma(\epsilon)\geqslant 0$.

The static energy is obtained by solving the following equations:

$$\epsilon_s=\phi(\mu),$$

$$\partial\phi(\mu)=0, \tag{4.9}$$

which corresponds to the usual prescription of maximizing the free energy over the replica-symmetry-breaking parameter m [43]. The dynamic energy is given by

$$\epsilon_d=\partial[\mu\phi(\mu)],$$

$$\partial^2[\mu\phi(\mu)]=0. \tag{4.10}$$

Finally, if $\epsilon_s=0$ the complexity of the ground state is $\Sigma(0)=-\lim_{\mu\to\infty}\mu\phi(\mu)$.

We were not able to exactly compute the 1RSB free energy $\phi(\mu)$. However excellent results can be obtained within an "almost factorized" variational ansatz, cf. Appendix . The picture that emerges is the following.

(1) In the low noise region ($p<p_d$), no metastable states exist. Local search algorithms should therefore be able to recover the erased bits.

(2) In the intermediate noise region ($p_d<p<p_c$) an exponentially large number of metastable states appear. They have energy densities ϵ in the range $\epsilon_s<\epsilon<\epsilon_d$, with $\epsilon_s>0$. Therefore the transmitted code word is still the only one compatible with the received message. Nonetheless a large number of extremely stable *pseudo-code-words* stop local algorithms. The number of violated parity checks in these code words cannot be reduced by means of local moves.

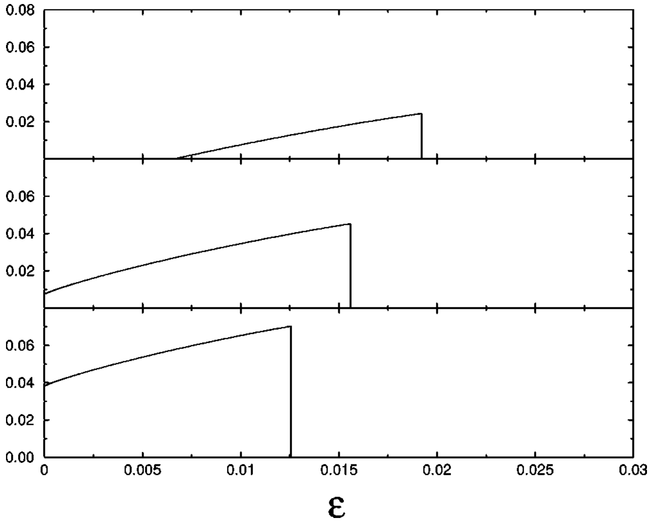

FIG. 6. The complexity $\Sigma(\epsilon)$ for (from top to bottom) $p=0.45$ (below p_c), $p=0.5$, and $p=0.55$ (above p_c).

(3) Above p_c we have $\epsilon_s=0$: a fraction of the metastable states is made of code words. Moreover, $\Sigma(0)$ (which gives the number of such code words) coincides with the paramagnetic entropy $s_{\text{words}}(\rho_*,\hat{\rho}_*)$ computed in the preceeding section.

As an illustration, let us consider the (6,3) regular code. In Fig. 6 we plot the resulting complexity curves $\Sigma(\epsilon)$ for three different values of the erasure probability p. In Fig. 7, left frame, we report the static and dynamic energies ϵ_s and ϵ_d as functions of p. In the right frame we present the total complexity $\Sigma_{\text{tot}}\equiv\max_\epsilon\Sigma(\epsilon)=\Sigma(\epsilon_d)$, and the zero-energy complexity $\Sigma(0)$.

D. Numerical results

In order to check analytical predictions and to better illustrate the role of metastable states, we have run a set of Monte Carlo simulations, with Metropolis dynamics, on the Hamiltonian (3.3) of the (6,3) regular code for the BEC. Notice that local search algorithms for the decoding problem have been already considered by the coding theory community [48].

We studied quite large codes ($N=10^4$ bits), and tried to decode it (i.e., to find a ground state of the corresponding spin model) with the help of simulated annealing techniques [49]. For each value of p, we start the simulation fixing a fraction $(1-p)$ of spins to $\sigma_i=+1$ (this part will be kept fixed all along the run). The remaining pN spins are the dynamical variables we change during the annealing in order to try to satisfy all the parity checks. The energy of the system counts the number of unsatisfied parity checks.

The cooling schedule has been chosen in the following way: τ Monte Carlo sweeps[8] (MCS) at each of the 1000

[7] Notice that one can give (at least) three possible definitions of the dynamic energy: (i) from the solution of the nonequilibrium dynamics: $\epsilon_d^{(d)}$, (ii) imposing the replicon eigenvalue to vanish: $\epsilon_d^{(r)}$, (iii) using, as in the text, the complexity $\Sigma(\epsilon)$, $\epsilon_d^{(c)}$. The three results coincide in the p-spin spherical fully connected model, however their equality in the present case is, at most, a conjecture.

[8] Each Monte Carlo sweep consists in N proposed spin flips. Each proposed spin flip is accepted or not accordingly to a standard Metropolis test.

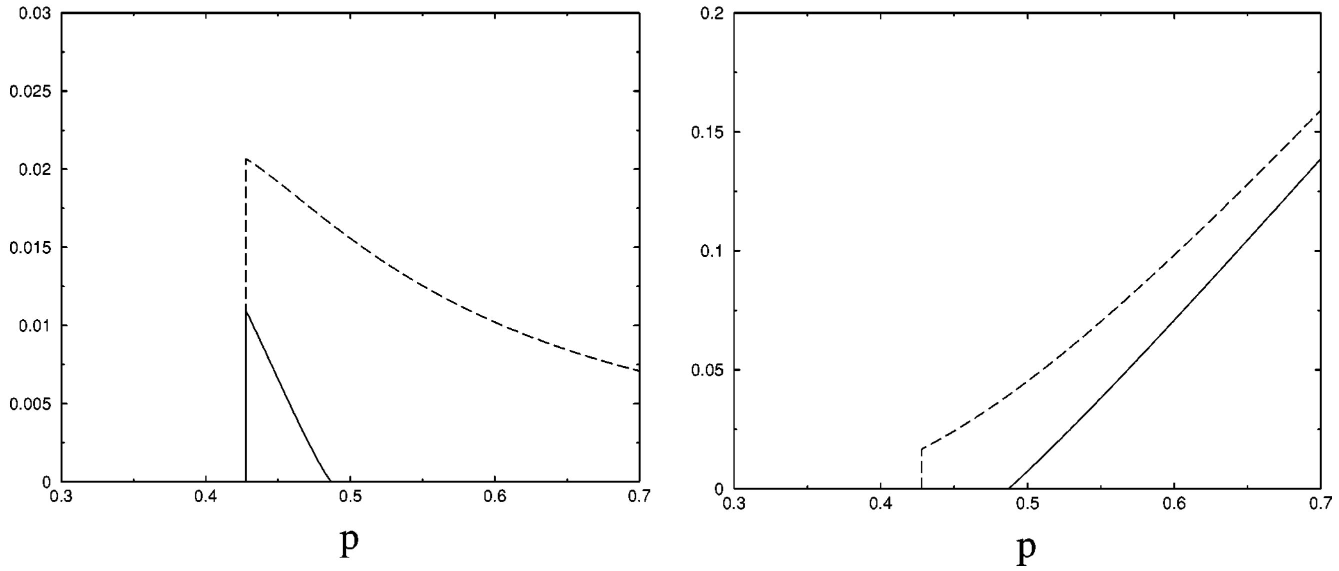

FIG. 7. Left-hand frame: the static and dynamic energies ϵ_s and ϵ_d of the metastable states (respectively, solid and dashed lines). Right-hand frame: the total complexity $\max_\epsilon \Sigma(\epsilon)$ and the zero-energy complexity $\Sigma(0)$.

equidistant temperatures between $T=1$ and $T=0$. The highest temperature is such that the system very rapidly equilibrates on the paramagnetic energy $\epsilon_P(T)$. Typical values for τ are from 1 to 10^3.

Notice that, for any fixed cooling schedule, the computational complexity of the simulated annealing method is linear in *N*. Then we expect it to be affected by metastable states of energy ϵ_d, which are present for $p>p_d$: the energy relaxation should be strongly reduced around ϵ_d and eventually be completely blocked.

In order to illustrate how the system relaxes during the simulated annealing, we show in Fig. 8 the energy density as a function of the temperature for $p=0.4$ (left) and $p=0.6$ (right) and various cooling rates, $\tau=10,10^2,10^3$ (each dataset is the average over many different samples).

For $p=0.4<p_d$ the final energy strongly depends on the cooling rate and the slowest cooling procedure is always able to bring the system to the ground state, corresponding to the transmitted code word. Decoding by simulated annealing is therefore successful.

For $p=0.6>p_d$ the situation drastically changes. Below a temperature T_d (marked by an arrow in Fig. 8, right frame) there is an almost complete stop of the energy relaxation. T_d marks the dynamical transition, and the corresponding energy $\epsilon_d(T_d)=\epsilon_P(T_d)$ is called the threshold energy. The energy of threshold states still varies a little bit with temperature, $\epsilon_d(T)$, and the final value reached by the simulated annealing algorithm is its zero-temperature limit $\epsilon_d(0)=\epsilon_d$. Remember that, by construction, ground states of zero energy are present for any *p* value, but they become unreachable for $p>p_d$, because they become shielded by metastable states of higher energy.

We show in Fig. 9 the lowest energy reached by the simulated annealing procedure for different *p* and τ values. While for $p<p_d$ all parity checks can be satisfied and the energy relaxes to zero in the limit of a very slow cooling, for

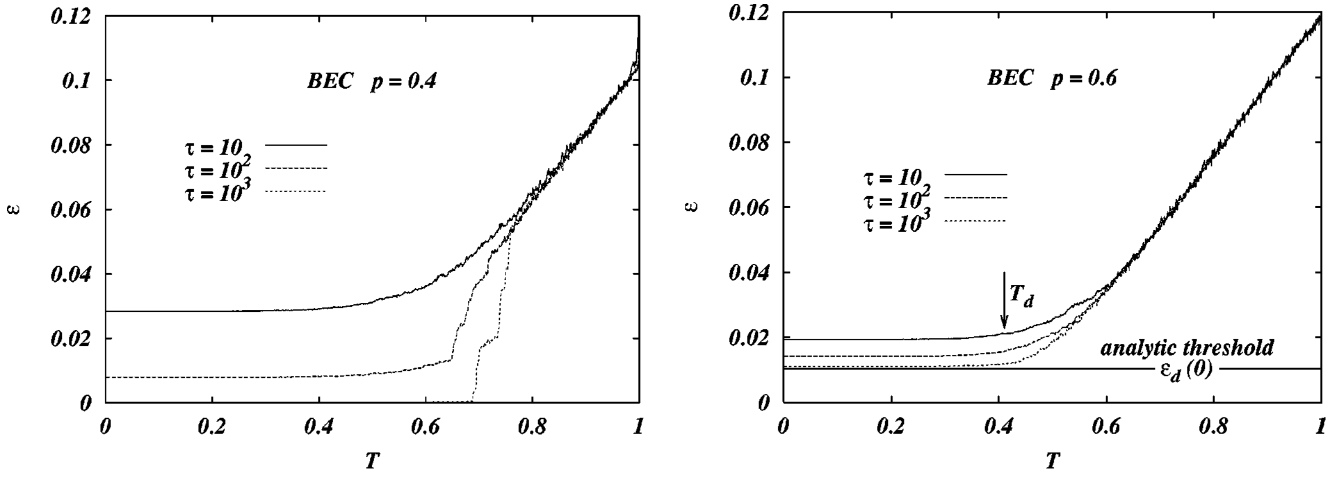

FIG. 8. Energy relaxation for the Hamiltonian of the (6,3) regular code during the simulated annealing with τ MCS per temperature and 1000 equidistant temperatures in [0,1]

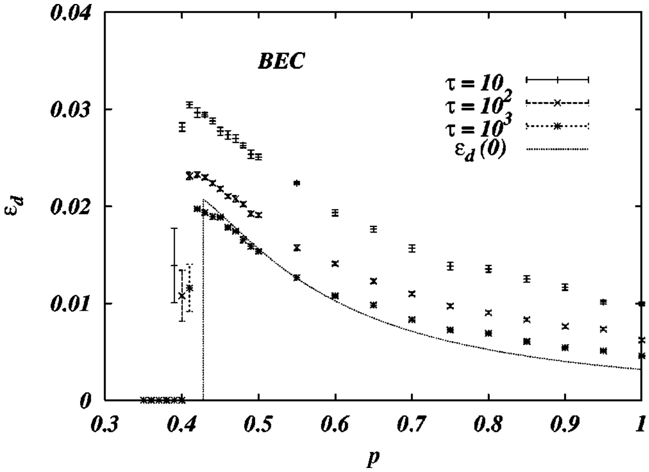

FIG. 9. Lowest energies reached by the simulated annealing. Errors are sample-to-sample fluctuations.

$p \geqslant p_d$ the simulation gets stuck in a metastable state of finite energy, that is, with a number of unsatisfied parity checks of order N. The agreement with the analytic prediction (dotted line) is quite good everywhere, but very close to p_d.

Discrepancies between analytical predictions and numerical results may be very well due to finite-size effects in the latter. One possible explanation for large finite-size effects near the dynamic critical point p_d is the following. Metastable states of energy ϵ_d are stable under any local dynamic, which may flip only a finite number of spins simultaneously, and under global dynamics flipping no more than ωN spins simultaneously. Physical intuition (threshold states become more robust increasing p) imply that the function $\omega(p)$ must monotonically increase for $p \in [p_d, 1]$. Moreover, continuity reasons tell us that $\omega(p_d) = 0$. The fact that $\omega(p)$ is very small close to p_d, together with the fact that in numerical simulations we are restricted to finite values of N, allows the local Monte Carlo dynamic to relax below the analytical predicted threshold energy. A more detailed characterization of this effect is presently under study and will be presented in a forthcoming publication.

V. THE GENERAL CHANNEL: ANALYTICAL AND NUMERICAL RESULTS

We considered the case of a general noisy channel using two different approaches: a finite-temperature and a zero-temperature approach. While the first one offers a clear connection with the dynamics of the decoding-by-annealing algorithm, the second one gives a nice geometrical picture of the situation.

A. Finite temperature

Suppose you received some message encoded using a Gallager code and you want to decode it, but no one explained to you the belief propagation algorithm, cf. Eqs. (2.4) and (2.5).

A physicist's idea would be the following. Write the corresponding Hamiltonian $H(\underline{\sigma})$, see Eq. (3.3), and run a Monte Carlo algorithm at inverse temperature β. If you wait enough, you will be able to sample the configuration $\underline{\sigma}$ according to the Boltzmann distribution $P_\beta(\underline{\sigma}) \propto e^{-\beta H(\underline{\sigma})}$. Then cool down the system adiabatically: i.e., change the temperature according to some schedule $\{\beta_1, \beta_2, \dots, \}$ with $\beta_k \uparrow \infty$, waiting enough at each temperature for the system to equilibrate.

As $\beta \to \infty$ the Boltzmann measure of the Hamiltonian (3.2) concentrates on the code words [for which the exchange term in Eq. (3.2) is equal to zero]. Moreover, each code word is given a weight that depends on its likelihood. In formulas,

$$\lim_{\beta \to \infty} P_\beta(\underline{\sigma}) = \frac{1}{Z_{\hat{\zeta}}} P(\underline{\sigma} | \underline{\mathbf{x}}^{\rm out})^{\hat{\zeta}}, \tag{5.1}$$

where $P(\underline{\sigma} | \underline{\mathbf{x}}^{\rm out})$ is the probability for $\underline{\sigma}$ to be the transmitted code word, conditional to the received message $\underline{\mathbf{x}}^{\rm out}$, and $Z_{\hat{\zeta}}$ is a normalization constant. Therefore when $\beta \gg 1$, our algorithm will sample a code word with probability proportional to $P(\underline{\sigma} | \underline{\mathbf{x}}^{\rm out})^{\hat{\zeta}}$. For good codes below the critical noise threshold p_c, the likelihood $P(\underline{\sigma} | \underline{\mathbf{x}}^{\rm out})$ is strongly concentrated[9] on the correct input code word. Therefore the system will spend most of its time on the correct code word as soon as $\beta \gg 1$ and $\hat{\zeta} \geqslant 1$ (for $\hat{\zeta} < 1$, p_c has a nontrivial dependence on $\hat{\zeta}$, cf. Ref. [32]).

This algorithm will succeed as long as we are able to keep the system in equilibrium at all temperatures down to zero. If some form of ergodicity breaking is present this may take an exponentially (in the size N) long time. Let us suppose that $O(N)$ computational time is spent at each temperature β_i of the annealing schedule (this is what happens in nature). We expect to be able to equilibrate the system only at low enough noise [let us say for $p < p_d(\hat{\zeta})$], when the magnetic field in Eq. (3.3) is strong enough to single out a unique ergodic component.

1. Theoretical dynamical line

The existence of metastable states can be detected within the replica formalism by the so-called marginal stability condition. One considers the saddle point equations for the 1RSB order parameter, fixing the RSB parameter $m = 1$, cf. Appendix B. The dynamical temperature $T_d(p)$ is the highest temperature for which a "nontrivial" solution of the equation exists. At this temperature, the ergodicity of the physical dynamics breaks down (at least this is what happens in infinite connectivity mean-field models) and we are no longer able to equilibrate the system within an $O(1)$ physical time [i.e., an $O(N)$ computational time].

We looked for a solution of Eqs. (B3)–(B6) using the population dynamics algorithm of Ref. [19]. We checked the "nontriviality" of the solution found by considering the vari-

[9] Namely, we have $P(\underline{\sigma}^{\rm in} | \underline{\mathbf{x}}^{\rm out}) = 1 - O(e^{-\alpha N})$. This happens because there is a minimum $O(N)$ Hamming distance between distinct code words [15].

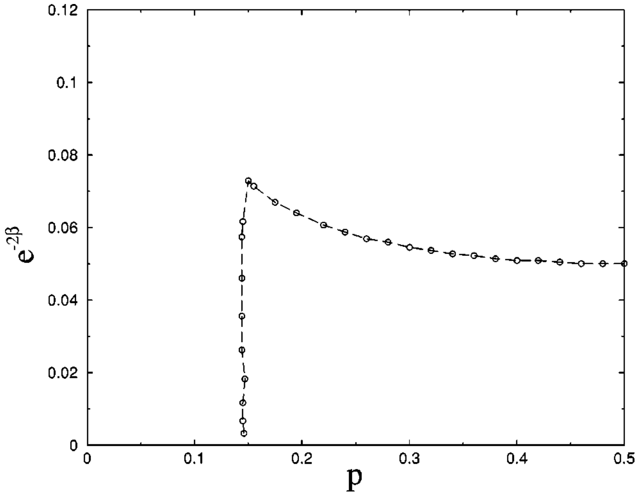

FIG. 10. The dynamical phase transition for a regular (6,5) code [cf. Eq. (3.2) with $k=6$ and $l=5$] with $\hat{\zeta}=1$.

ance of the distributions $\rho(x)$, $\hat{\rho}(y)$.

We consider the (6,5) regular code because it has well separated statical and dynamical thresholds p_c and p_d, cf. Table I. The resulting dynamical line for the Hamiltonian (3.2) with $\hat{\zeta}=1$ is reported in Fig. 10. The dynamic temperature $T_d(p)$ drops discontinuously below a noise $p_d(\hat{\zeta})$: for $p<p_d(\hat{\zeta})$ the dynamical transition disappears and the system can be equilibrated in linear computational time down to zero temperature. We get $p_d(1)\approx 0.14$, which is in good agreement with the coding theory results, cf. Table I.

2. Numerical experiments

We have repeated for the BSC the same kind of simulations already presented at the end of Sec. IV D for the BEC.

We have run a set of simulated annealings for the Hamiltonian (3.3) of the (6,5) regular code. System size is $N=12\,000$ and the cooling rates are the same as for the BEC, the only difference being the starting and the ending temperatures, which are now $T=1.2$ and $T=0.2$ (plus a quench from $T=0.2$ to $T=0$ at the end of each cooling). This should not have any relevant effect because $0.2 \ll T_d \approx 0.6$.

The important difference with respect to the BEC case is that now we have no fixed spins: all N spins are dynamical variables subject to a random external field of intensity $h=(1/\beta)\mathrm{arctanh}(1-2p)$, cf. Eq. (3.3).

Also here, as in the case of the BEC, the energy relaxation for $p>p_d$ undergoes a drastic arrest when the temperature is reduced below the dynamical transition at T_d, see Fig. 11.

Unfortunately, in this case, we are not able to calculate analytically the threshold energy $\epsilon_d(0)$, but only the dynamical critical temperature T_d and then the threshold energy at the transition $\epsilon_d(T_d)$, which is higher than $\epsilon_d(0)$. The difference $\Delta\epsilon=\epsilon_d(T_d)-\epsilon_d(0)$ is usually not very large (see, e.g., the BEC case), but it becomes apparent when p is decreased towards p_d. Indeed for $p=0.25$ (Fig. 11 left), the Metropolis dynamics is still able to relax the system for temperatures below T_d and then it reaches an energy well below $\epsilon_d(T_d)$. On the other hand, for $p=0.5$ (Fig. 11 right), where $\Delta\epsilon$ is small, the relaxation below T_d is almost absent and the analytic prediction is much more accurate. Notice that for this case we have run a still longer annealing with $\tau=10^4$: the asymptotic energy is very close to that for $\tau=10^3$ and hardly distinguishable from the analytical prediction.

In Fig. 12 we report the lowest energy reached by the simulated annealing for many values of p and $\tau=10,10^2,10^3$, together with the analytic calculation for the threshold energy at T_d.

B. Zero temperature

This approach follows from a physical intuition that is slightly different from that explained above. Once again we will formulate it algorithmically. For sake of simplicity we shall refer, in this section, to the BSC. We refer to Appendix for more general formulas.

The overlap between the transmitted code word and the received message

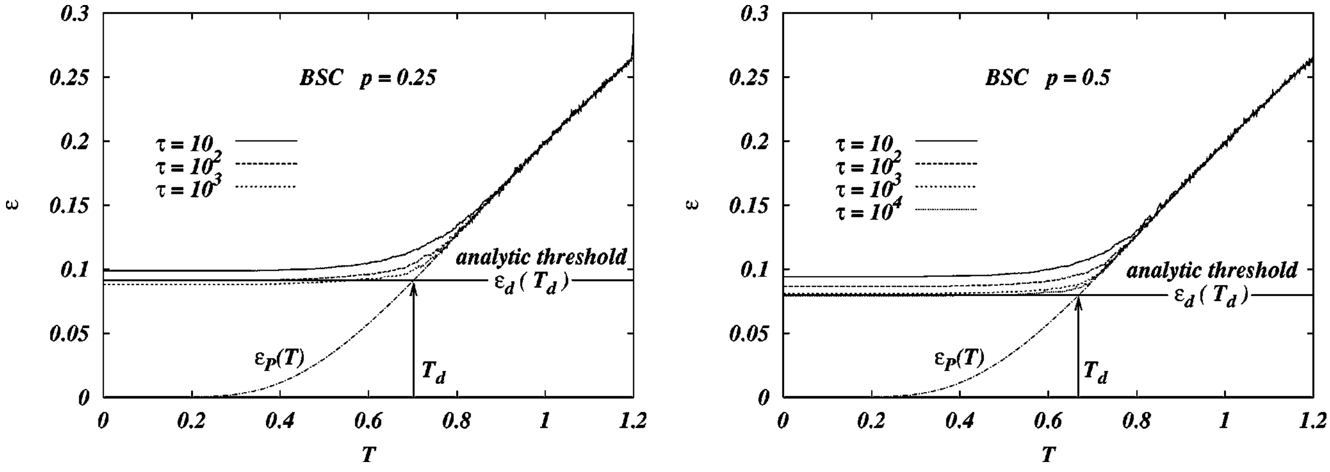

FIG. 11. Energy relaxation for the Hamiltonian of the (6,5) regular code during the simulated annealing with τ MCS per temperature and 1000 equidistant temperatures in [0.2,1.2]. Notice that in both cases, $p>p_d$. The dot-dashed line is the theoretical prediction for the paramagnetic exchange energy.

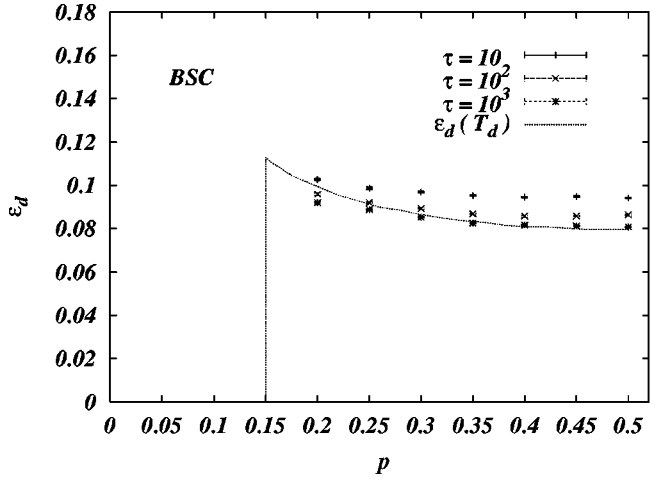

FIG. 12. Lowest energies reached by the simulated annealings. Errors are sample to sample fluctuations. The theoretical prediction $\epsilon_d(T_d)$ is computed using the results in Fig. 10 for $T_d(p)$.

$$q^{\mathrm{in,out}}=\frac{1}{N}\sum_{i=1}^{N}\sigma_i^{\mathrm{in}}\sigma_i^{\mathrm{out}}, \qquad (5.2)$$

is typically $q^{\mathrm{in,\,out}}=1-2p$. Given the received message, one can work in the subspace of all the possible configurations which have the prescribed overlap with it,[10] i.e., all the $\underline{\sigma}$ such that $(1/N)\Sigma_{i=1}^{N}\sigma_i\sigma_i^{\mathrm{out}}\approx(1-2p)$. Once this constraint has been imposed (for instance, in a Kawasaki-like Monte Carlo algorithm), one can restrict oneself to the exchange part of the Hamiltonian (3.2)

$$H_{\mathrm{exch}}(\underline{\sigma})=-\Sigma_k\Sigma_{(i_1\dots i_k)}\sigma_{i_1}\cdots\sigma_{i_k}$$

and apply the cooling strategy already described in the preceding section.

Below the static transition p_c there exists a unique code word having overlap $(1-2p)$ with the received signal. This is exactly the transmitted one $\underline{\sigma}^{\mathrm{in}}$. This means that $\underline{\sigma}^{\mathrm{in}}$ is the unique ground state of $H_{\mathrm{exch}}(\underline{\sigma})$ in the subspace we are considering. If we are able to keep our system in equilibrium down to $T=0$, the cooling procedure will finally yield the correct answer to the decoding problem. Of course, if metastable states are encountered in this process, the time required for keeping the system in equilibrium diverges exponentially in size.

We expect the number of such states to be exponentially large:[11]

$$\mathcal{N}_{MS}(\epsilon,q|p)\sim e^{N\Sigma_p(\epsilon,q)}, \qquad (5.3)$$

where ϵ is the exchange energy density $H_{\mathrm{exch}}(\underline{\sigma})/N$. Notice that we emphasized the dependence of these quantities upon the noise level p. In fact the noise level determines the statistics of the received message $\underline{\sigma}^{\mathrm{out}}$. The static threshold is the noise level at which an exponential number of code words with the same overlap as the correct one $(q=1-2p)$ appears: $\Sigma_p(0,1-2p)>0$. The dynamic transition occurs where metastable states with the same overlap begin to exist: $\Sigma_p(\epsilon,1-2p)>0$ for some $\epsilon>0$.

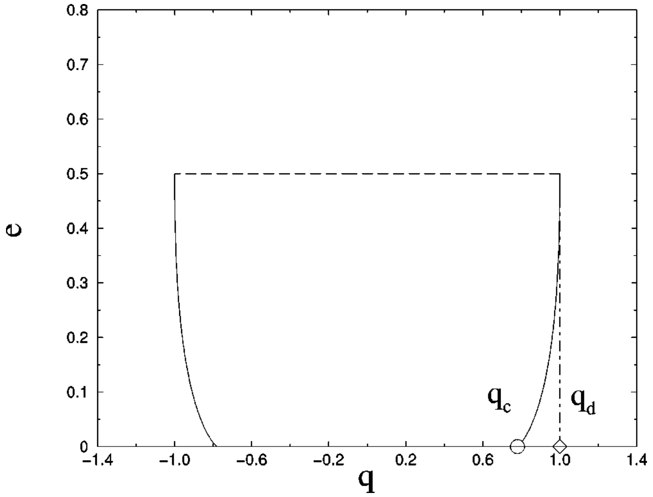

FIG. 13. Metastable states in the random linear code limit for $R=1/2$: their number is exponential between the continuous and the dashed lines. It vanishes discontinuously when the dashed line is crossed and continuously when the continuous line is crossed. The critical and dynamical overlaps are related to the statical and critical noise by $q_{c,d}=1-2p_{c,d}$. In this limit $p_d=0$ and $p_c=\delta_{GV}(1/2)\approx0.110\,025$.

1. The random linear code limit

It is quite easy to compute the complexity $\Sigma_p(\epsilon,q)$ in the limit $k,l\to\infty$ with the rate $R=1-l/k$ fixed. In particular, the zeroth-order term in a large k,l expansion can be derived by elementary methods.

Since the derivation is quite standard [32,50] we shall limit ourselves to quoting the result. Let us define the function

$$\tilde{\Sigma}(\epsilon,q)=h[(1-q)/2]+(1-R)h[\epsilon/2(1-R)]-(1-R). \qquad (5.4)$$

The number of metastable states is $\mathcal{N}_{MS}(\epsilon,q)\sim2^{N\Sigma(\epsilon,q)}$ with $\Sigma(\epsilon,q)=\tilde{\Sigma}(\epsilon,q)$ when $\tilde{\Sigma}(\epsilon,q)$, $\partial_\epsilon\tilde{\Sigma}(\epsilon,q)>0$, and $\Sigma(\epsilon,q)=-\infty$ otherwise.

In Fig. 13 we plot the region of the (ϵ,q) plane for which $\Sigma(\epsilon,q)>0$, for $R=1/2$ codes. Notice that in this limit $\Sigma(\epsilon,q)$ does not depend on the received message $\underline{\sigma}^{\mathrm{out}}$ (and, therefore, is independent of p). As expected, we get $p_c=\delta_{GV}(R)$ and $p_d=0$.

In order to get the first nontrivial estimate for the dynamical point p_d, we must consider the next term in the above expansion. This correction can be obtained within the replica formalism. The corresponding estimates for p_c and p_d are reported below for a few values of k and l:

[10] Of course this is true up to $O(N^{-1/2})$ corrections. For instance, one can work in the space of configurations $\underline{\sigma}$ such that $(1-2p-\delta)N<\Sigma_{i=1}^{N}\sigma_i\sigma_i^{\mathrm{out}}<(1-2p+\delta)N$, for some small number δ.

[11] For a related calculation in a fully connected model see Ref. [51].

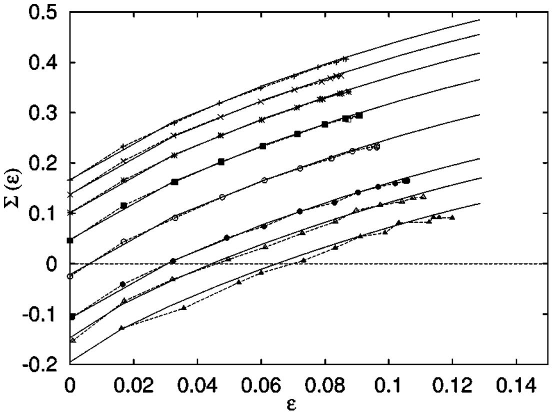

FIG. 14. The configurational entropy versus the energy for the (6,5) regular code. Symbols refer to various noise levels. From top to bottom $p=0.5,0.4,0.35,0.3,0.25,0.2,0.18,0.155$. Continuous lines give the result of a variational computation [20].

(k,l)	p_c	$p_d(1)$
(6,3)	0.097	0.071
(10,5)	0.108	0.060
(14,7)	0.109	0.049
(6,5)	0.264	0.108

2. The complete calculation

The full 1RSB solution can be obtained through the population dynamics method [19]. Here, as in Sec. V A 1, we focus on the example of the (6,5) code. In Fig. 14 we plot the configurational entropy as a function of the energy of the states along the lines of constant q, together with the corresponding results obtained within a simple variational approach. The approximate treatment is in quantitative agreement with the complete calculation for $\epsilon<\epsilon_d$, but predicts a value for the threshold energy, which is larger than the correct one: $\epsilon_d^{var}>\epsilon_d$. Here $\epsilon_d^{var}\approx 0.127$ and almost p independent.

Unhappily the estimate of the dynamic energy obtained from this curve is not very precise. Moreover, at least two more considerations prevent us from comparing these results with those of simulated annealing simulations, cf. Sec. V A 2: (i) In our annealing experiments the overlap with the received message $\underline{\sigma}^{\rm out}$ is free to fluctuate, (ii) We cannot exclude the fact that the 1RSB solution become unstable at low temperature.

However, the population dynamics solution give the estimate $p_d \lesssim 0.155$. This allows us to confirm that the point $p_d=0.139$ where the decoding algorithm fails to decode, cf. Table I, coincides with the point where the metastable states appear.

VI. CONCLUSIONS AND FUTURE PERSPECTIVES

In this work we have studied the dynamical phase transition for a large class of diluted spin models in a random field, the main motivation being their correspondence with very powerful error correcting codes.

In the particular case of a binary erasure channel, we were able to show that the dynamic critical point coincides exactly with the critical noise level for an important class of decoding algorithms, namely, belief propagation (cf. Sec. IV and Appendix A). Above this threshold, metastable states of high energy exist in an exponentially large number and they inhibit the decoding algorithm from converging in linear time. The presence of such metastable states has been verified by extensive Monte Carlo simulations: The energy at which very slow simulated annealings get stuck is compatible with the analytic prediction.

For a general model of the noisy channel, we could not present a completely explicit proof of the coincidence between the decoding threshold and the dynamical transition. However, also for the binary symmetric channel, we have obtained, within numerical precision, identical values for the algorithmic and the statistical mechanics thresholds.

It may be worth listing a few interesting problems which emerge from our work:

(1) We show explicitly that the identity between statistical mechanics and algorithmic thresholds holds in general. From a technical point of view, this is a surprising fact because the two thresholds are obtained, respectively, within replica symmetric, cf Eqs. (2.10) and (2.11), and a one-step replica-symmetry-breaking-calculations.

(2) We considered message-passing and simulated annealing algorithms. Extend the above analysis to other classes of algorithm (and, eventually, to any linear-time algorithm).

(3) Message-passing decoding algorithms get stuck because they are unable to decode some fraction of the received message, the ‘‘hard’’ bits, while they have been able to decode the other ones, the ‘‘easy’’ bits, cf. Appendix I A 1. A closer look at this heterogeneous behavior would be very fruitful (see Ref. [54] for a first attempt).

ACKNOWLEDGMENTS

It is a pleasure to thank Riccardo Zecchina who participated in the early stages of this work A.M. also thanks M. Mézard and R. Zecchina for fruitful suggestion [53].

APPENDIX A: CALCULATIONS, BINARY ERASURE CHANNEL

In this appendix we give the details of the replica calculation for the BEC. Notice that although we use the regular (6,3) code as a generic example, all the computations are presented for general degree distributions $\{c_k\}$ and $\{v_l\}$. Since the replica symmetric case can be regarded as a particular limit of the RSB one, we shall limit ourselves to detailing the last one.

Replica symmetry breaking

The exact computation of the 1RSB free energy is a difficult task for a finite connectivity model [18]. Good results can be obtained from following variational ansatz (see Ref. [52] for the general philosophy of the variational approach)

$$\lambda(\vec{\sigma})=(1-p)\delta_{\vec{\sigma},\vec{\sigma}_0}+pf(\underline{\sigma}^{(1)})\cdots f(\underline{\sigma}^{(n/m)}), \qquad (A1)$$

$$\hat\lambda(\vec\sigma)=\hat f(\underline{\sigma}^{(1)})\cdots\hat f(\underline{\sigma}^{(n/m)}),\tag{A2}$$

where $\underline{\sigma}^{(\alpha)}=(\sigma^{(\alpha-1)m+1},\dots,\sigma^{\alpha m})$. This amounts to considering a fraction of the spins (namely, those with an infinite magnetic field) as frozen in the $+1$ state, and assuming all the other spins to be equivalent. In the $n\to0$ limit we get $\partial_n S[\lambda,\hat\lambda]\to\phi[f,\hat f]$ with

$$\phi[f,\hat f]=\frac{\overline{l}p}{m}\ln\left(\sum_{\underline\sigma} f(\underline\sigma)\hat f(\underline\sigma)\right)-\frac{p}{m}\sum_{l=2}^{\infty}v_l\ln\left(\sum_{\underline\sigma}\hat f(\underline\sigma)^l\right)-\frac{\overline{l}}{\overline{k}m}\sum_{\nu=0}^{\infty}g_\nu\ln\left[\sum_{\underline\sigma_1\cdots\underline\sigma_k}J_\beta^{(m)}(\underline\sigma_1,\dots,\underline\sigma_k)\times f(\underline\sigma_1)\cdots f(\underline\sigma_k)\right],\tag{A3}$$

where $\underline\sigma$ are m-component replicated spins and

$$g_\nu\equiv\sum_{k=\nu}^{\infty}c_k\binom{k}{\nu}p^\nu(1-p)^{k-\nu}.\tag{A4}$$

The generating function of the coefficients $\{g_\nu\}$ is given by $g(x)=c(1-p+px)$. Notice that $\{g_\nu\}$ is the effective degree distribution of parity check nodes (i.e., the analogous of $\{c_k\}$), once the received bits have been eliminated.

Notice that the energy (A3) is invariant under a multiplicative rescaling of $f(\underline\sigma)$ and $\hat f(\underline\sigma)$. We shall fix this freedom by requiring that $\Sigma_{\underline\sigma}f(\underline\sigma)=\Sigma_{\underline\sigma}\hat f(\underline\sigma)=1$.

Substituting

$$f(\underline\sigma)\equiv\int dx\,\rho(x)\frac{\exp\left(\beta x\sum_{a=1}^{m}\sigma^a\right)}{(2\cosh\beta x)^m},$$

$$\times\hat f(\underline\sigma)\equiv\int dy\,\hat\rho(y)\frac{\exp\left(\beta y\sum_{a=1}^{m}\sigma^a\right)}{(2\cosh\beta y)^m},\tag{A5}$$

we obtain

$$\beta\phi[\rho,\hat\rho]=\frac{\overline{l}p}{m}\ln\left[\int d\rho(x)d\hat\rho(y)(1+\tanh\beta x\tanh\beta y)^m\right]-\frac{\overline{l}}{\overline{k}}\ln\left(\frac{1+e^{-2\beta}}{2}\right)-\frac{\overline{l}}{\overline{k}m}\sum_{\nu=0}^{\infty}g_\nu\ln\left[\int\prod_{i=1}^{\nu}d\rho(x_i)\times(1+\tanh\beta\tanh\beta x_1\cdots\tanh\beta x_\nu)^m\right]-\frac{p}{m}\sum_{l=2}^{\infty}v_l\ln\left\{\int\prod_{i=1}^{l}d\hat\rho(y_i)\left[\prod_{i=1}^{l}(1+\tanh\beta y_i)+\prod_{i=1}^{l}(1-\tanh\beta y_i)\right]\right\}\tag{A6}$$

and the corresponding saddle point equations:

$$\frac{\rho(x)}{(2\cosh\beta x)^m}=\frac{1}{\mathbb{Z}\overline{l}}\sum_{l=2}^{\infty}v_l l B_l^{-1}\int\prod_{i=1}^{l-1}\frac{d\hat\rho(y_i)}{(2\cosh\beta y_i)^m}\times\delta\left(x-\sum_{i=1}^{l-1}y_i\right),\tag{A7}$$

$$\hat\rho(y)=\frac{1}{\mathbb{Q}}\sum_{\nu=1}^{\infty}f_{\nu-1}A_\nu^{-1}\int\prod_{i=1}^{\nu-1}d\rho(y_i)\,\delta\left(y-\frac{1}{\beta}\text{arctanh}\times[\tanh\beta\tanh\beta y_1\cdots\tanh\beta y_{\nu-1}]\right),\tag{A8}$$

where $f_{\nu-1}\equiv g_\nu\nu/(p\overline{k})$, and

$$B_l\equiv\int\prod_{i=1}^{l}d\hat\rho(y_i)\left[\prod_{i=1}^{l}(1+\tanh\beta y_i)+\prod_{i=1}^{l}(1-\tanh\beta y_i)\right]^m,\tag{A9}$$

$$A_\nu\equiv\int\prod_{i=1}^{\nu}d\rho(x_i)[1+\tanh\beta\tanh\beta x_1\cdots\tanh\beta x_\nu]^m.\tag{A10}$$

The constants $\mathbb{Z}$ and $\mathbb{Q}$ can be chosen to enforce the normalization condition $\int d\rho(x)=\int d\hat\rho(y)=1$.

In the $\beta\to\infty$ limit, we keep $m\beta=\mu$ fixed and adopt the dollowing ansatz for $\rho(x)$ and $\hat\rho(y)$:

$$\rho(x)=\sum_{q=-\infty}^{+\infty}\rho_q\delta(x-q),$$

$$\hat\rho(y)=\hat\rho_+\delta(y-1)+\hat\rho_0\delta(y)+\hat\rho_-\delta(y+1).\tag{A11}$$

Moreover, we define $\rho_+\equiv\Sigma_{q>0}\rho_q$ and $\rho_-=\Sigma_{q<0}\rho_q$.

We finally obtain the following expression for the free energy:

$$\phi(\mu)=\frac{\overline{l}p}{\mu}\ln\{1+(e^{-2\mu}-1)[\rho_+\hat\rho_-+\rho_-\hat\rho_+]\}-\frac{\overline{l}}{\overline{k}\mu}\sum_{\nu=0}^{\infty}g_\nu\ln\left\{1+\frac{1}{2}(e^{-2\mu}-1)[(\rho_++\rho_-)^\nu-(\rho_+-\rho_-)^\nu]\right\}-\frac{p}{\mu}\sum_{l=2}^{\infty}v_l\times\ln\left\{\sum_{n_+,n_0,n_-}{}'\frac{l!}{n_+!n_0!n_-!}\hat\rho_+^{n_+}\hat\rho_0^{n_0}\hat\rho_-^{n_-}\times e^{-2\mu\min(n_+,n_-)}\right\},\tag{A12}$$

the sum Σ' being restricted to the integers $n_+,n_0,n_-\geqslant0$ such that $n_++n_0+n_-=l$. The saddle point equations are

$$\hat{\rho}_+ = \frac{1}{2\mathbb{Q}} \sum_{\nu=1}^{\infty} f_{\nu-1} A_\nu^{-1} [(\rho_+ + \rho_-)^{\nu-1} + (\rho_+ - \rho_-)^{\nu-1}], \tag{A13}$$

$$\hat{\rho}_- = \frac{1}{2\mathbb{Q}} \sum_{\nu=1}^{\infty} f_{\nu-1} A_\nu^{-1} [(\rho_+ + \rho_-)^{\nu-1} - (\rho_+ - \rho_-)^{\nu-1}], \tag{A14}$$

$$\rho_+ = \frac{1}{\mathbb{Z}l} \sum_{l=2}^{\infty} v_l l B_l^{-1} \sum_{n_+ > n_- ; n_0} \frac{(l-1)!}{n_+! n_0! n_-!} \times \hat{\rho}_+^{n_+} \hat{\rho}_0^{n_0} \hat{\rho}_-^{n_-} e^{-2\mu n_-} \delta_{n_+ + n_0 + n_-, l-1}, \tag{A15}$$

$$\rho_- = \frac{1}{\mathbb{Z}l} \sum_{l=2}^{\infty} v_l l B_l^{-1} \sum_{n_- > n_+ ; n_0} \times \frac{(l-1)!}{n_+! n_0! n_-!} \hat{\rho}_+^{n_+} \hat{\rho}_0^{n_0} \hat{\rho}_-^{n_-} e^{-2\mu n_+} \delta_{n_+ + n_0 + n_-, l-1}, \tag{A16}$$

where

$$A_\nu = 1 + \frac{1}{2}(e^{-2\mu} - 1)[(\rho_+ + \rho_-)^\nu - (\rho_+ - \rho_-)^\nu], \tag{A17}$$

$$B_l = \sum_{n_+, n_0, n_-} \frac{l!}{n_+! n_0! n_-!} \times \hat{\rho}_+^{n_+} \hat{\rho}_0^{n_0} \hat{\rho}_-^{n_-} e^{-2\mu \min(n_+, n_-)} \delta_{n_+ + n_0 + n_-, l}, \tag{A18}$$

$$\mathbb{Q} = \sum_{\nu=1}^{\infty} f_{\nu-1} A_\nu^{-1}, \tag{A19}$$

$$\mathbb{Z} = \frac{1}{l} \sum_{l=2}^{\infty} v_l l B_l^{-1} \sum_{n_+, n_0, n_-} \frac{(l-1)!}{n_+! n_0! n_-!} \times \hat{\rho}_+^{n_+} \hat{\rho}_0^{n_0} \hat{\rho}_-^{n_-} e^{-2\mu \min(n_+, n_-)} \delta_{n_+ + n_0 + n_-, l-1}. \tag{A20}$$

We look for a "glassy" (i.e., with $\rho_+, \rho_- > 0$) solution of Eqs. (A13)–(A16). Such a solution exists in some interval $\mu_1(p) < \mu < \mu_2(p)$. For $p < p^*$ no physical solution exists for any value of μ. For $p^* < p < p_d$, $0 = \mu_1(p) < \mu_2(p)$ and $\phi(\mu)$ is a monotonically increasing function between $\mu_1(p)$ and $\mu_2(p)$. A physical solution exists but we cannot associate with it any well-behaved complexity. Above p_d we have $0 < \mu_1(p) < \mu_2(p) = \infty$ and a "well-behaved" complexity can be computed by Legendre transforming $\mu\phi(\mu)$,[12] cf. Eq. (4.8). The complexity $\Sigma(\epsilon)$ is nonzero between ϵ_s and ϵ_d. At $p = p_c$ the static energy ϵ_s vanishes: more than one code word (more precisely, about $\exp[N\Sigma(0)]$ code words) is consistent with the received message.[13]

Beyond the factorized ansatz

The general one-step replica-symmetry-breaking order parameter [18] is

$$\lambda(\vec{\sigma}) = \int DQ[\rho] \prod_{\mathcal{G}=1}^{n/m} \left[\int d\rho(x) \frac{\exp\left(\beta x \sum_{a \in \mathcal{G}} \sigma^a\right)}{(2\cosh \beta x)^m} \right],$$

$$\hat{\lambda}(\vec{\sigma}) = \int D\hat{Q}[\hat{\rho}] \prod_{\mathcal{G}=1}^{n/m} \left[\int d\hat{\rho}(y) \frac{\exp\left(\beta y \sum_{a \in \mathcal{G}} \sigma^a\right)}{(2\cosh \beta y)^m} \right]. \tag{A21}$$

The saddle point equations for functional order parameters $Q[\rho]$ and $\hat{Q}[\hat{\rho}]$ are given in the following section for a general channel, cf. Eqs. (B3) and (B4).

In the preceding section we used a quasifactorized ansatz of the form

$$Q[\rho] = (1-p)\,\delta[\rho - \delta_\infty] + p\,\delta[\rho - \rho_0], \quad \hat{Q}[\hat{\rho}] = \delta[\hat{\rho} - \hat{\rho}_0], \tag{A22}$$

where $\delta[\cdot]$ is a functional delta function, and $\delta_\infty(x)$ is the ordinary Dirac delta centered at $x = +\infty$. This ansatz does not satisfy the saddle point equations (B3) and (B4), but yields very good approximate results.

Some exact results (within an 1RSB scheme) can be obtained by writing the general decomposition

$$Q[\rho] = u\, Q_s[\rho] + (1-u)\, Q_a[\rho],$$

$$\hat{Q}[\hat{\rho}] = \hat{u} \hat{Q}_s[\hat{\rho}] + (1-\hat{u}) \hat{Q}_a[\hat{\rho}], \tag{A23}$$

where $Q_s[\rho]$ and $\hat{Q}_s[\hat{\rho}]$ are concentrated on the subspace of symmetric distributions [for which $\rho(x) = \rho(-x)$, $\hat{\rho}(y) = \hat{\rho}(-y)$], while $Q_a[\rho]$ and $\hat{Q}_a[\hat{\rho}]$ have zero weight on this subspace. Using this decomposition in Eqs. (B3) and (B4),

[12] The situation around p_d is more complicate than the one we described. This is an artifact of the variational approximation we adopted for computing the 1RSB free energy. Here is a sketch of what happens. At $p \approx 0.419$ a maximum of $\phi(\mu)$, which is still defined between 0 and $\mu_2(p) < \infty$, appears. At $p \approx 0.424$ the function $\phi(\mu)$ breaks down into two branches: a small μ [defined between 0 and $\mu_1(p) > 0$] and a large μ [defined between $\mu_1(p)$ and $\mu_2(p) < \infty$] continuation. This second branch has a maximum for some μ^*. At $p \approx 0.427\,15$, $\mu_2(p) \to \infty$. This threshold can be computed by studying the asymptotic problem defined by Eqs. (A13)–(A16) in the limit $\mu \to \infty$. Finally, at $p = p_d \approx 0.429\,440$, the small μ branch disappears.

[13] Once again, because of the variational approximation we made in computing $\phi(\mu)$, we obtain $\epsilon_s = 0$ above $p > p_c' \approx 0.486\,97$.

we get, for the BEC, a couple of equations for u and $\hat{u}$, which are identical to the replica symmetric ones, cf. Eq. (4.3).

The meaning of this result is clear. For $p > p_d$ the system decomposes into two parts. There exists a *core* which the iterative algorithms are unable to decode, and which is completely glassy. This part is described by the functionals $Q_s[\rho]$ and $\hat{Q}_s[\hat{\rho}]$. The rest of the system (the *peripheral* region) can be decoded by the belief propagation algorithm and, physically, is strongly magnetized. This corresponds to the functionals $Q_a[\rho]$ and $\hat{Q}_a[\hat{\rho}]$ (a more detailed study shows that the asymmetry of ρ and $\hat{\rho}$ is, in this case, typically positive).

APPENDIX B: CALCULATIONS, THE GENERAL CHANNEL

In this appendix we give some details of the replica calculation for a general noisy channel [i.e., for a general distribution $p(h)$ of the random fields]. In contrast with the BEC case, cf Eqs. (A11), the local field distributions do not have a simple form even in the zero-temperature limit. Therefore our results are mainly based on a numerical solution of the saddle point equations.

1. Finite temperature

The one-step replica-symmetry-breaking ansatz is given in Eqs. (A21). Inserting in Eq. (3.5) and taking the $n \to 0$ limit, we get $S[\lambda,\hat{\lambda}] = n\phi[Q,\hat{Q}] + O(n^2)$, with

$$\begin{aligned}\phi[Q,\hat{Q}] = {} & \frac{\overline{l}}{m}\int DQ[\rho]\int D\hat{Q}[\hat{\rho}]\ln\Bigg\{\int d\rho(x)\int d\hat{\rho}(y) \\ & \times[1+t_\beta(x)t_\beta(y)]^m\Bigg\} \\ & -\frac{\overline{l}}{\overline{k}m}\sum_{k=3}^{\infty} c_k\int\prod_{i=1}^{k} DQ[\rho_i] \\ & \times\ln\Bigg\{\int\prod_{i=1}^{k} d\rho_i(x_i)[1+t_\beta t_\beta(x_1)\cdots t_\beta(x_k)]^m\Bigg\} \\ & -\frac{1}{m}\sum_{l=2}^{\infty} v_l\int\prod_{i=1}^{l} D\hat{Q}[\hat{\rho}_i] \\ & \times\Bigg\langle \ln\Bigg\{\int\prod_{i=1}^{l} d\hat{\rho}_i(y_i)\mathbb{F}_{l+1}\left(\frac{\hat{\zeta}h}{\beta},y_1,\ldots,y_l\right)^m\Bigg\}\Bigg\rangle \\ & \times_h - \langle\ln\cosh(\hat{\zeta}h)\rangle_h + \frac{\overline{l}}{\overline{k}}\ln(1+t_\beta), \end{aligned} \tag{B1}$$

where we used the shorthand $t_\beta(x) = \tanh(\beta x)$, $t_\beta = \tanh(\beta)$, and defined

$$\mathbb{F}_n(y_1,\ldots,y_n) \equiv \prod_{i=1}^{n}[1+t_\beta(y_i)] + \prod_{i=1}^{n}[1-t_\beta(y_i)]. \tag{B2}$$

The saddle point equations are

$$\begin{aligned} Q[\rho] = {} & \frac{1}{l}\sum_{l=2}^{\infty} v_l l\int dp(h)\int\prod_{i=1}^{l-1} D\hat{Q}[\hat{\rho}_i] \\ & \times\delta[\rho - \rho_h^{(l)}[\hat{\rho}_1,\ldots,\hat{\rho}_{l-1}]], \end{aligned} \tag{B3}$$

$$\hat{Q}[\hat{\rho}] = \frac{1}{k}\sum_{k=3}^{\infty} c_k k\int\prod_{i=1}^{k-1} DQ[\rho_i]\,\delta[\hat{\rho} - \hat{\rho}^{(k)}[\rho_1,\ldots,\rho_{k-1}]], \tag{B4}$$

where $\delta[\cdots]$ denotes the functional delta function, and $\rho_h^{(l)}[\cdots]$, $\hat{\rho}^{(k)}[\cdots]$ are defined as follows:

$$\begin{aligned} \frac{\rho_h^{(l)}(x)}{(2\cosh\beta x)^m} = {} & \frac{1}{\mathcal{Z}}\int\prod_{i=1}^{l-1}\frac{d\hat{\rho}_i(y_i)}{(2\cosh\beta y_i)^m} \\ & \times\delta\left(x - \frac{\hat{\zeta}h}{\beta} - y_1 - \cdots - y_{l-1}\right), \end{aligned} \tag{B5}$$

$$\begin{aligned} \hat{\rho}^{(k)}(y) = {} & \int\prod_{i=1}^{k-1} d\rho_i(x_i)\,\delta\Bigg[y - \frac{1}{\beta}\text{arctanh} \\ & \times[t_\beta t_\beta(x_1)\cdots t_\beta(x_{k-1})]\Bigg]. \end{aligned} \tag{B6}$$

These equations can be solved numerically using the population dynamics algorithm of Ref. [19]. Some outcomes of this approach are reported in Sec. V A 1.

2. Zero temperature

In this appendix we compute the number of metastable states having a fixed overlap with a random configuration[14] $\underline{\sigma}^{\text{out}}$. The dynamical and statical thresholds for the BSC can be deduced from the results of this computation, cf. Sec. V B. The generalization to other statistical models for the noisy channel is straightforward (but slightly cumbersome from the point of view of notation).

In order to study the existence of metastable states, we consider the constrained partition function

$$Z(q;\underline{\sigma}^{\text{out}}) = \sum_{\underline{\sigma}} e^{-\beta H_{\text{exch}}(\underline{\sigma})}\,\delta\left(Nq - \sum_{i=1}^{N}\sigma_i^{\text{out}}\sigma_i\right), \tag{B7}$$

where the received bits σ_i^{out} are i.i.d. quenched variables: $\sigma_i^{\text{out}} = +1$ (-1) with probability $1-p$ (p). We introduce m “real” weakly coupled replicas of the system:

[14] Notice that such states are not necessarily stable with respect to moves that change their overlap with $\underline{\sigma}^{\text{out}}$.

$$Z_m(q;\underline{\sigma}^{\mathrm{out}})=\int_{-i\infty}^{+i\infty}\prod_{a=1}^{m}\beta\frac{dh_a}{2\pi}\exp\left(-Nq\sum_a h_a\right)\times\sum_{\{\underline{\sigma}^a\}}\exp\left[-\beta\sum_{a=1}^{m}H_{\mathrm{exch}}(\underline{\sigma}^a)+\beta\sum_{a=1}^{m}\sum_{i=1}^{N}h_a\sigma_i^{\mathrm{out}}\sigma_i^a\right]. \qquad \text{(B8)}$$

For a general channel we should look at the likelihood rather than at the overlap.

We make the hypothesis of symmetry among the m coupled replicas. In particular, we use the same value of the Lagrange multiplier for all of them: $h_a=h_0/\beta m$. We are therefore led to compute

$$\phi(m;h_0)=-\lim_{n\to 0}\frac{1}{n}\ln\overline{\tilde{Z}_m(h_0;\underline{\sigma}^{\mathrm{out}})^{n/m}}, \qquad \text{(B9)}$$

where

$$\tilde{Z}_m(h_0;\underline{\sigma}^{\mathrm{out}})=\sum_{\{\underline{\sigma}^a\}}\exp\left[-\beta\sum_{a=1}^{m}H_{\mathrm{exch}}(\underline{\sigma}^a)+(h_0/m)\sum_{a=1}^{m}\sum_{i=1}^{N}\sigma_i^{\mathrm{out}}\sigma_i^a\right]. \qquad \text{(B10)}$$

Next we take the zero-temperature limit keeping $m\beta=\mu$ fixed. With a slight change of notation, we have $m\phi(m;h_0)\to\mu\phi(\mu;h_0)$. The entropy of metastable states, cf. Eq. (5.3), is obtained as the Legendre transform of $\mu\phi(\mu;h_0)$:

$$\Sigma_p(\epsilon,q)=\mu\epsilon-h_0q-\mu\phi(\mu;h_0), \qquad \text{(B11)}$$

with $\epsilon=\partial_\mu[\mu\phi(\mu;h_0)]$ and $q=-\partial_{h_0}[\mu\phi(\mu;h_0)]$.

The replica expression for $\phi(\mu;h_0)$ is easily obtained by taking the zero temperature limit on the results of the preceding section. The free energy reads [for sake of simplicity we write it for a regular (k,l) code; the generalization is trivial by making use of Eq. (B1)]

$$\mu\phi[Q,\hat{Q}]=l\int DQ[\rho]\int D\hat{Q}[\hat{\rho}]\ln\left\{1+\int d\rho(x)\times\int d\hat{\rho}(y)\,\theta(-xy)[e^{-2\mu\min(|x|,|y|)}-1]\right\}-\frac{l}{k}\int\prod_{i=1}^{k}DQ[\rho_i]\ln\left\{1+\int\prod_{i=1}^{k}d\rho_i(x_i)\,\theta(-x_1\cdots x_k)\times[e^{-2\mu\min(1,|x_1|,\dots,|x_k|)}-1]\right\}-\int\prod_{i=1}^{l}D\hat{Q}[\hat{\rho}_i]\sum_{\sigma^{\mathrm{out}}}p_{\sigma^{\mathrm{out}}}\ln\left\{\int\prod_{i=1}^{l}d\hat{\rho}_i(y_i)\times\exp[-2\mu\mathbb{E}_{\sigma^{\mathrm{out}}}(y_1\cdots y_l)]\right\}-h_0, \qquad \text{(B12)}$$

where $p_{\sigma^{\mathrm{out}}}=1-p$ for $\sigma^{\mathrm{out}}=+1$, and $p_{\sigma^{\mathrm{out}}}=p$ for $\sigma^{\mathrm{out}}=-1$ and

$$\mathbb{E}_\sigma(y_1,\dots,y_l)=\min\left[\sum_{i:y_i\sigma<0}|y_i|;h_0/\mu+\sum_{i:y_i\sigma>0}|y_i|\right]. \qquad \text{(B13)}$$

The saddle point equations become in this limit

$$Q[\rho]=\frac{1}{l}\sum_{l=2}^{\infty}v_l l\sum_{\sigma^{\mathrm{out}}}p_{\sigma^{\mathrm{out}}}\int\prod_{i=1}^{l-1}D\hat{Q}[\hat{\rho}_i]\times\delta[\rho-\rho^{(l)}_{\sigma^{\mathrm{out}}}[\hat{\rho}_1,\dots,\hat{\rho}_{l-1}]], \qquad \text{(B14)}$$

$$\hat{Q}[\hat{\rho}]=\frac{1}{k}\sum_{k=3}^{\infty}c_k k\int\prod_{i=1}^{k-1}DQ[\rho_i]\,\delta[\hat{\rho}-\hat{\rho}^{(k)}[\rho_1,\dots,\rho_{k-1}]]. \qquad \text{(B15)}$$

The functionals $\rho^{(l)}_{\sigma^{\mathrm{out}}}[\cdots]$, $\hat{\rho}^{(k)}[\cdots]$ are defined as follows:

$$\rho^{(l)}_{\sigma^{\mathrm{out}}}(x)=\frac{1}{\mathcal{Z}}\int\prod_{i=1}^{l-1}d\hat{\rho}_i(y_i)\exp\left(\mu|x|-\mu\sum_i|y_i|\right)\times\delta(x-(h_0/\mu)\sigma^{\mathrm{out}}-y_1-\cdots-y_{l-1}), \qquad \text{(B16)}$$

$$\hat{\rho}^{(k)}(y)=\int\prod_{i=1}^{k-1}d\rho_i(x_i)\,\delta[y-\mathrm{sgn}(x_1\cdots x_{k-1})\times\min(1,|x_1|,\dots,|x_{k-1}|)]. \qquad \text{(B17)}$$

[1] Theor. Comput. Sci. **265** (1-2) (2001), special issue on phase transitions in combinatorial problems, edited by O. Dubois, R. Monasson, B. Selman and R. Zecchina.

[2] J. -P. Bouchaud, L. F. Cugliandolo, J. Kurchan, and M. Mézard, in *Spin Glasses and Random Fields,* edited by A. P. Young (World Scientific, Singapore, 1997).

[3] A. Barg, in *Handbook of Coding Theory,* edited by V. S. Pless and W. C. Huffman (Elsevier Science, Amsterdam, 1998).

[4] D. A. Spielman, Lect. Notes Comput. Sci. **1279**, 67 (1997).

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[6] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding* (McGraw-Hill, New York, 1979).

[7] R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York, 1968).

[8] C.E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).

[9] N. Sourlas, Nature (London) **339**, 693 (1989).

[10] N. Sourlas, in *Statistical Mechanics of Neural Networks*, edited by L. Garrido, Lecture Notes in Physics Vol. 368 (Springer, New York, 1990).

[11] N. Sourlas, in *From Statistical Physics to Statistical Inference and Back*, edited by P. Grassberger and J.-P. Nadal (Kluwer Academic, Dordrecht, 1994).

[12] P. Ruján, Phys. Rev. Lett. **70**, 2968 (1993).

[13] N. Sourlas, Europhys. Lett. **25**, 159 (1994).

[14] C. Berrou, A. Glavieux, and P. Thitimajshima, in *Proceedings of the 1993 International Conference Comm.* (IEEE, Piscataway, NJ, 1993), pp.1064–1070.

[15] R. G. Gallager, *Low Density Parity-Check Codes* (MIT Press, Cambridge, MA, 1963).

[16] D. J. C. MacKay, IEEE Trans. Inf. Theory **45**, 399 (1999).

[17] S. M. Aji, G. B. Horn, D. MacKay, and R. J. McEliece, in *Codes, Systems, and Graphical Models*, edited by B. Marcus and J. Rosenthal (Springer, New York, 2001).

[18] R. Monasson, J. Phys. A **31**, 513 (1998).

[19] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).

[20] S. Franz, M. Leone, F. Ricci-Tersenghi, and R. Zecchina, Phys. Rev. Lett. **87**, 127209 (2001).

[21] S. Franz, M. Mézard, F. Ricci-Tersenghi, M. Weigt, and R. Zecchina, Europhys. Lett. **55**, 465 (2001).

[22] J. S. Yedidia, W. T. Freeman, and Y. Weiss, in *Advances in Neural Information Processing Systems*, edited by T. K. Leen, T. G. Dietterich, and V. Tresp (MIT Press, Cambridge, MA, 2001), Vol. 13.

[23] J. S. Yedidia, W. T. Freeman, and Y. Weiss, MERL Technical Report No. TR 2001-22 (unpublished), available at http://www. merl. com/papers/TR2001-22.

[24] I. Kanter and D. Saad, Phys. Rev. Lett. **83**, 2660 (1999).

[25] R. Vicente, D. Saad, and Y. Kabashima, Phys. Rev. E **60**, 5352 (1999).

[26] I. Kanter and D. Saad, Phys. Rev. E **61**, 2137 (1999).

[27] A. Montanari and N. Sourlas, Eur. Phys. J. B **18**, 107 (2000).

[28] A. Montanari, Eur. Phys. J. B **18**, 121 (2000).

[29] Y. Kabashima, T. Murayama, and D. Saad, Phys. Rev. Lett. **84**, 1355 (2000).

[30] I. Kanter and D. Saad, J. Phys. A **33**, 1675 (2000).

[31] R. Vicente, D. Saad, and Y. Kabashima, Europhys. Lett. **51**, 698 (2000).

[32] A. Montanari, Eur. Phys. J. B **23**, 121 (2001).

[33] Y. Kabashima, N. Sazuka, K. Nakamura, and D. Saad, e-print cond-mat/0010173.

[34] T. Richardson and R. Urbanke, in *Codes, Systems, and Graphical Models* (Ref. [17]).

[35] R. M. Tanner, IEEE Trans. Inf. Theory **27**, 533 (1981).

[36] G. D. Forney, Jr., IEEE Trans. Inf. Theory **47**, 520 (2001).

[37] S.-Y. Chung, G.D. Forney, Jr., T.J. Richardson, and R. Urbanke, IEEE Commun. Lett. **5**, 58 (2001).

[38] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, IEEE Trans. Inf. Theory **47**, 569 (2001).

[39] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, IEEE Trans. Inf. Theory **47**, 585 (2001).

[40] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference* (Kaufmann, San Francisco, 1988).

[41] Y. Kabashima and D. Saad, Europhys. Lett. **44**, 668 (1998).

[42] T. Richardson and R. Urbanke, IEEE Trans. Inf. Theory **47**, 599 (2001).

[43] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[44] See, for instance, http://www. digitalfountain. com/technology/index. htm.

[45] C. Di, D. Proietti, E. Telatar, T. Richardson, and R. Urbanke, IEEE Trans. Inf. Theory (to be published).

[46] R. Monasson, Phys. Rev. Lett. **75**, 2847 (1995).

[47] S. Franz and G. Parisi, J. Phys. I **5**, 1401 (1995).

[48] M. Sipser and D. A. Spielman, IEEE Trans. Inf. Theory **42**, 1710 (1996).

[49] S. Kirkpatrick, C. D. Vecchi, and M. P. Gelatt, Science **220**, 671 (1983).

[50] B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[51] A. Cavagna, J. P. Garrahan, and I. Giardina, J. Phys. A **32**, 711 (1999).

[52] G. Biroli, R. Monasson, and M. Weigt, Eur. Phys. J. B **14**, 551 (2000).

[53] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002); M. Mézard and R. Zecchina, e-print cond-mat/0207194.

[54] A. Montanari and F. Ricci-Tersenghi, e-print cond-mat/0207416.