

Statistical mechanics of error exponents for error-correcting codes

Thierry Mora

Laboratoire de Physique Théorique et Modèles Statistiques, Bât. 100, Université Paris-Sud, F-91405 Orsay, France

Olivier Rivoire

Laboratory of Living Matter, The Rockefeller University, 1230 York Avenue, Box 34, New York, New York 10021, USA

(Received 27 June 2006; published 15 November 2006)

Error exponents characterize the exponential decay, when increasing message length, of the probability of error of many error-correcting codes. To tackle the long-standing problem of computing them exactly, we introduce a general, thermodynamic, formalism that we illustrate with maximum-likelihood decoding of low-density parity-check codes on the binary erasure channel and the binary symmetric channel. In this formalism, we apply the cavity method for large deviations to derive expressions for both the average and typical error exponents, which differ by the procedure used to select the codes from specified ensembles. When decreasing the noise intensity, we find that two phase transitions take place, at two different levels: a glass to ferromagnetic transition in the space of codewords and a paramagnetic to glass transition in the space of codes.

DOI: [10.1103/PhysRevE.74.056110](https://doi.org/10.1103/PhysRevE.74.056110)

PACS number(s): 89.90+n, 89.70+c, 05.50+q

I. INTRODUCTION

Communicating information requires a physical channel whose inherent noise impairs the transmitted signals. Reliability can be improved by adding redundancy to the messages, thus allowing the receiver to correct the effects of the noise. This procedure has the drawbacks of increasing the cost of generating and sending the messages and of decreasing the speed of transmission. At first sight, better accuracy seems achievable only at the expense of lesser efficiency. Remarkably, Shannon showed that, in the limit of infinite-length messages, error-free communication is possible using only limited redundancy [1]. His proof of principle has triggered many efforts to construct actual error-correcting schemes that would approach the theoretical bounds. A renewal of interest in the subject has taken place during the last ten years, as new error-correcting codes were finally discovered [2], or rediscovered [3], which showed practical performances close to Shannon's bounds.

In this paper, we analyze a major family of such codes, the low-density parity-check (LDPC) codes, also known as Gallager codes, from the name of their inventor [4]. Our focus is on the characterization of rare decoding errors, in situations where most realizations of the noise are accurately corrected. Error-free communication, as guaranteed by Shannon's theorem, indeed results from a law of large number and is achieved only with infinite-length messages. Accordingly, any error-correcting scheme acting on finite-length messages has a nonzero error probability, which generically vanishes exponentially with the message length. Such error probabilities are described by *error exponents*, giving their rate of exponential decay. Two kinds of error exponents are usually distinguished: *average* error exponents, where the average is taken over an ensemble of codes, and *typical* error exponents, where the codes are typical elements of their ensemble.

The study of error exponents attracted early on considerable attention in the information theory community, but exact expressions have turned out to be particularly difficult to derive (see, e.g., [5] and [6] for concise and nontechnical

reviews with entries in the literature). Exact asymptotic results are known in the limit of the so-called *random linear model* [7] (presented in Appendix B), but only loose bounds (presented in Appendix C) have been established for more general codes. Recently, a systematic finite-length analysis of LDPC codes under iterative decoding was carried out for the binary erasure channel (BEC) [8,9], yielding exact, yet non-explicit, formulas for the average error probability. Up to now, little has, however, been known of the error probability under maximum-likelihood decoding, except for the work of [10] dealing with the binary symmetric channel (BSC).

We address here the problem of computing error exponents of LDPC codes under maximum-likelihood decoding, over both the BEC and BSC (all the necessary definitions are recalled below). We adopt a statistical physics point of view, which exploits the well-established [11] mapping between error-correcting codes and spin glasses [12]. A thermodynamic formalism is introduced where error exponents are expressed as large deviation functions [13], which we compute by means of the extension of the cavity method [14] proposed in [15]. This approach offers an alternative to the related replica method employed in [10] and allows us to address both average and typical error exponents. We thus obtain an interesting phase diagram, with two very distinct phase transitions occurring when the intensity of the noise in the channels is varied.

A brief summary of our results can be found in [16]. We present in what follows a much more detailed account of our approach. In a first part, we define LDPC codes, recall their mapping to some models of spin glasses and optimization problems, and give a general overview of our thermodynamic (large deviation) formalism. The two subsequent parts apply this framework to the analysis of LDPC codes over the BEC and BSC, respectively. We sum up our results in a conclusion where we also point out some open questions. Most of the technical calculations are relegated to the Appendixes, which also contain a detailed discussion of the limiting case of random linear codes.

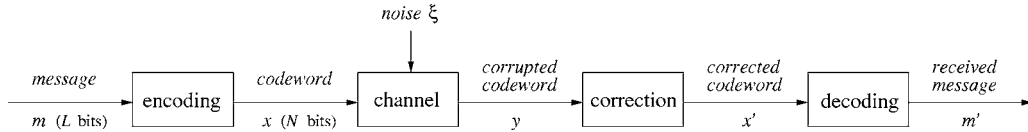


FIG. 1. Error correction scheme. A message m composed of L bits, $m \in \{0, 1\}^L$, is first encoded in a codeword of longer size N with $R=L/N < 1$, defining the rate of the code. The noise ξ of the channel corrupts the transmitted codeword which becomes y (see Fig. 2 for examples of channels). This output is generically not a codeword, and the correction consists in inferring the most probable codeword to which it comes from. Finally, the inferred codeword x' is converted back into its corresponding message m' . The communication is successful if $m' = m$.

II. ERROR-CORRECTING CODES AND THE LARGE DEVIATION FORMALISM

A. Error-correcting codes

Error-correcting codes are based on the idea that adding sufficient redundancy to the messages can allow the receiver to reconstruct them, even if they have been partially corrupted by the noisy channel [17]. A schematic view of how these codes operate is presented in Fig. 1. Given a *message* composed of L bits, an encoding map $\{0, 1\}^L \rightarrow \{0, 1\}^N$ first introduces redundancy by converting the L bits of the message into a longer sequence of N bits, called a *codeword*. The ratio $R \equiv L/N$ defines the *rate* of the code and should ideally be as large as possible to reduce communication costs, yet small enough to allow for corrections. Corrections are implemented downstream the noisy channel and specified by a decoding map $\{0, 1\}^N \rightarrow \{0, 1\}^L$ whose purpose is to reconstruct the original message from the received corrupted codeword. Decoding is composed of two steps: first, the most probable codeword is inferred, and second, it is converted into its corresponding message.

In this scheme, messages and codewords are related by the one-to-one encoding map, and translating messages into codewords or conversely is relatively straightforward. The computationally most demanding part is concentrated on inferring the most probable codeword sent, given the corrupted codeword received. In what follows, we shall focus exclusively on this problem, which requires manipulating only codewords.

B. Communication channels

Formally, a noisy channel is characterized by a transition probability $Q(\mathbf{y}|\mathbf{x})$ giving the probability for its output to be \mathbf{y} given that its input was \mathbf{x} . For the sake of simplicity, we confine ourselves to *memoryless* channels where the noise affects each bit independently of the others—i.e., $Q(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^N Q(y_i|x_i)$ with $Q(y_i|x_i)$ independent of i .

We shall consider more specifically two examples of memoryless channels. The first one is the *binary erasure channel* where a bit is erased with probability p —that is, $Q(*|x) = p$ and $Q(x|x) = 1-p$ where $*$ represents an erased bit (see Fig. 2). The second is the *binary symmetric channel* where a bit is flipped with probability p —that is, $Q(0|1) = Q(1|0) = p$ and $Q(0|0) = Q(1|1) = 1-p$ (see Fig. 2).

C. LDPC codes and code ensembles

Shannon first formalized the problem of error correction and determined the lowest achievable rate R allowing error-

free correction [1]. He found a general expression for this limit, called the *channel capacity*, which depends only on the nature of the channel and takes the form $C_{\text{BEC}}(p) = 1-p$ and $C_{\text{BSC}}(p) = 1-p \ln p - (1-p) \ln(1-p)$ for the BEC and BSC, respectively. Shannon's proof for the existence of codes achieving the channel capacity was nonconstructive and his analysis restricted to the limit of infinitely long messages, $L \rightarrow \infty$. Among the various families of codes proposed to practically perform error correction, one of the most promising is the family of *low-density parity-check* codes [4].

A LDPC code is defined by a sparse matrix A where “sparse” means that A is mostly composed of 0's, with otherwise a few 1's. The *parity-check matrix* A has size $M \times N$ with $M = N - L$ and is associated with a *generator matrix* G of size $L \times N$ such that $GA = 0$ (see, e.g., [3] for explicit constructions); the encoding map is taken to be the linear map $x = Gm$ and the rate of the code is $R = L/N = 1 - M/N$. By construction, an N -bit codeword x satisfies the M parity-check equations $Ax = 0$, or, in other words, the set of codewords is the kernel of A . The parity-check matrix A is usually represented graphically by a *factor graph*, as in Fig. 3: the columns of A are associated with *check nodes* labeled with $a \in \{1, \dots, M\}$ and represented by squares, and the lines of A are associated with *variable nodes* labeled with $i \in \{1, \dots, N\}$ and represented by circles. A nonzero element of the matrix A such as $A_{ia} = 1$ appears as a link between the variable node i and the check node a .

A particularly powerful approach for analyzing error-correcting codes is the probabilistic method where, instead of considering a single code, one studies an *ensemble* of codes. With LDPC codes, code ensembles correspond to sets of matrices or, equivalently, sets of factor graphs. A popular choice is to consider the ensemble of factor graphs with given connectivities c_k and v_ℓ , which is the set of factor graphs having $c_k M$ check nodes with connectivity k and $v_\ell N$ variable nodes with connectivity ℓ , where $\sum_k c_k = \sum_\ell v_\ell = 1$. A convenient representation is by means of the generating

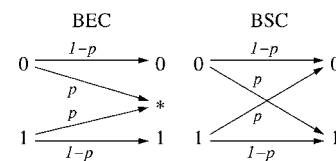


FIG. 2. Communication channels. On the left the BEC (binary erasure channel) erases a bit with probability p and leaves it unchanged with probability $1-p$. On the right the BSC (binary symmetric channel) flips a bit with probability p and leaves it unchanged with probability $1-p$.

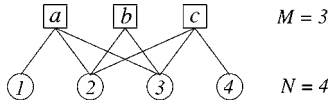


FIG. 3. Factor graph (Tanner graph [18]). The circles represent the variable nodes, associated with the N bits $\{x_i\}$, and the squares represent the M parity check. In the example given, the constraints read: (a) $x_1+x_2+x_3=0$, (b) $x_2+x_3=0$, and (c) $x_2+x_3+x_4=0$ (modulo 2).

functions $c(x)=\sum_k c_k x^k$ and $v_\ell=\sum_\ell v_\ell x^\ell$; these notations allow one, for instance, to write the mean connectivities as $\langle k \rangle = c'(1)$ and $\langle \ell \rangle = v'(1)$. Due to their simplicity, particular attention will be devoted to *regular* codes, whose check nodes have all same degree k and variable nodes same degree ℓ , corresponding to $c_{k'}=\delta_{k,k'}$ and $v_{\ell'}=\delta_{\ell,\ell'}$ or, equivalently, $c(x)=x^k$ and $v(x)=x^\ell$.

The mathematical fact underlying the probabilistic method is the phenomenon of *measure concentration* which occurs in the limit where $N \rightarrow \infty$ and $M \rightarrow \infty$ with fixed ratio $\alpha=M/N$: in this limit, many properties are shared by *almost all* elements of the ensemble (i.e., all but a subset of measure zero). As a consequence, by studying average properties over an ensemble, one actually has access to properties of typical elements of this ensemble. This fact is one of the building blocks of random graph theory [19] and is also central to the physics of disordered systems where it is known as the *self-averaging property* [20].

While the factor graph representation makes obvious the connection between LDPC codes and random graph theory, it will also turn particularly fruitful to exploit the close ties of LDPC codes with both optimization problems [21] and spin-glass systems [20]. LDPC codes are indeed intimately related to a class of combinatorial optimization problems known as XORSAT problems where, given a sparse matrix A and a vector τ , one is to find solutions σ to the equation $A\sigma=\tau$. Although algorithmically relatively simple (Gauss method provides an answer in a time polynomial in the size of the matrix), XORSAT problems share many common features with notably more difficult, NP-complete [21], problems such as K -SAT. A recent physical approach to XORSAT problems makes use of their formal equivalence with a class of spin-glass systems known as p -spin models [22–24]. We shall follow this line of investigation and apply the *cavity method* [14,25] from spin-glass theory to analyze LDPC codes. We note that alternative, sometimes equivalent, physical approaches have previously been applied to LDPC codes; we refer the reader to [26] for a review of the subject.

The distinctive feature of XORSAT at the root of its computational simplicity is the presence of an underlying group symmetry that relates all solutions. In the context of LDPC codes, it corresponds to the fact that the set of codewords is the kernel of the parity-check matrix A ; we shall refer to the XORSAT problem $A\sigma=0$ whose solutions define the set of codewords as the *encoding constraint satisfaction problem* (CSP) of the LDPC code with check matrix A . The group symmetry has a number of interesting consequences which will crucially simplify the analysis.

Most of the interest in LDPC codes stems from the possibility to decode them using efficient, iterative algorithms

(described in Sec. III A 3). Unless otherwise stated, we shall, however, be here concerned with the theoretically simpler, yet computationally much more demanding, *maximum-likelihood decoding* procedure. It consists in systematically decoding a received message to the most probable codeword (a task that iterative algorithms are in some cases unable to perform, as recalled in Sec. III A 3).

Finally, it is interesting to note that in the limit where $\langle k \rangle, \langle \ell \rangle \rightarrow \infty$ with fixed ratio, LDPC codes define the *random linear model* (RLM) whose typical elements have been shown by Shannon to achieve the channel capacity. This particular limit, where many quantities can be computed by invoking only elementary combinatorial arguments, is discussed in detail in Appendix B.

D. Typical properties and phase transitions

The performance of a particular code over a given channel is measured by its error probability—i.e., the probability that it fails to correctly decode a corrupted codeword. More precisely, if $d(\mathbf{y})$ denotes the inferred codeword when \mathbf{x} is sent and \mathbf{y} received, one defines the *block error probability* for \mathbf{x} as

$$P_N^{(B)}(\mathbf{x}) = \sum_{\mathbf{y}} Q(\mathbf{y}|\mathbf{x}) 1_{d(\mathbf{y}) \neq \mathbf{x}} \quad (1)$$

and the average block error probability as

$$P_N^{(B)} = \mathbb{E}_{\mathbf{x}}[P_N^{(B)}(\mathbf{x})], \quad (2)$$

where $\mathbb{E}_{\mathbf{x}}$ denotes the expectation (average) over the set of codewords. With LDPC codes, this average is trivial since, due to the group symmetry, all codewords are equivalent, and $P_N^{(B)}(\mathbf{x})$ is independent of \mathbf{x} .

The concentration phenomenon alluded to above means here that $P_N^{(B)} \rightarrow p_B$ with $N \rightarrow \infty$ within a given code ensemble defined by generating functions $c(x)$ and $v(x)$. As the level of the noise p is increased, a phase transition is generically observed: a critical value p_c exists above which error-free correction is no longer possible ($p_B=0$ for $p < p_c$ and $p_B=1$ for $p > p_c$). The formalism to be presented in the next sections will yield in particular the value of p_c for given code ensembles and channels. Obviously, the presence of this phase transition indicates that, when using a channel with noise level p , one should choose a code from an ensemble for which $p < p_c$. The phase transition is, however, occurring only in the limit of infinite codewords (thermodynamic limit) whereas practical coding inevitably deals with finite N . This leads to the fact that the block error probability is not exactly zero, even in the regime $p < p_c$.

For a given code of finite but large block-length N , error can thus be caused by rare, atypical, realizations of the noise. Similarly, when picking a code at random from a code ensemble of finite size, one can observe properties differing from the typical properties predicted by the law of large numbers. We show in what follows how these two atypical features induced by finite-size effects can be analyzed in a common framework.

TABLE I. The analogy with spin glasses or, more generally, the statistical physics of disordered system with quenched disorder.

	Spin glass	Average	Typical	Multistep, step 1	Multistep, step 2
Disorder	Couplings J_{ij}		Typical codes \mathcal{C}^0	Codes \mathcal{C} at y	
Configurations	Spins $\{\sigma_i\}_i$	Noise+codes (ξ, \mathcal{C})	Noise ξ	Noise ξ	Codes \mathcal{C}
Observable	$E = \sum_{ij} J_{ij} \sigma_i \sigma_j$	$S_N(\xi, \mathcal{C})$	$S_N(\xi, \mathcal{C}^0)$	$S_N(\xi, \mathcal{C})$	$L_{\mathcal{C}}(s)$
Entropy	$s(e = E/N)$	$L_0(s = S_N/N)$	$L(s = S_N/N)$	$\mathcal{L}(\phi, x)$	
Temperature ⁻¹	$\beta = \partial_e s$	$x = \partial_s L_1$	$x = \partial_s L_0$	$x = \partial_s L$	$y = \partial_\phi \mathcal{L}$
Potential	$\beta f = \beta e - s$	$\phi_1 = xs - L_1$	$\phi_0 = xs - L_0$	$\phi = xs - L$	$\psi = y\phi - \mathcal{L}$

E. Large deviations

At this stage, it is useful to make explicit the three different levels of statistics involved in the analysis of error-correcting codes: (i) statistics over the codes \mathcal{C} in a defined code ensemble \mathcal{C} , (ii) statistics over the set of transmitted codewords \mathbf{x} of a particular code, and (iii) statistics over the noise ξ of the channel, with a specified p . For given \mathcal{C} , \mathbf{x} , and ξ , a fourth level of statistics is involved in the decoding process, over the possible codewords $y \in \{0, 1\}^N$ from which the received corrupted codeword originates. The group structure of the set of codewords of LDPC codes makes level (ii) trivial since all codewords are in fact equivalent (isomorphic). We will consequently ignore it and address only levels (i) and (iii).

The problem of evaluating the probability that, due to finite-size effects, a property differs from the typical case belongs to *large deviation* theory [13]. To give here a general presentation of the concepts and methods to be used, we assume that the success of the decoding is measured by a function $S_N(\xi, \mathcal{C})$ extensive in N and such that $S_N(\xi, \mathcal{C}) \leq 0$ if the code \mathcal{C} correctly decodes a message subject to noise ξ and $S_N(\xi, \mathcal{C}) > 0$ otherwise; in the next sections, we will show explicitly how such an observable can be defined with LDPC codes, for both the BEC and BSC channels. In terms of S_N , the decoding phase transition takes the following form: in the limit $N \rightarrow \infty$, the distribution of the density S_N/N concentrates around a typical value $s_{\text{typ}}(p)$ which verifies $s_{\text{typ}}(p) \leq 0$ if $p < p_c$, and $s_{\text{typ}}(p) > 0$ if $p > p_c$, where p denotes as before the level of noise of the channel (see Fig. 2 for examples).

For typical codes in their ensemble, denoted \mathcal{C}^0 , we describe large deviations of S_N with respect to the noise ξ by a *rate function* $L_0(s)$ such that the probability to observe $S_N(\xi, \mathcal{C}^0)/N = s$ satisfies

$$\mathbb{P}_N[\xi; S_N(\xi, \mathcal{C}^0)/N = s] \asymp e^{-NL_0(s)}. \quad (3)$$

Here the symbol $a_N \asymp b_N$ refers to an exponential equivalence, $\ln a_N / \ln b_N \rightarrow 1$ as $N \rightarrow \infty$. Viewed as a function of the noise level p , the rate function $E_{\text{typ}}(p) = L_0(s=0)$ is known in the coding literature as the *typical error exponent* [5]. The exponential decay with N of atypical properties is quite generic when dealing with large deviations, but this scaling is not necessarily ensured, as discussed in more detail in Appendix A. In the thermodynamic formalism that we shall

adopt, rate functions are computed by introducing a potential $\Phi_{\mathcal{C}}(x)$ defined by

$$\Phi_{\mathcal{C}}(x) = \ln(\mathbb{E}_{\xi} [e^{xS_N(\xi, \mathcal{C})}]). \quad (4)$$

In the limit $N \rightarrow \infty$ limit, the density $\Phi_{\mathcal{C}}(x)/N$ tends to a typical value $\phi_0(x)$, which is related to the rate function $L_0(s)$ by

$$e^{N\phi_0(x)} \asymp \int ds e^{N[xs - L_0(s)]}. \quad (5)$$

Equivalently, by taking the saddle point,

$$\phi_0(x) = xs - L_0(s), \quad x = \partial_s L_0(s). \quad (6)$$

The rate function $L_0(s)$ can thus be reconstructed from $\phi_0(x)$ by inverting the Legendre transformation,

$$L_0(s) = sx - \phi_0(x), \quad s = \partial_x \phi_0(x). \quad (7)$$

The analogy with the usual thermodynamics is summarized in Table I.

From a theoretical perspective, it is simpler to make an average over the codes and compute the rate function $L_1(s)$ defined as

$$\mathbb{P}_N[\xi, \mathcal{C}; S_N(\xi, \mathcal{C})/N = s] \asymp e^{-NL_1(s)}. \quad (8)$$

This procedure yields the so-called *average error exponent* $E_{\text{av}} = L_1(s=0)$. In the thermodynamical formalism, $L_1(s)$ is conjugated to the potential $\phi_1(x)$ satisfying

$$e^{N\phi_1(x)} = \mathbb{E}_{(\xi, \mathcal{C})} [e^{xS_N(\xi, \mathcal{C})}] = \int ds e^{N[xs - L_1(s)]}. \quad (9)$$

The two rate functions $L_0(s)$ and $L_1(s)$ may differ, meaning that the average exponent can be associated with atypical codes. Such atypical codes correspond themselves to large deviations of the potential $\Phi_{\mathcal{C}}(x)$. For fixed values of x , we define a rate function $\mathcal{L}(\phi, x)$ as

$$\mathbb{P}_N[\mathcal{C}; \Phi_{\mathcal{C}}(x)/N = \phi] \asymp e^{-N\mathcal{L}(\phi, x)}. \quad (10)$$

In a thermodynamic formalism, $\mathcal{L}(\phi, x)$ is again associated with a potential $\psi(x, y)$ defined by

TABLE II. Analogy with the replica approach of spin glasses. The replica-symmetric method prescribes that the typical partition function Z_0 of a disordered system is given by $Z_0 \sim \mathbb{E}[Z_N^n]^{1/n}$ with $n \rightarrow 0$ or, more precisely, if $\Lambda_N = \ln Z_N$, the typical value of $\lambda = \Lambda_N/N$ is $\lambda_0 = \lim_{n \rightarrow 0} \lim_{N \rightarrow \infty} (1/Nn) \ln \mathbb{E}[e^{n\Lambda_N}]$. This is mathematically justified by the Gärdner-Ellis theorem which moreover provides a rigorous basis for the interpretation of nonzero values of n in terms of large deviations, as discussed in the text. According to this theorem, if the function $\phi(x) = \lim_{N \rightarrow \infty} (1/N) \ln \mathbb{E}[e^{x\Lambda_N}]$ exists and is regular enough (see, e.g., [13] for a rigorous presentation), then a large deviation principle holds for λ with a rate function being the Legendre transform of $\phi(x)$; if we assume the functions differentiable, $L(\lambda) = \lambda x - \phi(x)$ with $\lambda = \partial_x \phi(x)$. As a corollary of this theorem, the typical value λ_0 , which by definition satisfies $L(\lambda_0) = 0$ and $x = \partial_\lambda L(\lambda_0) = 0$, is given by $\lambda_0 = \partial_x \phi(x=0) = \lim_{x \rightarrow 0} [\phi(x)/x]$ ($x=0$), as predicted by the replica method. Note also that $n=1$, with $Z_1 = \mathbb{E}[Z_N]$, corresponds to the so-called annealed approximation.

Replica (symmetric) theory of spin glasses	Multistep large deviations for LDPC codes
Hamiltonian $H_J[\sigma] = \sum_{ij} J_{ij} \sigma_i \sigma_j$	$S_N(\xi, \mathcal{C})$
Disorder $\{J_{ij}\}_{ij}$	Codes \mathcal{C}
Configurations $\{\sigma_i\}_i$	Noise ξ
Number of replicas n	Temperature $^{-1}$ y
Physical temperature $^{-1}$ β	Temperature $^{-1}$ x
Annealed approximation $n=1$	Average codes $y=1$
Quenched computation $n \rightarrow 0$	Typical codes $y \rightarrow 0$

$$e^{N\psi(x,y)} = \mathbb{E}_{\mathcal{C}}[\mathbb{E}_{\xi}[e^{xS_N(\xi, \mathcal{C})}]^y] = \mathbb{E}_{\mathcal{C}}[e^{y\Phi_{\mathcal{C}}(x)}] = \int d\phi e^{N[y\phi - \mathcal{L}(\phi, x)]}. \quad (11)$$

We refer to this hierarchical embedding of large deviations as a *multistep large deviation* structure [15], a term meant to reflect the formal equivalence with the multistep replica symmetry breaking scenario developed for spin glasses [20] (see Table II). In the limit $N \rightarrow \infty$ where the integral is dominated by its saddle point we obtain the Legendre transformation

$$\psi(x, y) = y\phi - \mathcal{L}(\phi, x), \quad y = \partial_{\phi} \mathcal{L}(\phi, x). \quad (12)$$

Within this extended framework, we recover the average case by taking $y=1$. Indeed, from the definitions (9) of $\phi_1(x)$ and (11) of $\psi(x, y)$ it follows that

$$e^{N\psi(x,1)} = \mathbb{E}_{\mathcal{C}}[\mathbb{E}_{\xi} e^{S_N(\xi, \mathcal{C})}] = \mathbb{E}_{(\xi, \mathcal{C})}[e^{xS_N(\xi, \mathcal{C})}] = e^{N\phi_1(x)}, \quad (13)$$

that is,

$$\psi(x, y=1) = \phi_1(x). \quad (14)$$

This average case differs in general from the typical case which corresponds to $y=0$. Indeed, by definition [see Eq. (10)], typical codes are associated with the potential ϕ_0 minimizing $\mathcal{L}(\phi, x)$, with $\mathcal{L}(\phi_0, x) = 0$, yielding $y = \partial_{\phi} \mathcal{L} = 0$. Note that the potential ϕ_0 is related to $\psi(x, y)$ by $\phi_0(x) = \lim_{y \rightarrow 0} (1/y) \psi(x, y)$, which can also be viewed as a corol-

lary of Gärdner-Ellis theorem [13], best known in statistical physics as the replica trick [20] (see Table II). In the language of the replica method, the average case ($y=1$) and the typical case ($y=0$) are, respectively, referred to as the annealed and quenched computations.

The previous discussion assumed that the potentials were analytical functions of their parameters x and y , but this may not be the case, and we will find that phase transitions can occur when these temperatures are varied. In such cases, taking naively the limit $y \rightarrow 0$ leads to erroneous results. We will discuss how to overcome such difficulties when encountering them.

III. LDPC CODES OVER THE BEC

We now proceed to illustrate our formalism with LDPC codes over the binary erasure channel. We start with rederiving the typical phase diagram by means of the cavity method, a slightly different approach than the replica method originally used in [27]. This sets the stage for the analysis of error exponents that follows.

A. Typical phase diagram

1. Formulation

Consider a LDPC code \mathcal{C} with parity-check matrix A ; its *encoding CSP* (the constraint satisfaction problem whose SAT assignments define the codewords) has cost function

$$H_{\mathcal{C}}[\sigma] = \sum_{a=1}^M E_a[\sigma], \quad \text{with } E_a[\sigma] = \sum_{i=1}^N A_{ai} \sigma_i \pmod{2}. \quad (15)$$

Since $E_a[\sigma] \in \{0, 1\}$, the cost function $H_{\mathcal{C}}[\sigma]$ counts the number of constraints violated by the assignment $\sigma = \{\sigma_i\}_{i=1, \dots, N}$ (where $\sigma_i \in \{0, 1\}$). When a codeword σ^* , satisfying $H_{\mathcal{C}}[\sigma^*] = 0$, goes through a BEC, each of its bits σ_i has probability p to be erased. A given realization of the noise can be characterized by a vector $\xi = (\xi_1, \dots, \xi_N)$ with $\xi_i = 1$ implying that the bit σ_i^* is lost and $\xi_i = 0$ that it is unaffected. If we denote by \mathcal{E} the set of indices i for which $\xi_i = 1$ (erased bits), the decoding task consists in reconstructing $\{\sigma_i^*\}_{i \in \mathcal{E}}$ from the received bits $\{\sigma_i^*\}_{i \notin \mathcal{E}}$ and knowledge of the encoding CSP $H_{\mathcal{C}}$. This decoding problem defines a new constraint satisfaction problem, the *decoding CSP*, obtained from the encoding CSP by fixing the values of the noncorrupted bits. More explicitly, the decoding CSP has cost function $H_{\mathcal{C}}^{(\xi)}[\sigma^{(\xi)}] = \sum_a E_a^{(\xi)}[\sigma^{(\xi)}]$ where $\sigma^{(\xi)} = \{\sigma_i\}_{i \in \mathcal{E}}$ and

$$E_a^{(\xi)}[\sigma^{(\xi)}] = \sum_{i \in \mathcal{E}} A_{ai} \sigma_i + \sum_{i \notin \mathcal{E}} A_{ai} \sigma_i^* \pmod{2}. \quad (16)$$

Decoding is possible if and only if $\{\sigma_i^*\}_{i \in \mathcal{E}}$ is the only SAT assignment of the decoding CSP.

If $\mathcal{N}_N(\xi, \mathcal{C})$ denotes the number of solutions of the decoding CSP, $S_N(\xi, \mathcal{C})$ can be taken as $S_N(\xi, \mathcal{C}) \equiv \ln \mathcal{N}_N(\xi, \mathcal{C})$. This entropy fulfills the desired properties: namely, $S_N(\xi, \mathcal{C}) \leq 0$ if decoding is successful, and $S_N(\xi, \mathcal{C}) > 0$ otherwise.

The particularity of LDPC codes compared to other error-correcting codes is that the decoding CSP has same form as the encoding CSP (both are XORSAT problems). As a consequence, the \mathbb{Z}_2 symmetry of the group of codewords is always preserved, at variance with what happens in other CSP's where fixing variables breaks a symmetry. The BEC is also particular compared with other channels, since the set \mathcal{E} of corrupted bits is known to the receiver (this will not be the case with the BSC, where identifying the corrupted bits is part of the decoding problem). This entails that bits can only be fixed to their correct value.

2. Cavity approach

Before considering large deviations, it is instructive to recall the typical results—i.e., the values taken by $S_N(\xi, \mathcal{C}^0)$ when \mathcal{C}^0 is a typical code from a given ensemble specified by $c(x)$ and $v(x)$, and ξ a typical realization of the noise from the probability distribution specified by p . We resort here to the *cavity method at zero temperature* [14], whose validity is based on the treelike structure of the factor graphs associated with typical LDPC codes. The essentially equivalent replica method has been used in the past: in [28], $S_N(\xi, \mathcal{C})$ is thus obtained by first computing a free energy with the replica method and then taking the zero-temperature limit to obtain $S_N(\xi, \mathcal{C})$, viewed as the entropy of the zero-energy ground states.

The approach we follow here, which corresponds to a particular implementation of the entropic cavity method presented in [29], has several advantages over the replica approach: it involves neither a zero-replica limit nor a zero-temperature limit, it emphasizes the specificities of LDPC codes associated with the underlying \mathbb{Z}_2 symmetry, and it naturally connects to the algorithmic analysis of single codes. In the common language of the replica and cavity methods, the calculation to be done is coined *one-step replica symmetry breaking* (1RSB) and the entropy $s=S_N/N$ is referred to as a *complexity*. This is reflected in what follows by the fact that we strictly restrict to SAT assignments and assume that all constraints are satisfied (the reweighting parameter μ , as denoted in [25], is here infinite, $\mu=\infty$). This 1RSB approach is known to exactly describe XORSAT problems [23,24].

Let $P_i(\sigma_i)$ be the probability, taken over the set of solutions of the decoding CSP, that the bit i assumes the value $\sigma_i \in \{0, 1\}$. Due to the preservation of the \mathbb{Z}_2 symmetry, no bit can be nontrivially biased: either it is fixed to 0 or 1, corresponding to $P_i=\delta_0$ and $P_i=\delta_1$, respectively, or it is completely free, corresponding to $P_i=(\delta_0+\delta_1)/2$, where we denote $\delta_r(\sigma)=\delta_{r,\sigma}$. In technical terms, the evanescent fields that are generically required to compute entropies in CSP [29] have here a trivial distribution, thus explaining that they can be safely ignored, as was done in [28].

Let ν be the probability, taken over the N nodes of a typical factor graph, that a bit i is free—i.e., that $P_i=(\delta_0+\delta_1)/2$. Since a free node has equal probability to be 0 or 1, its contribution to the entropy is $\ln 2$ and the mean entropic contribution per node is $\nu \ln 2$. This value is, however, only an upper bound (known as the annealed, or first moment,

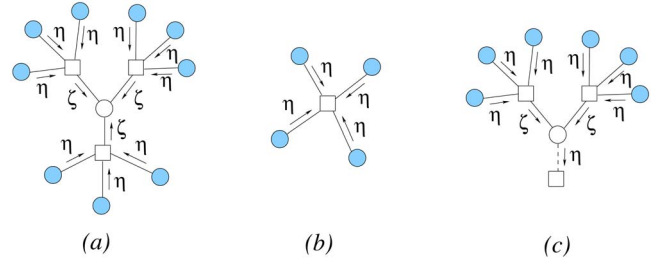


FIG. 4. (Color online) Illustration of cavity fields: (a) addition of a variable node, (b) addition of a parity check, and (c) cavity iteration.

bound) on the entropy density $s=S_N/N$ that we wish to calculate. In fact, it holds only if the bits are independent: indeed, two bits may both be free but, by fixing one, the second may be constrained to a unique value, in which case the joint entropic contribution of the two nodes is $\ln 2$ and not $2 \ln 2$. The correct expression is given by the Bethe formula, which can be heuristically derived as follows. First, we sum the entropic contributions $\Delta S_{\circ+\square \in \circ}$ of each node \circ , including the corrections due to its adjacent parity checks $\square \in \circ$. Second, we note that each parity check \square is involved in k_\square terms, with k_\square being the connectivity of \square . To count it only once, we therefore subtract $(k_\square-1)$ times the entropic contribution ΔS_\square of each parity check \square . This leads to

$$s = \frac{1}{N} \left(\sum_{\circ} \Delta S_{\circ+\square \in \circ} - \sum_{\square} (k_\square - 1) \Delta S_\square \right) = \langle \Delta S_{\circ+\square \in \circ} \rangle - \frac{\langle \ell \rangle}{\langle k \rangle} \sum_k c_k (k-1) \langle \Delta S_\square^{(k)} \rangle, \quad (17)$$

where $\langle \Delta S_{\circ+\square \in \circ} \rangle$ represents the average of $\Delta S_{\circ+\square \in \circ}$ over the nodes \circ and $\langle \Delta S_\square^{(k)} \rangle$ the average of ΔS_\square over the parity checks \square with connectivity $k_\square=k$; the factor $\langle \ell \rangle / \langle k \rangle$ accounts for the ratio of the number M of parity checks over the number N of nodes.

To compute $\Delta S_{\circ+\square \in \circ}$, we need to know whether the bits of the nodes adjacent to \circ are fixed or not, in the absence of the “cavity node” \circ . As the cavity node is connected to its neighbors through parity checks [see Fig. 4(a)], we can decompose the computation in two steps. First, we observe that a given neighboring parity check constrains the value of the cavity node if and only if all the other nodes to which it is connected have themselves their bit fixed *in the absence of the cavity node*. Denoting by ζ the probability of this event and by η the probability for a node to be free *in the absence of one of its adjacent parity check*, we thus have

$$\zeta = \sum_k \frac{kc_k}{\langle k \rangle} [1 - (1 - \eta)^{k-1}] = 1 - \frac{c'(1 - \eta)}{\langle k \rangle}, \quad (18)$$

where $kc_k/\langle k \rangle$ is the probability for a parity check be connected to $k-1$ nodes *in addition to the cavity node* [see Fig. 4(a)] and $1 - (1 - \eta)^{k-1}$ is the probability that at least one of these $k-1$ nodes is free in the absence of the parity check. Next, we observe that the probability for the cavity node to

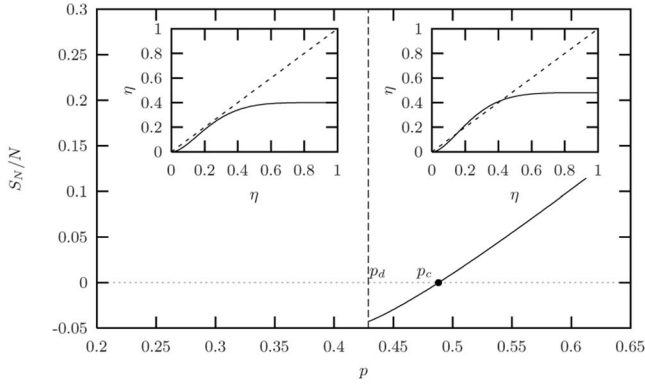


FIG. 5. Reduced entropy vs noise level p for an LDPC code with $k=6$ and $\ell=3$. When $p=0.4 < p_d$ (left inset), $\eta=0$ is the only solution to the cavity equation (24), yielding $s=0$. When $p=0.48 > p_d$ (right inset), two more solutions appear, one of which is stable. The entropy of this solution crosses zero at the critical noise p_c , above which the entropy become strictly positive, causing failure of decoding.

be free is the probability that none of its adjacent parity checks is constraining—that is,

$$\nu = p \sum_{\ell} v_{\ell} \zeta^{\ell} = p v(\zeta). \quad (19)$$

In order to close the equations, we also need the probability for the cavity node to be free in the absence of one of its connected parity check [see Fig. 4(c)], which is

$$\eta = p \sum_{\ell} \frac{\ell v_{\ell}}{\langle \ell \rangle} \zeta^{\ell-1} = p \frac{v'(\zeta)}{\langle \ell \rangle}, \quad (20)$$

where $\ell v_{\ell} / \langle \ell \rangle$ represents the probability for a node to be connected to $\ell-1$ parity checks in addition to the one ignored. The “cavity fields” η and ζ , determined by Eqs. (18) and (20), contain all the information needed to evaluate the entropy. Thus $\langle \Delta S_{\square} \rangle$ is given by

$$\langle \Delta S_{\square} \rangle = (\ln 2) [p v(\zeta) - \langle \ell \rangle \zeta]. \quad (21)$$

The first term $(\ln 2) p v(\zeta)$ corresponds to $(\ln 2) \nu$ see [Eq. (19)], the average entropic contribution of a node \circ , and the second term $-(\ln 2) \langle \ell \rangle \zeta$ subtracts the entropic reductions of its adjacent parity-check nodes; indeed, they are $\langle \ell \rangle$ on average and each is constraining the cavity node with probability ζ . Similarly, the average entropic reduction due to a parity check alone is

$$\langle \Delta S_{\square}^{(k)} \rangle = -(\ln 2) [1 - (1 - \eta)^k] \quad (22)$$

since $1 - (1 - \eta)^k$ is the probability that at least one of the k connected nodes is free in the absence of the parity check [see Fig. 4(b)]. To sum up, the entropy is determined by the formulas

$$s = (\ln 2) \left[p v \left(1 - \frac{c'(1-\eta)}{\langle k \rangle} \right) - \frac{\langle \ell \rangle}{\langle k \rangle} [1 - c(1-\eta) - \eta c'(1-\eta)] \right], \quad (23)$$

$$\eta = p \frac{v'(1 - c'(1-\eta)/\langle k \rangle)}{\langle \ell \rangle}. \quad (24)$$

Equation (24) can admit two kinds of solution (see Fig. 5). The first kind, referred to as *ferromagnetic*, describes the situation where decoding is possible, with only one codeword being solution of the decoding CSP: this solution has $\eta=0$ (all bits are fixed to σ^*) and $s=0$. The second kind, referred to as *paramagnetic* (but strictly speaking corresponding to a 1RSB glassy solution), describes the situation where decoding is impossible and has $\eta > 0$. It is found to exist only for p greater than the so-called *dynamical threshold*, denoted by p_d . It is, however, relevant only when associated with a positive entropy, $s > 0$, a condition which defines the *static threshold*, denoted by p_c and satisfying $p_c > p_d$. The static threshold corresponds to the threshold above which decoding is doomed to fail, as confirmed by rigorous studies.

3. Algorithmic interpretation

The cavity method is related to a particular decoding algorithm known as belief propagation (BP). Its principle is the following: starting from a configuration where only the noncorrupted bits are fixed to their values, one goes through each node of the factor graph, checks if its immediate neighboring environment constrains it to a unique value, fixes it to this value if it is the case, and iterates the whole procedure until convergence. At the end, some bits may still not be fixed, which certainly occurs if the decoding CSP has not a unique solution, but if all the bits end up fixed, one is ensured to have correctly decoded. Similar message-passing algorithms can be defined with different channels. They are responsible for the practical interest of LDPC codes as they provide algorithmically efficient decoding (yet suboptimal, as discussed below). With the BEC, these algorithms are particularly easy to analyze thanks to the fact that one can never be fooled by fixing bits to an incorrect value. To perform the analysis of the possible outcomes of the belief propagation algorithm, we can assume without loss of generality that the transmitted message is $(0, \dots, 0)$ (the \mathbb{Z}_2 symmetry implies that all codewords are equivalent). We thus start with $\sigma_i = *$ if $i \in \mathcal{E}$ and $\sigma_i = 0$ otherwise. Cavity fields are attributed to each oriented link of the factor graphs and are updated with the following rules, where t indexes iteration steps:

$$h_{i \rightarrow a}^{(t+1)} = \begin{cases} 0 & \text{if } \sigma_i = 0 \text{ or if } u_{b \rightarrow i}^{(t)} = 1 \text{ for some } b \in i - a, \\ * & \text{otherwise,} \end{cases}$$

$$u_{a \rightarrow i}^{(t+1)} = \begin{cases} 1 & \text{if } h_{j \rightarrow a}^{(t)} = 0 \text{ for all } j \in a - i, \\ * & \text{otherwise.} \end{cases} \quad (25)$$

Here, $u_{a \rightarrow i}^{(t)} = 1$ (*) means that the parity check a is constraining (is not constraining) i . $h_{i \rightarrow a}^{(t)} = 0$ (*) means that σ_i is fixed (not determined) to its correct value 0 without taking

into account the constraints due to a . The algorithm is analyzed statistically by introducing

$$\eta^{(t)} = \frac{1}{\langle \ell \rangle N} \sum_{(i,a)} \delta(h_{i \rightarrow a}^{(t)}, 0), \quad \zeta^{(t)} = \frac{1}{\langle k \rangle M} \sum_{(i,a)} \delta(u_{a \rightarrow i}^{(t)}, 1). \quad (26)$$

As suggested by our notations, the evolution for these quantities exactly mimics the derivation of the formulas for the cavity fields, yielding

$$\eta^{(t+1)} = p \frac{v'(\zeta^{(t)})}{\langle \ell \rangle}, \quad \zeta^{(t+1)} = 1 - \frac{c'(1 - \eta^{(t)})}{\langle k \rangle}. \quad (27)$$

The fixed point is given by Eq. (24). When $p < p_d$, the algorithm converges towards the unique, ferromagnetic, fixed point $\eta^{(\infty)} = \zeta^{(\infty)} = 0$ and decoding is successfully achieved. When $p_d < p < p_c$, a paramagnetic fixed point appears in addition to the ferromagnetic fixed point and the iteration leads to this second paramagnetic fixed point. The belief propagation algorithm thus fails to decode above the dynamical threshold p_d , before reaching the static threshold p_c below which no algorithm can possibly be successful (in this sense, BP is suboptimal).

B. Average error exponents

1. Entropic (IRSB) large deviations

The previous section recalled the properties of typical codes subject to typical noise. With finite codewords, $N < \infty$, failure to decode may also be due to atypical noise with unusually destructive effects. This is the purpose of our large deviation approach to investigate such events. We first focus on the simplest case: namely, the computation of the average error exponent where both the codes \mathcal{C} and the noise ξ are treated on the same footing (see Sec. II E). Our procedure to deal with the statistics over atypical factor graphs is an application of the cavity method for large deviations proposed

in [15]. For the sake of simplicity, we restrain ourselves here to regular codes, where nodes and check nodes have both fixed connectivity, ℓ and k , respectively, and defer the generalization to irregular codes to Appendix D.

As explained in Sec. II E, the thermodynamic formalism assigns a Boltzmann weight $e^{x S_N(\mathcal{C}, \xi)}$ to each ‘‘configuration’’ (\mathcal{C}, ξ) . The parameter x plays the role of an inverse temperature or, in other words, is a Lagrange multiplier enforcing the value of S_N . Taking the infinite-temperature limit $x=0$ (no constraint on the value of S_N) will thus lead us back to the typical case discussed above.

The cavity equations are as before derived by considering the effect of the addition of a node. As adding a new node, along with its adjacent parity checks, inevitably increases the degrees of the other nodes, strictly restraining to regular graphs is not possible and we must work in a larger framework. Accordingly, we consider ensembles where the degree of parity checks is fixed to k , but where the degree of nodes has a distribution $\{v_L\}$ (meaning that degree L has probability v_L , independently for each node). We will describe the regular ensemble by taking $v_L = \delta_{\ell, L}$ in the final formulas. Adding a new node with ℓ parity checks brings us from an ensemble characterized by v_L to an ensemble characterized by v'_L , with

$$v'_L = \left(1 - \frac{\ell(k-1)}{N}\right)v_L + \frac{\ell(k-1)}{N}v_{L-1} = v_L + \frac{\ell(k-1)}{N}\delta v_L, \quad (28)$$

where $\delta v_L = v_{L-1} - v_L$, since $\ell(k-1)$ nodes have their degree increased by 1. Let denote by $L(s, \{v_L\})$ the rate function for the probability to observe $S_N/N = s$ in an ensemble characterized by $\{v_L\}$ —that is,

$$\mathbb{P}_N[(\mathcal{C}, \xi): S_N(\mathcal{C}, \xi)/N = s | \{v_L\}] \asymp e^{-NL(s, \{v_L\})}. \quad (29)$$

We introduce $P_{\circ+\square\in\circ}^{(\ell)}(\Delta S)$, the probability distribution of the entropy contribution caused by the addition of the new nodes along with its ℓ adjacent parity checks. The passage from N nodes to $N+1$ nodes can then be described by

$$\begin{aligned} \mathbb{P}_{N+1}(s = S/(N+1) | \{v_L\}) &\asymp e^{-(N+1)L(S/(N+1), \{v_L\})} = \sum_{\ell} v_{\ell} \int d\Delta S P_{\circ+\square\in\circ}^{(\ell)}(\Delta S) \mathbb{P}_N[s = (S - \Delta S)/N | \{v_L - \ell(k-1)/N \delta v_L\}] \\ &\asymp \sum_{\ell} v_{\ell} \int d\Delta S P_{\circ+\square\in\circ}^{(\ell)}(\Delta S) e^{-NL[(S - \Delta S)/N, \{v_L - \ell(k-1)/N \delta v_L\}]}. \end{aligned} \quad (30)$$

Expanding for large N , one gets

$$\begin{aligned} \phi_s(x) &= xs - L(s, \{v_L\}) \\ &= \ln \sum_{\ell} v_{\ell} \int d\Delta S P_{\circ+\square\in\circ}^{(\ell)}(\Delta S) e^{x\Delta S + z\ell(k-1)}, \end{aligned} \quad (31)$$

with

$$z = \sum_L \delta v_L \frac{\partial L(s, \{v_L\})}{\partial v_L}. \quad (32)$$

The parameter z is determined by noting that the addition of a new parity check changes the node degree distribution in the same way as in Eq. (28), with $v'_L = v_L + (k/N)\delta v_L$, yielding

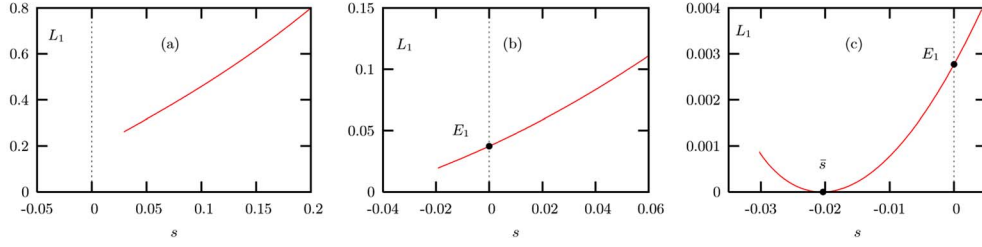


FIG. 6. (Color online) Rate function $L(s)$ as a function of the entropy s , here illustrated with a regular code with $k=6$ and $\ell=3$ (for the BEC channel). The three regimes are represented. (a) $p=0.2 < p_{1RSB}$: the spinodal of the paramagnetic solution is for $s_d > 0$. (b) $p=0.35 \in [p_{1RSB}, p_d]$: the spinodal is now for $s_d < 0$. (c) $p=0.45 \in [p_d, p_c]$: the spinodal is preceded by a minimum (the typical value), with $x_d = \partial_s L(s=s_d) < 0$. The typical dynamical and static transitions can be read on the $s=0$ axis: by definition of p_d and p_c , this equation has a solution \bar{s} for $p > p_d$ and this solution is positive, $\bar{s} > 0$, for $p > p_c$ (not represented here).

$$e^{-NL(S/N, \{v_L\})} \asymp \int d\Delta S P_{\square}(\Delta S) e^{-NL[(S-\Delta S)/N, \{v_L - (k/N)\delta v_L\}]}, \quad (33)$$

where $P_{\square}(\Delta S)$ is the probability of the entropy reduction caused by the addition of a new parity check. Expanding here also for large N leads to an equation for z ,

$$z = -\frac{1}{k} \ln \int d\Delta S P_{\square}(\Delta S) e^{x\Delta S}. \quad (34)$$

Following the same line of reasoning as in the typical case, the two distributions $P_{\square}^{(\ell)}$ and P_{\square} can be expressed by means of cavity fields η and ζ . First consider the addition of a node: If the bit of the new node is fixed, either because it was not erased or because one of its adjacent parity checks constrains it, there is an entropic reduction $-\ln 2$ per nonconstraining adjacent parity check and thus a weight 2^{-x} . Otherwise, if the new node is free, which occurs with probability $p\zeta^{\ell}$, the entropy shift is $(\ln 2)(1-\ell)$, giving a weight $2^{x(1-\ell)}$. Taking $v_L = \delta_{L,\ell}$, Eq. (31) therefore reads

$$\phi_s(x) = \ln[(\zeta 2^{-x} + 1 - \zeta)^{\ell} - p(\zeta 2^{-x})^{\ell} + p\zeta^{\ell} 2^{x(1-\ell)}] + \ell(k-1)z, \quad (35)$$

with

$$\zeta = 1 - (1 - \eta)^{k-1}. \quad (36)$$

Similarly, a new parity check removes a degree of freedom if and only if one of its adjacent node is free, which happens with probability $1 - (1 - \eta)^k$, yielding

$$z = -\frac{1}{k} \ln\{1 - [1 - (1 - \eta)^k] + [1 - (1 - \eta)^k]2^{-x}\}. \quad (37)$$

Finally, we obtain a self-consistent equation for η by considering the addition of a new (cavity) node in the absence of one of its adjacent parity checks:

$$\eta = \text{P(cavity node free)} \\ \propto \int d\Delta S P_{\circ \rightarrow \square}(\Delta S | \text{cavity node free}) e^{x\Delta S + z(\ell-1)(k-1)}, \quad (38)$$

$$1 - \eta = \text{P(cavity node fixed)}$$

$$\propto \int d\Delta S P_{\circ \rightarrow \square}(\Delta S | \text{cavity node fixed}) e^{x\Delta S + z(\ell-1)(k-1)}, \quad (39)$$

where $P_{\circ \rightarrow \square}$ corresponds to $P_{\circ \rightarrow \square}^{(\ell-1)}$, taken either under the condition that the cavity node be free or that be fixed. We obtain

$$\eta = \frac{p2^x(\zeta 2^{-x})^{\ell-1}}{(\zeta 2^{-x} + 1 - \zeta)^{\ell-1} + p(2^x - 1)(\zeta 2^{-x})^{\ell-1}}. \quad (40)$$

Alternatively, these equations can be obtained by differentiation of Eq. (35), which is variational with respect to the cavity η . The large deviation cavity equations (36) and (40) allow us to compute the generating function $\phi_s(x)$ using Eqs. (35) and (37), from which the rate function $L(s | \{v_l = \delta_{l,\ell}\})$ is deduced by Legendre transformation as discussed in Sec. II E.

Again, two kinds of solutions, paramagnetic or ferromagnetic, can be present. For a given value of p , we find that a nontrivial, paramagnetic solution to Eq. (40) exists only for $x \geq x_d(p)$. In agreement with the observation reported in the previous section that the paramagnetic solution typically exists only when $p < p_d$, we have $x_d(p) < 0$ for $p > p_d$ and $x_d(p) > 0$ for $p < p_d$ (the typical case is indeed associated with $x=0$). We obtain the average error exponent by selecting the value of $L(s)$ where $s=0$: our results are illustrated in Fig. 6. By extension of the concept of dynamical threshold p_d , one could define a ‘‘dynamical’’ error exponent as $E_d(p) = L(x_d(p)) = x_d(p)s(x_d(p)) - \phi(x_d(p))$ with $x_d(p)$ corresponding to the temperature of the spinodal for the paramagnetic solution. The relevance of this concept is, however, limited by the fact that the algorithmic interpretation presented in Sec. III A 3 does not extend to large deviations (see also Sec. III C 3).

More interestingly, we find an additional threshold (see Table III), denoted p_{1RSB} , below which the equation $s(x)=0$ has no longer a solution (see Fig. 6). This inconsistency of the 1RSB solution is indicative of the presence of a phase transition occurring at some $p_e > p_{1RSB}$. The following section is devoted to computing p_e and describing the nature of the new phase present for $p < p_e$.

TABLE III. Values of some thresholds p_{1RSB} , p_{RS} , p_e , p_d , and p_c for different regular ensembles of LDPC codes on the BEC.

(k, ℓ)	p_{1RSB}	p_{RS}	p_e	p_d	p_c
(4,3)	0.325 262 970 9	0.546 574 881 1	0.606 872 016 6	0.647 425 6494	0.746 009 7025
(6,3)	0.266 856 875 4	0.337 837 464 1	0.349 188 490 2	0.429 439 8144	0.488 150 8842
(6,5)	0.013 008 205 24	0.427 701 036 8	0.714 365 751 3	0.551 003 5344	0.833 315 3204
(10,5)	0.044 128 845 46	0.243 565 689 4	0.334 772 117 6	0.341 550 0230	0.499 490 7179

2. Energetic (RS) large deviations

The previous ‘‘entropic (1RSB) approach’’ attributed errors to the presence of an exponential number of solutions in the decoding CSP. The same assumption was underlying the analysis of the typical case, in Sec. III A 2, where rigorous studies support the conclusions drawn from this hypothesis. This view is also consistent with the phase diagram of XOR-SAT problems to which the encoding CSP belongs. The structure of the well-separated codewords corresponds in this context to a ‘‘frozen 1RSB glassy’’ phase. As p departs from the value $p=1$, however, the decoding CSP deviates increasingly in nature from the initial encoding CSP. As the number of constraints increases (as p decreases), the presence of an exponential number of solutions (glassy phase) in addition to the isolated correct codeword becomes less and less probable. An alternative rare event possibly dominating the probability of error at low p is the presence of a second isolated (ferromagnetic) codeword close to the correct one. This can lead to a new phase transition that has no counterpart in the typical phase diagram, reflected by a nonanalyticity of the error exponent.

In our framework, investigating an alternative source of error requires considering for S_N another quantity than the entropy of the number of solutions. A possible choice, associated with a replica symmetric (RS) ansatz, is the energy E_N of the ground state of the decoding CSP, giving the minimal number of violated parity checks. Ignoring the correct codeword, a second isolated codeword is present if and only if $E_N=0$ (otherwise $E_N>0$). Large deviations of this energy are described by the rate function $L_1(e)$ defined as

$$P[\xi, \mathcal{C}: E_N(\xi, \mathcal{C})/N = e] \asymp e^{-NL_1(e)}. \quad (41)$$

The generating function for the rate function $L_1(e)$, defined by

$$e^{N\phi_e(x)} = \mathbb{E}_{\xi, \mathcal{C}}[e^{xE_N(\xi, \mathcal{C})}] = \int de e^{N(xe - L_1(e))}. \quad (42)$$

is given by (see [24] for a similar calculation)

$$\begin{aligned} \phi_1(x) = \ln & \left\{ p \int \prod_{a=1}^{\ell} du_a Q(u_a) \exp \left[-x \left(\sum_{a=1}^{\ell} |u_a| - \left| \sum_{a=1}^{\ell} u_a \right| \right) \right] \right. \\ & \left. + (1-p) \int \prod_{a=1}^{\ell} du_a Q(u_a) \exp \left(-2x \sum_{a=1}^{\ell} \delta_{u_a, -1} \right) \right\} \\ & - \frac{\ell(k-1)}{k} \ln \int \prod_{i=1}^k dh_i P(h_i) \exp \left[-x \delta \left(\prod_{i=1}^k h_i - 1 \right) \right], \end{aligned} \quad (43)$$

with

$$\begin{aligned} P(h \neq +\infty) & \propto p \int \prod_{a=1}^{\ell-1} du_a Q(u_a) \\ & \times \exp \left[-\frac{x}{2} \left(\sum_{a=1}^{\ell-1} |u_a| - \left| \sum_{a=1}^{\ell-1} u_a \right| \right) \right] \\ & \times \delta \left(h - \sum_{a=1}^{\ell-1} u_a \right), \end{aligned} \quad (44)$$

$$P(h = +\infty) \propto (1-p) \int \prod_{a=1}^{\ell-1} du_a Q(u_a) \exp \left(-x \sum_{a=1}^{\ell-1} \delta_{u_a, -1} \right), \quad (45)$$

$$Q(u) = \int \prod_{i=1}^{k-1} dh_i P(h_i) \delta \left[u - S \left(\prod_{i=1}^{k-1} h_i \right) \right], \quad (46)$$

where $S(x)=1$ if $x>0$, -1 if $x<0$, and 0 if $x=0$. Since u only takes values in $\{-1, 0, +1\}$ and h is restrained to integer values, we can introduce

$$Q(u) = q_+ \delta(u-1) + q_- \delta(u+1) + q_0 \delta(u) \quad (47)$$

and

$$p_+ = \int_{h>0} dh P(h), \quad p_- = \int_{h<0} dh P(h), \quad p_0 = 1 - p_+ - p_-. \quad (48)$$

Our interest is here in zero-energy ground states, described by the limit $x \rightarrow \infty$, where the equations simplify to

$$\begin{aligned} \phi_e(x = +\infty) = -L(e = 0) = & \ln[(1 - q_-)^\ell + p(1 - q_+)^\ell - pq_0^\ell] \\ & - \frac{\ell(k-1)}{k} \ln \left[1 - \frac{1}{2} [(p_+ + p_-)^k - (p_+ - p_-)^k] \right], \end{aligned} \quad (49)$$

with

$$p_+ \propto (1 - q_-)^{\ell-1} - pq_0^{\ell-1}, \quad (50)$$

$$p_- \propto p(1 - q_+)^{\ell-1} - pq_0^{\ell-1}, \quad (51)$$

$$p_0 \propto pq_0^{\ell-1}, \quad (52)$$

$$q_+ = \frac{1}{2} [(p_+ + p_-)^{k-1} + (p_+ - p_-)^{k-1}], \quad (53)$$

$$q_- = \frac{1}{2} [(p_+ + p_-)^{k-1} - (p_+ - p_-)^{k-1}], \quad (54)$$

$$q_0 = 1 - (p_+ + p_-)^{k-1}. \quad (55)$$

We find that the only stable solution to these cavity equations satisfies $q_0 = p_0 = 0$, which allows us to further simplify the formulas

$$\begin{aligned} \phi_e(+\infty) = & \ln[q_+^\ell + p(1 - q_+)^\ell] \\ & - \frac{\ell(k-1)}{k} \ln \left[\frac{1}{2} [1 + (2p_+ - 1)^k] \right], \end{aligned} \quad (56)$$

with

$$p_+ = \frac{q_+^{\ell-1}}{q_+^{\ell-1} + p(1 - q_+)^{\ell-1}}, \quad (57)$$

$$q_+ = \frac{1}{2} [1 + (2p_+ - 1)^{k-1}]. \quad (58)$$

The resulting RS average error exponent, given by $E_e(p) = -\phi(+\infty)$, is represented in Fig. 7.

We identify the transition p_e as the point where the 1RSB and RS error exponents coincide, which satisfies $p_e > p_{1RSB}$. We find that the RS solution is limited by a spinodal point and is only defined for $p \geq p_{RS}$. While we conjecture that the 1RSB estimate is exact for $p > p_e$, the existence of p_{RS} suggests that either an additional phase transition occurs at some $p'_e > p_{RS}$ or, more radically, that our description of the phase $p < p_e$ is incorrect. The limit case of random codes, however, indicates that the energetic method is valid in the limit $k, \ell \rightarrow \infty$.

3. Limit of random codes

The only limiting case where the average error exponent has been obtained integrally so far is the fully connected limit where $k, \ell \rightarrow \infty$ with $\ell/k = \alpha = 1 - R$ fixed. This limit corresponds to the *random linear model*, where each parity check is connected to each node with probability $1/2$. In this limit, the entropic 1RSB approach gives

$$E_s(k, \ell \rightarrow \infty) = L(s = 0) = D(1 - R \| p), \quad (59)$$

where $D(q \| p) = q \ln(q/p) + (1 - q) \ln[(1 - q)/(1 - p)]$ is known as the Kullback-Leibler divergence, while the energetic RS approach gives

$$E_e(k, \ell \rightarrow \infty) = -\phi_e(+\infty) = -(R - 1) \ln 2 - \ln(1 + p) \quad (60)$$

(with $p_+ = 1/1 + p$ and $q_+ = 1/2$). The two expressions coincide at the critical noise p_e , with

$$p_e = (1 - R)/(1 + R). \quad (61)$$

We thus predict the average error exponent of the RLM to be

$$E_1(\text{RLM}) = \begin{cases} (1 - R) \ln 2 - \ln(1 + p) & \text{if } p < \frac{1 - R}{1 + R}, \\ D(1 - R \| p) & \text{if } \frac{1 - R}{1 + R} < p < 1 - R. \end{cases} \quad (62)$$

This result coincides with the exact expression (see Appendix B for a direct combinatorial derivation), thus validating our approach in this particular case.

As explained above, we are not able to fully account for the small noise regime as soon as k and ℓ are finite, even though the solutions are found to be stable with respect to further replica symmetry breakings in the space of codewords [30]. This does not exclude that a similar replica symmetry breaking occurs in the space of codes. Remarkably, previous attempts reported in the literature have also failed to obtain error exponents in the low p regime.

C. Typical error exponents

1. Cavity equations

The typical error exponent is encoded into a potential $\psi(x, y)$, as defined in Eq. (13). The equations for $\psi(x, y)$ are obtained from the cavity method for large deviations by following very closely the path leading to $\phi(x)$ [31]. As noticed in Sec. II, the formalism with finite y provides a generalization of the average case which is recovered by taking $y = 1$, with $\psi(x, y = 1) = \phi(x)$. We will therefore quote our results. In the entropic (1RSB) case, we find

$$\psi_s(x,y) = \ln[(\zeta 2^{-xy} + 1 - \zeta)^\ell - (\zeta 2^{-xy})^\ell + \zeta^\ell (p 2^x + 1 - p)^y 2^{-\ell xy}] - \frac{\ell(k-1)}{k} \ln\{(1-\eta)^k + [1 - (1-\eta)^k] 2^{-xy}\}, \quad (63)$$

with

$$\eta = \frac{\zeta^{\ell-1} (p 2^x)^y 2^{-(\ell-1)xy}}{(\zeta 2^{-xy} + 1 - \zeta)^{\ell-1} - (\zeta 2^{-xy})^{\ell-1} + \zeta^{\ell-1} (p 2^x + 1 - p)^y 2^{-(\ell-1)xy}},$$

$$\zeta = 1 - (1 - \eta)^{k-1}. \quad (64)$$

In the energetic (RS) case with $x = +\infty$, we find

$$\psi_e(x = +\infty, y) = \ln[q_+^\ell + p^y (1 - q_+)^{\ell}] - \frac{\ell(k-1)}{k} \ln\left[\frac{1}{2}[1 + (2p_+ - 1)^k]\right], \quad (65)$$

with

$$p_+ = \frac{q_+^{\ell-1}}{q_+^{\ell-1} + p^y (1 - q_+)^{\ell-1}}, \quad (66)$$

$$q_+ = \frac{1}{2}[1 + (2p_+ - 1)^{k-1}]. \quad (67)$$

In each case, from the potential $\psi(x,y)$, the rate function is obtained as $\mathcal{L}(\phi, x) = y\phi - \psi(x,y)$, with $\phi(x) = \partial_y \psi(x,y)$. By definition, a typical code corresponds to a minimum of \mathcal{L} , with $\mathcal{L} = 0$, which, when \mathcal{L} is analytical at this minimum, is associated with $y = \partial_\phi \mathcal{L} = 0$.

As a generic feature, we find that $\mathcal{L}(y,x)$ is an increasing function of y for fixed x , going from negative values for $y < y_c(x)$ to positive ones for $y > y_c(x)$. Negative rate functions, as thus obtained, are certainly unphysical. As negative entropies in the usual cavity-replica method, we attribute them to analytical continuations of physical solutions. The simplest way to circumvent them is, as with the frozen 1RSB ansatz in the replica method, to select $y_c(x)$ with $\mathcal{L}(y_c(x), x) = 0$. When $y_c(x) < 1$, meaning that $\mathcal{L}(y=1, x) > 0$, we consider that the average exponent is associated with atypical codes and therefore differs from the typical exponent, described by $\mathcal{L}(y_c(x), x) = 0$. Using this criterion, we find that the two exponents indeed differ for the lowest values of p , when $p < p_y$, where $p_y < p_e$ (see Fig. 8 for an illustration). In general the situation is complicated by the fact that the cavity equations may fail to provide solutions in this regime, as already seen in the average case when $p < p_{RS}$ (corresponding here to $y=1$); the random code limit, where this complication is absent, is thus the most instructive.

2. Limit of random codes

In the limit $k, \ell \rightarrow \infty$, we obtain the following results. In the entropic regime, $p > p_e$, the average and typical exponents are found to coincide. This conclusion extends in the energetic regime only for a restricted interval $[p_y, p_e]$. When $p < p_y$, we have $y_c(x) < 1$ and average and typical error ex-

ponents differ. The formula we obtain for the typical error exponent reads

$$E_{\text{typ}}(\text{RLM}) = \begin{cases} -\delta_{GV}(R) \ln p & \text{if } p < p_y, \\ E_{\text{av}}(\text{RLM}) & \text{if } p_y < p < p_c, \end{cases} \quad (68)$$

with

$$p_y = \frac{\delta_{GV}(R)}{1 - \delta_{GV}(R)}. \quad (69)$$

$\delta_{GV}(R)$ denotes the smallest solution to $(R-1) \ln 2 + H(\delta) = 0$, whose interpretation is discussed in Appendix B. This result, which does not seem to have been reported previously in the literature, coincides with the union bound presented in Appendix C, which strongly suggests that it is indeed exact.

For LDPC with finite connectivity, a similar phase diagram is expected. In the entropic regime, we find indeed that average and typical exponents are identical. In the energetic regime, we face the problem that the cavity equations have no solution below some value of p , which precludes us from estimating p_y .

3. Algorithmic implications

The cavity formalism has the attractive property of corresponding formally to message passing algorithms. Based on this analogy, new algorithmic procedures have been systematically proposed to analyze single finite graphs; each time the cavity approach was found to operate at the ensemble level. With a phase transition occurring at the ensemble level, we have, however, here a system where such a correspondence is no longer meaningful. Following the usual procedure, it is indeed straightforward to implement the cavity approach for average error exponent on a single graph, but in the regime $p < p_y$, this algorithm is doomed to fail: for any typical graph, in the limit of large size, the message passing algorithm will yield the average error exponent, which, as we have seen, is distinct for the correct, typical, error exponent.

IV. LDPC CODES OVER THE BSC

A. Definition

We now turn to error exponents for LDPC codes on the binary symmetric channels. One motivation for repeating the analysis with this channel is that it is representative of a broader class of channels, where bits are not simply erased as

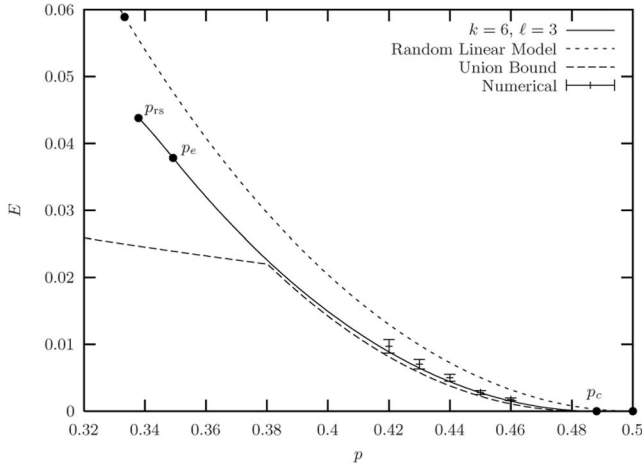


FIG. 7. Average error exponent as a function of the noise level p for the regular code ensemble with $k=6$ and $\ell=3$, on the BEC. Numerical estimates of the error probability, based on 10^6 runs of exact maximum-likelihood decoding (using Gauss elimination) on samples of sizes ranging from $N=500$ to $N=1500$, yield reasonably good estimates of the error exponent using an exponential fit. These numerical results agree well with our theoretical prediction. The union bound (C11) and the random linear limit (62) are also represented for comparison.

with the BEC, but can be *corrupted*, in the sense that their content 0 or 1 is changed to other admissible values. This clearly complicates the decoding as corrupted bits cannot be straightforwardly identified; in fact, with the BSC, no scheme can guarantee to identify the corrupted bits and the receiver is never certain that his decoding is correct. We will, however, see that the overall phase diagram is very similar to that obtained with the BEC.

By definition, maximum-likelihood decoding consists in inferring the most probable realization of the noise *a posteriori*. The *a posteriori* probability can be expressed from the *a priori* probability thanks to Bayes' theorem. If \mathbf{x} denotes the transmitted message and \mathbf{y} the received message, the *a priori* probability to receive \mathbf{y} given \mathbf{x} is

$$Q(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^N (1-p)^{\delta_{x_i,y_i}} p^{1-\delta_{x_i,y_i}}. \quad (70)$$

To make contact with physical models of disordered systems [12], it is convenient to adopt a spin convention $\sigma_i = (-1)^{x_i}$, $\tau_i = (-1)^{y_i}$, and to rewrite the previous relation as

$$Q(\boldsymbol{\sigma}|\boldsymbol{\tau}) \propto \exp\left(\sum_{i=1}^N h_i \tau_i\right), \quad h_i \equiv h_0 \sigma_i, \quad h_0 \equiv \frac{1}{2} \ln\left(\frac{1-p}{p}\right). \quad (71)$$

This formulation emphasizes the analogy with the random field Ising model [32], a prototypical disordered system. Using the group symmetry of the set of codewords, we can assume, without loss of generality, that the sent codeword is $\boldsymbol{\sigma} = (+1, \dots, +1)$. With this simplification, the random field takes value $h_i = h_0$ with probability $1-p$ and $-h_0$ with prob-

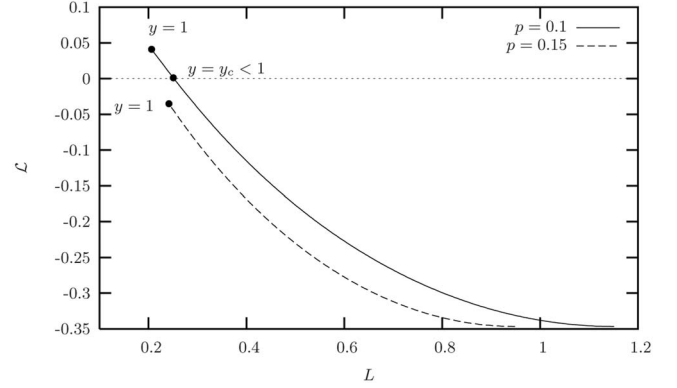


FIG. 8. Rate function $\mathcal{L}(L_e) = \mathcal{L}[-\phi_e(+\infty)]$ of the energetic error exponent for an LDPC code with $k=24$, $\ell=12$ on the BEC. When $p > p_y$ (solid curve), the rate function is negative (and therefore unphysical) for all $0 < y < 1$, entailing that the typical and average error exponents should coincide. When $p < p_y$ (dashed curve), we postulate that the typical error exponent is given by the inverse “freezing temperature” y_c at which the rate function cancels.

ability p . Bayes' formula for the *a posteriori* probability that the message $\boldsymbol{\tau}$ was sent reads

$$P(\boldsymbol{\tau}|\boldsymbol{\sigma}) = \frac{P(\boldsymbol{\sigma}|\boldsymbol{\tau})P(\boldsymbol{\tau})}{\sum_{\boldsymbol{\tau}'} P(\boldsymbol{\sigma}|\boldsymbol{\tau}')P(\boldsymbol{\tau}')} \\ = \frac{1}{Z(\beta)} \exp\left(\beta \sum_{i=1}^N h_i \tau_i\right) \prod_{a=1}^M \delta(\tau_a = 1), \quad (72)$$

where τ_a is a shorthand for $\prod_{i \in a} \tau_i$: in the present spin convention, the constraint induced by the parity check a indeed reads $\tau_a = 1$. To continue the analogy with statistical mechanics, we have also introduced a temperature β , called the decoding temperature, whose value is here fixed to $\beta=1$ (Nishimori temperature—see [11]). Given the *a posteriori* probability, the selection of the most probable codeword $\mathbf{d}(\boldsymbol{\sigma})$ can still be done according to different criteria, among which are the following.

(i) Word maximum *a posteriori* (word MAP), where one maximizes the posterior probability in block by taking $\mathbf{d}_{\text{block}}(\boldsymbol{\sigma}) = \arg\max_{\boldsymbol{\tau}} P(\boldsymbol{\tau}|\boldsymbol{\sigma})$. This scheme minimizes the block-error probability $P_{\text{block}} = (1/M) \sum_{\boldsymbol{\tau}} P[\mathbf{d}(\boldsymbol{\sigma}) \neq \boldsymbol{\tau}]$.

(ii) Symbol maximum *a posteriori* (symbol MAP), where one maximizes the posterior probability bit per bit by taking $\mathbf{d}_{\text{bit}}(\boldsymbol{\sigma}) = \arg\max_{\tau_i} \sum_{\boldsymbol{\tau}_{j \neq i}} P(\boldsymbol{\tau}|\boldsymbol{\sigma})$. This scheme minimizes the bit-error probability $P_{\text{bit}} = (1/M) \sum_{\boldsymbol{\tau}} (1/N) \sum_i P[\mathbf{d}(\boldsymbol{\sigma})_i \neq \tau_i]$.

In physical terms, the word-MAP procedure consists in finding the ground state of the system with partition function $Z(\beta)$ given by the normalization in Eq. (72); this amounts to studying the zero-temperature limit $\beta \rightarrow \infty$. Conversely, symbol MAP is equivalent to taking the sign of the local magnetizations at temperature $\beta=1$,

$$\tau_i^{\text{bit}} = \text{sgn}(\langle \tau_i \rangle) = \text{sgn}\left[\sum_{\boldsymbol{\tau}} \tau_i P(\boldsymbol{\tau}|\boldsymbol{\sigma})\right]. \quad (73)$$

We will treat the two cases in a common framework by considering an arbitrary temperature $\beta \geq 1$.

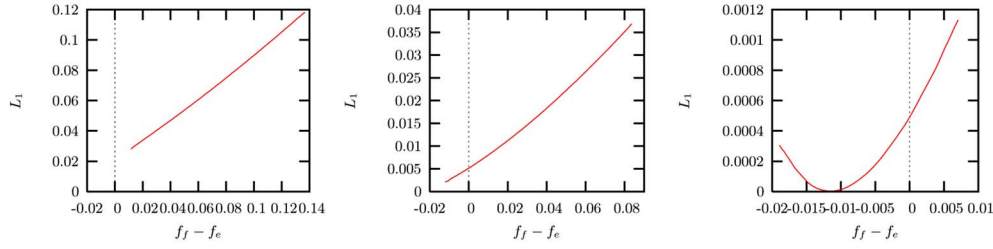


FIG. 9. (Color online) Large deviation rate $L_1(f_f - f_e, s_e = 0)$ as a function of the difference between the ferromagnetic and the nonferromagnetic free energies, here for regular codes with $k=6$ and $\ell=3$ on the BSC. The thresholds are $p_{1RSB} \approx 0.058$ and $p_c \approx 0.100$. The three regimes are represented. From left to right, $p=0.045$, $p=0.07$, and $p=0.09$.

From the physical perspective, the original codeword is recovered if it dominates the Gibbs measure defined in Eq. (72). This can be expressed by decomposing the partition function $Z(\beta)$ as

$$Z(\beta) = Z_{\text{corr}}(\beta) + Z_{\text{err}}(\beta), \quad Z_{\text{corr}}(\beta) = e^{\beta \sum_i h_i},$$

$$Z_{\text{err}}(\beta) = \sum_{\tau \neq 1} e^{\beta \sum_i h_i \tau_i} \prod_a \delta(\tau_a - 1). \quad (74)$$

We define the corresponding free energies $F_{\text{corr}}(\beta) = -(1/\beta) \ln Z_{\text{corr}}(\beta)$ and $F_{\text{err}}(\beta) = -(1/\beta) \ln Z_{\text{err}}(\beta)$. The first one corresponds physically to a ferromagnetic phase (as with the BEC), while the second will be shown to correspond to either a paramagnetic or a glassy phase, depending on the values of β and p . Decoding is successful if, and only if, the ferromagnetic phase has lower free energy, $F_{\text{corr}} < F_{\text{err}}$. The quantity $S_N(\xi, \mathcal{C})$ introduced in Sec. II E can therefore be defined here as

$$S_N = F_{\text{corr}}(\beta) - F_{\text{err}}(\beta), \quad (75)$$

where the dependence in the noise ξ and the code \mathcal{C} is implicitly understood.

B. Cavity analysis and the 1RSB frozen ansatz

As with the BEC, explicit calculations can be performed by means of the replica or cavity methods. Details can be found in Appendix E, and we only discuss here the points where differences with the BEC arise. For any fixed p , a replica-symmetric calculation, whose derivation follows the derivation of the paramagnetic solution with the BEC, is found to undergo an entropy crisis—i.e., $s_{\text{RS}}(\beta) = \beta^2 \partial_{\beta} f_{\text{RS}}(\beta) < 0$ for $\beta > \beta_g$. This feature is indicative of the presence of a glassy phase and points to the need to break the replica symmetry. The glassy phase of LDPC codes is, however, of the “frozen 1RSB” type, which implies that the glassy free energy f_{err} can be completely inferred from the replica-symmetric solution f_{RS} . This simplicity stems from the “hard” nature of the constraints: changing a bit automatically violates all its surrounding checks, forcing the rearrangement of many variables [33,34]. When the degree of all nodes is $\ell_i \geq 2$, one can indeed show [24] that changing one bit while keeping all checks satisfied requires the rearrangement of an extensive ($\propto N$) number of variables (in the language of [24], factor graphs of LDPC codes have no leaves).

The consequence, expressed in the replica language, is that the 1RSB “states” are reduced to single configurations and thus have zero internal entropy. The 1RSB potential $\phi(\beta, m)$ whose optimization over $m \in [0, 1]$ is predicted to yield f_{err} [20] thus simplifies to $\phi(\beta, m) = f_{\text{RS}}(\beta m)$ [35], since

$$e^{-N\beta m \phi(\beta, m)} \equiv \sum_{\text{states } \alpha} e^{-N\beta m f_{\alpha}(\beta)} = \sum_{\alpha} e^{-N\beta m e_{\alpha}} = e^{-\beta m f_{\text{RS}}(\beta m)}. \quad (76)$$

According to whether one is above or below the freezing temperature β_g^{-1} , defined by

$$s_{\text{RS}}(\beta_g) = \beta_g^2 \partial_{\beta} f_{\text{RS}}(\beta_g) = 0, \quad (77)$$

the free energy $f_{\text{err}}(\beta)$ is given either by $f_{\text{RS}}(\beta)$ (paramagnetic phase) or by $f_{\text{RS}}(\beta_g)$ (glassy phase). This is summarized as follows:

$$f_{\text{err}}(\beta) = \max_{\beta' < \beta} f_{\text{RS}}(\beta') = \begin{cases} f_{\text{RS}}(\beta) & \text{if } \beta < \beta_g, \\ f_{\text{RS}}(\beta_g) & \text{if } \beta > \beta_g. \end{cases} \quad (78)$$

Finally, we note that as in the BEC case, a nonferromagnetic solution $f_{\text{RS}}(\beta)$ exists only for large enough p . The threshold $p_d(\beta)$ giving the smallest noise level at which a nonferromagnetic solution exists is again called the dynamical threshold and can be shown here also to coincide with the dynamical arrest of BP [28].

C. Average error exponent: LDPC codes

In the region relevant for error exponents, where $p < p_c$ and $\beta \geq 1$, the ferromagnetic solution is typically dominant (this is the definition of $p < p_c$) and metastable phases described by f_{err} are typically glassy, since $\beta_g < 1$. Therefore, to compute error exponents, we have to consider $f_{\text{err}}(\beta) = f_{\text{RS}}(\beta_g)$ and not $f_{\text{err}}(\beta) = f_{\text{RS}}(\beta)$. This leads us to introduce an extra temperature β_e distinct from the decoding temperature β , which is to be set to β_g by requiring that the entropy s_{RS} be zero. Similarly, we introduce a ferromagnetic tem-

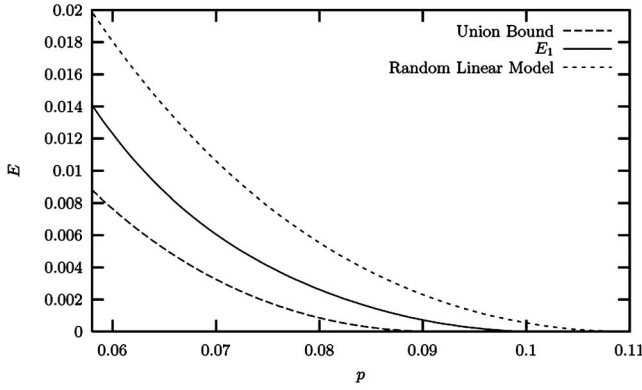


FIG. 10. Average error exponent as a function of the noise level p for the regular code ensemble with $k=6$ and $\ell=3$ through the BSC. Here $p_{1\text{rsb}} \approx 0.058$. The union bound (C17) and the random linear model ($k, \ell \rightarrow \infty$) limit (B14) are also represented for comparison.

perature β_f , set to $\beta_f = \beta$, and define the rate function $L_1(f_e, f_f)$ and its Legendre transform as

$$\begin{aligned} \mathbb{P}[\xi, \mathcal{C}: F_{\text{RS}}(\beta_e)/N = f_e, F_{\text{corr}}(\beta_f)/N = f_f] &\asymp e^{-NL_1(f_e, f_f)}, \\ e^{N\phi_1(\beta_e, \beta_f, x_e, x_f)} &= \mathbb{E}_{\xi, \mathcal{C}}[e^{-x_e \beta_e F_{\text{RS}}(\beta_e) - x_f \beta_f F_{\text{corr}}(\beta_f)}] \\ &= \int df_e df_f e^{N[-x_e \beta_e f_e - x_f \beta_f f_f - L_1(f_e, f_f)]}. \end{aligned} \quad (79)$$

The potential ϕ_1 contains all the necessary information about both solutions:

$$-\beta_a \phi_a = \partial_{x_a} \phi_1, \quad s_a = \partial_{x_a} \phi_1 - \frac{\beta_a}{x_a} \partial_{\beta_a} \phi_1, \quad (80)$$

where the index $a=e, f$ corresponds to the two possible phases. For the purpose of computing error exponents, we need only to control $f_e - f_f$ and s_e for all temperatures $\beta_e < \beta$. Note that the ferromagnetic solution f_f has no entropy, $s_f=0$, which is here reflected by the fact that the potential ϕ_1 depends upon β_f and x_f only through $m_f \equiv \beta_f x_f$. These observations allow us to focus on a simplified potential

$$\hat{\phi}(\beta_e, m) = \phi_1\left(\beta_e, x_e = \frac{m}{\beta_e}, m_f = -m\right), \quad (81)$$

which satisfies

$$\partial_m \hat{\phi} = f_f - f_e, \quad \partial_{\beta_e} \hat{\phi} = -ms_e. \quad (82)$$

As with the BEC, the average error exponent is identified with the smallest value of L_1 such that $s_e \geq 0$ and $f_f - f_e \geq 0$. The present formulation is in fact equivalent to the presentation based on the replica method given in [10]. A remarkable consequence of the analysis is that the average error exponent is predicted to be the same for any $\beta \geq 1$. Indeed, both the glassy and the ferromagnetic free energies are temperature independent for $\beta \geq \beta_g$. In particular, symbol and word MAP are predicted to have same error exponents.

Based on the cavity equations given in Appendix E, the potential $\hat{\phi}$ can be computed numerically by population dy-

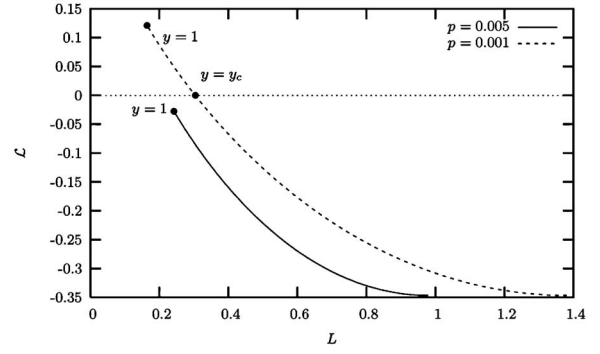


FIG. 11. Rate function $\mathcal{L}(L)$ for the RLM on the BSC with $R=1/2$ and $p=0.005 > p_y$ (solid curve) and $p=0.001 < p_y$ (dashed curve).

namics. As an illustration, we plot in Fig. 9 the rate function $L_1(f_f - f_e, s_e=0)$ for a regular code with $k=6$, $\ell=3$. As in the case of BEC, three regimes can be distinguished, according to the value of p .

- (i) $p < p_{1\text{rsb}}$: no zero-entropy RS solution typically exists and $f_e < f_f$ for the metastable solutions.
- (ii) $p_{1\text{rsb}} < p < p'_d$: no zero-entropy RS solution typically exists but the dominant metastable solutions have $f_e > f_f$.
- (iii) $p'_d < p < p_c$: a zero-entropy RS solution is typically present.

The major difference with the BEC is that the threshold p'_d , defined by $p'_d = p_d(\beta_g(p'_d))$ does not coincide with the dynamical threshold $p_d(\beta)$: indeed here p'_d is defined in relation to the existence of a solution with positive entropy, while, in the framework of BP, the dynamical arrest p_d is related to the existence of a paramagnetic solution at decoding temperature β^{-1} [28]. In Fig. 10, we plot the average error exponent for regular codes with $k=6$, $\ell=3$.

D. Random code limit

1. Average error exponent

As with the BEC, the $k, \ell \rightarrow \infty$ limit can be computed exactly, yielding

$$E_1^{(1)} = L_1(f_f = f_e, s_e = 0) = D(\delta_{GV}(R) \| p), \quad (83)$$

where $\delta_{GV}(R)$ denotes the smallest solution to $R - 1 + H(\delta) = 0$. In this regime, errors are most likely to be caused by large noises driving the received message beyond the typical nearest-codeword distance.

As pointed out in [10], a second ferromagnetic solution is present in this limit (see Appendix E for details), yielding the error exponent

$$E_1^{(2)} = -\ln \frac{1}{2} [1 + 2\sqrt{p(1-p)}] - R \ln 2. \quad (84)$$

Such a solution also exists for finite k, ℓ , but is clearly unphysical (it predicts negative exponents for $k=6, \ell=3$). Yet it correctly describes the low p phase (B14) in the $k, \ell \rightarrow \infty$ limit, where failure is caused by the existence of one (or a

few) unusually close codewords. In that sense it plays the same role as the energetic solution in the BEC analysis, with the difference that it is not extensible to any case with finite connectivities. The critical noise p_e below which such a scenario occurs is given by

$$\frac{\sqrt{p_e}}{\sqrt{p_e + \sqrt{1 - p_e}}} = \delta_{GV}(R). \quad (85)$$

We thus predict the average error exponent to be

$$E_1(\text{RLM}) = \begin{cases} D(\delta_{GV}(R) \parallel p) & \text{if } p < p_e < p_c, \\ -\ln \frac{1}{2}[1 + 2\sqrt{p(1-p)}] - R \ln 2 & \text{if } p < p_e. \end{cases} \quad (86)$$

This expression coincides with the exact result (B14) of the RLM.

2. Typical error exponent

The typical exponent of the RLM can be evaluated using the two-step potential:

$$e^{N\psi(\beta_e, m, y)} = \mathbb{E}_c[e^{N y \hat{\phi}(\beta_e, m)}] = \int d\hat{\phi} e^{N[y\hat{\phi} - \mathcal{L}(\hat{\phi}, \beta_e, m)]}. \quad (87)$$

The details of the calculations by the cavity method are given in Appendix E. As in the average case, two distinct solutions appear. The first one is the counterpart of the solution discussed in Sec. IV C. It yields, in the random linear limit,

$$\psi(\beta_e, m, y) = y \hat{\phi}(\beta_e, m). \quad (88)$$

A consequence of the linear dependence on y is that $\hat{\phi}$ always takes the value obtained from the average calculation,

irrespectively of y . Therefore, the average and typical error exponents coincide in this regime and are given by Eq. (83).

This solution is, however, only valid in the high-noise regime ($p > p_e$). As in the average case, for low p , the errors in decoding are dominated by the presence of a subexponential (zero entropy) number of close codewords. The associated solution has for potential

$$\psi(y) = -yL - \mathcal{L} = (R-1)\ln 2 + \ln\{1 + [2\sqrt{p(1-p)}]^y\}. \quad (89)$$

We observe two types of behavior according to the value of p : for $p_y < p < p_e$, $\mathcal{L}(y)$ is negative for $0 \leq y \leq 1$, whereas for $p < p_y$, it crosses 0 at $y_c < 1$ (see Fig. 11). Interpreting, as in the BEC analysis (see Sec. III C 1), negative values of \mathcal{L} as evidence of a glassy transition in the space of codes, we deduce that the typical error exponent is given by $L(y_c)$ when $y_c < 1$, in which case it differs from the average error exponent. To sum up,

$$E_0(\text{RLM}) = \begin{cases} L(y_c) = -\delta_{GV}(R)\ln[2\sqrt{p(1-p)}] & \text{if } p < p_y, \\ L(y=1) = E_1(\text{RLM}) & \text{if } p_y < p < p_c, \end{cases} \quad (90)$$

where the critical noise $p_y(R)$ is a solution of

$$\frac{2\sqrt{p_y(1-p_y)}}{1 + 2\sqrt{p_y(1-p_y)}} = \delta_{GV}(R). \quad (91)$$

This exponent coincides with the RLM limit of the union bound (C18) and is rigorously established [7] to be the correct typical error exponent on the BSC.

V. CONCLUSION

Since Shannon laid the basis for information theory, the analysis of error-correcting codes has been a major subject of study in this field of science [4]. Error-correcting codes aim

at reconstructing signals altered by noise. Their performance is measured by their error probability—i.e., the probability that they fail in accomplishing this task. For block codes, where the messages are taken from a set of 2^M codewords of length N , it is known that when the rate $R = M/N$ is below the channel capacity R_c , the probability of error behaves, in the limit of large N , at best, as $P_e \sim \exp[-NE(R)]$ [4]. This error exponent $E(R)$, also called reliability function, provides a particularly concise characterization of performance.

For a given code ensemble, two classes of error exponents can generally be distinguished, due to the presence of two levels of “disorder,” one associated with the choice of the code itself and a second associated with the realization of the noise. *Average error exponents* correspond to take the error

probability P_e with respect to these two levels simultaneously, while *typical error exponents* refer to fixed, typical, codes.

In the present paper, we tackled the computation of these two error exponents for a particular class of block codes, the low-density parity-check codes, with two particular channels, the binary erasure channel and the binary symmetric channel. We considered decoding under maximum-likelihood decoding, the best conceivable decoding procedure. We framed the problem in terms of large deviations and applied a recently proposed extension of the cavity method designed to probe atypical events in systems defined on random graphs [15]. This method provides an alternative to the replica method used in [10] to address similar problems, with the advantage of being based on explicitly formulated probabilistic assumptions. With respect to this earlier contribution, our work offers several clarifications, notably on the nature of the different phases, and various extensions, notably to the BEC channel. With this particular channel, our results are analytical, and in the high-noise regime, we conjecture them to be exact. Recent mathematical results on the typical phase diagram [36] foster hope for a confirmation of our results in that context.

From a statistical physics perspective, error exponents are interesting for the richness of their phase diagram, which comprises two phase transitions of different natures. These transitions are observed when the level of noise p is varied at fixed rate R (or, equivalently in the special case of random codes, when the rate R is varied at fixed p). Close to the static threshold, for $p_e < p < p_c$, errors are mostly due to the proliferation of many incorrect codewords in the vicinity of the received message. We interpreted this feature in terms of the presence of a glassy phase, and accordingly, we were able to describe this regime by considering a one-step replica symmetry breaking approach. Below p_e , errors become dominated by the effect of single isolated codewords, which we attributed to a transition towards a ferromagnetic state or 1RSB to RS transition. The noise p_e has its counterpart in the “critical rate” R_e of information theory [4], which marks the point below which only bounds on the reliability function are known. The replica-symmetric approach we employed to investigate the regime $p < p_e$ also turns out to be only approximate, except in the limit of infinite connectivity, where we recovered the error exponents of random linear codes [7]. We also described a second transition occurring at $p_y < p_e$, below which atypical codes come to dominate the average exponent, causing it to differ from the typical error exponent. As it takes place in the space of graphs, this is an example of a critical phenomenon whose description is not accessible to the standard cavity method [14], but only to its extension to large deviations [15] (see also [37] for an other example). However, this second transition should be taken with utmost care, as it relies on an approximate ansatz.

The numerous efforts made in the information theory community to account for the low rate regime $R < R_e$ have so far resulted only in upper and lower bounds for the reliability function [6]. Maybe not too surprisingly, this is also the region of the phase diagram where our methods encounter difficulties. Several examples are, however, now available which demonstrate that statistical physics methods can pro-

vide exact solutions to notoriously difficult mathematical problems. The solutions thus obtained generally sharpen our comprehension both of the system at hand and of the techniques themselves, besides often paving the way for rigorous derivations. In the light of some recent such achievements, extending the present statistical physics approach to reach a thorough understanding of error exponents seems to us a valuable challenge.

ACKNOWLEDGMENTS

The work of T.M. was supported in part by the EC through the network MTR 2002-00319 “STIPCO” and the FP6 IST consortium “EVERGROW.” O.R. thanks the Human Frontier Science Program for support.

APPENDIX A: A NOTE ON THE EXPONENTIAL SCALING

The thermodynamic approach is based on the assumption that the leading contribution to the probability of error decays exponentially with N . However, as initially shown by Gallager, for ensembles of LDPC codes, the probability of error decays only polynomially in N to the leading order. In physical terms, this is due to a few codes (whose number is a polynomial in N) which display a second, metastable, ferromagnetic state at a smaller distance from the ground state (corresponding to the correct codeword) than the numerous configurations forming the paramagnetic state.

To overpass this spurious effect in the simplest, yet purely theoretical way, Gallager focused on the so-called “expurgated ensemble” where the half of the codes with smallest minimum distance is disregarded. On this restricted ensemble which excludes the codes with multiple ferromagnetic states, the error probability decays now exponentially in N at the leading order and can be characterized with an average error exponent. Needless to say, this construction only makes sense as a convenient theoretical way to access good codes.

As the large deviation method automatically overlooks any polynomial contribution, its results actually apply to the “expurgated ensemble.” This is, however, only true to the extent that the expurgation does not affect the distribution of graphs in the ensemble (i.e., does not change the distribution of degrees, of loops, etc.). This is presumably the case, as supported by the construction presented in [38], where an expurgated ensemble much tighter than Gallager’s one is defined by explicitly associating to any random code an expurgated code obtained by modifying only a number $O(1)$ of small loops.

APPENDIX B: RANDOM LINEAR MODEL

Definition

A parity-check code is defined by a $M \times N$ matrix A over \mathbb{Z}_2 and its codewords are the vectors $\mathbf{x}=(x_1, \dots, x_N)$ satisfying $A\mathbf{x}=0$. Code ensembles are therefore subsets of the set of

all 2^{MN} possible matrices. Taking this complete set (with all possible matrices having same probability) defines the so-called *random linear model*. In contrast with LDPC codes, since a typical matrix from the RLM is not sparse, the belief propagation algorithm cannot be used to decode. While of little practical interest due to this absence of efficient decoding algorithm, the RLM has, however, two major theoretical advantages, both originating from its “maximally random” nature: typical codes from the RLM saturate the Shannon bounds, and error exponents can be derived rigorously. We review here some of the established results, which we used in the main text as a reference point to compare our nonrigorous results. Error exponents for the RLM are indeed expected to provide upper bounds for error exponents of LDPC ensemble, which are reached only in the limit of infinite connectivity $k, l \rightarrow \infty$ (this limit is similar to that in which p -spin models approach the random energy model when $p \rightarrow \infty$ [27]).

Weight enumerator function

We first characterize the geometry of the space of codewords by means of the so-called *weight enumerator function*. Given a code C with matrix A , this function gives the number $\mathcal{N}_C(d)$ of codewords \mathbf{x} at (Hamming) distance $d = |\mathbf{x}| \equiv \sum_{i=1}^N x_i$ from the origin:

$$\mathcal{N}_C(d) = \sum_{\mathbf{x}} \delta\left(d, \sum_{i=1}^N x_i\right) \delta(A\mathbf{x}, \mathbf{0}), \quad (\text{B1})$$

where the sum is over all codewords and $\delta(x, y)$ enforces the constraint $x = y$. The *average* weight enumerator function is obtained by averaging over the code ensemble and satisfies

$$\bar{\mathcal{N}}(d) \equiv \mathbb{E}_C[\mathcal{N}_C(d)] = \binom{N}{d} 2^{-M} \asymp e^{N\Sigma(R, \delta=d/N)},$$

$$\Sigma(R, \delta) = (R-1)\ln 2 + H(\delta), \quad (\text{B2})$$

where the limit of infinite block length, $N \rightarrow \infty$, is taken with $M = N(1-R)$ and $d = Nx$. The exponent $\Sigma(R, x)$ defines the so-called *average weight enumerator exponent*. A critical distance is the distance $\delta_{GV}(R)$ defined as the smallest $\delta > 0$ such that $\Sigma(R, \delta) = 0$. Codewords at distance $d = N\delta$ with $\delta > \delta_{GV}(R)$ proliferate exponentially. On the other hand, the probability of existence of a codeword at distance $d = N\delta$ with $\delta < \delta_{GV}(R)$ is upper-bounded by $\bar{\mathcal{N}}(d)$ and thus decays exponentially with N . Consequently, for any $\epsilon(N)$ such that $\epsilon(N) \rightarrow \infty$ [e.g., $\epsilon(N) = \sqrt{N}$], only an exponentially small fraction of the codes in the ensemble have a minimal nonzero distance $d = N\delta$ smaller than $N\delta_{GV}(R) - \epsilon(N)$. Excluding these “worst” codes from the RLM defines the *expurgated RLM ensemble*.

Average error exponent over the BEC

Due to the group symmetry of the set of codewords, we can assume without loss of generality that the transmitted codeword is $(0, \dots, 0)$. For a given realization of the disorder

due to a BEC, we denote by $E \subset \{1, \dots, N\}$ the subset of erased bits in the received string and d the number of elements in E . If A is the $M \times N$ matrix representing the code, the submatrix \tilde{A}^E induced by A on E defines the decoding CSP problem: decoding is impossible if and only if the kernel of \tilde{A}^E is nonzero. When all matrices A are sampled with uniform probabilities as in the RLM, the submatrices \tilde{A}^E are also represented with uniform probability. Given a noise realization E of magnitude d , the error probability is the probability that a random $M \times d$ matrix \tilde{A}^E is noninjective,

$$\mathbb{E}_C[\mathbb{P}_N^{(B)}(\mathbf{0})] = \sum_{d=0}^N \binom{N}{d} p^d (1-p)^{N-d} \times \mathbb{P}(\exists \mathbf{x} \neq \mathbf{0} \text{ such that } \tilde{A}^E \mathbf{x} = \mathbf{0}). \quad (\text{B3})$$

When $d > M$, \tilde{A}^E is necessarily noninjective. When $d \leq M$, on the other hand, a straightforward inductive argument [8] gives

$$\mathbb{P}(\exists \mathbf{x} \neq \mathbf{0} \text{ such that } \tilde{A}^E \mathbf{x} = \mathbf{0}) = 1 - \prod_{i=0}^{d-1} (1 - 2^{i-M}). \quad (\text{B4})$$

Consequently, the *exact* expression for the average error probability of the RLM reads

$$\mathbb{E}_C[\mathbb{P}_N^{(B)}(\mathbf{0})] = \sum_{d=0}^M \binom{N}{d} p^d (1-p)^{N-d} \left(1 - \prod_{i=0}^{d-1} (1 - 2^{i-M}) \right) + \sum_{d=M+1}^N \binom{N}{d} p^d (1-p)^{N-d}. \quad (\text{B5})$$

In the $N \rightarrow \infty$, this expression can be evaluated by the saddle-point method. When $p < (1-R)/(1+R)$, the dominant contribution comes from the first sum, with

$$\sum_{d=0}^M \binom{N}{d} p^d (1-p)^{N-d} \left(1 - \prod_{i=0}^{d-1} (1 - 2^{i-M}) \right) \asymp e^{-N[(1-R)\ln 2 - \ln(1+p)]} \quad (\text{B6})$$

and typical number of errors $d = N2p/(1+p)$. When $p > (1-R)/(1+R)$ (and $p < 1-R$ to stay below the capacity), the dominant contribution comes from the second sum, with

$$\sum_{d=M+1}^N \binom{N}{d} p^d (1-p)^{N-d} \asymp e^{-ND(1-R)p} \quad (\text{B7})$$

and the typical number of errors $d = N(1-R)$. We thus obtain for the average error exponent of the RLM the expression given in Eq. (62),

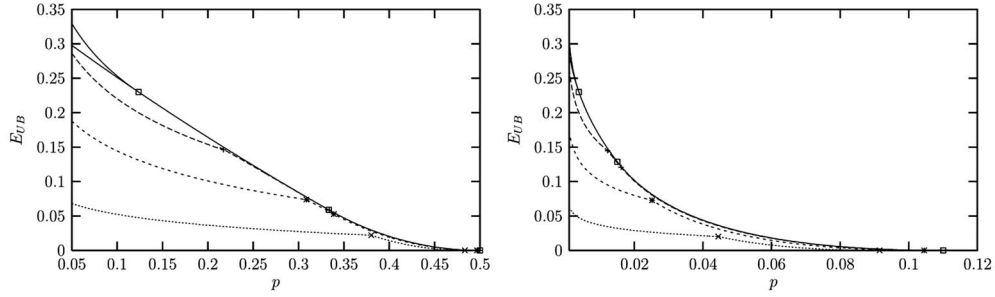


FIG. 12. Expurgated union bounds for the BEC (left) and BSC (right). From bottom to top, $(k, \ell) = (6, 3), (8, 4), (12, 6)$ and the RLM limit, expurgated (top solid curve) and not expurgated (bottom solid curve) with $R = 1/2$. The points indicate the transition between the three regimes, as well as e_{UB} .

$$E_1(\text{RLM}) = \begin{cases} (1-R)\ln 2 - \ln(1+p) & \text{if } p < \frac{1-R}{1+R}, \\ D(1-R \parallel p) & \text{if } \frac{1-R}{1+R} < p < 1-R. \end{cases} \quad (\text{B8})$$

In physical terms, the transition between the two regimes can be interpreted as a transition between a ferromagnetic (RS) phase and a glassy (1RSB) phase. In the high-noise regime $p > (1-R)/(1+R)$, the error is indeed most probably due to the noise driving the received string into a “glassy phase” of exponentially numerous incorrect codewords, as reflected by the fact that then $\text{P}(\exists \mathbf{x} \neq \mathbf{0} \text{ such that } \tilde{A}^E \mathbf{x} = \mathbf{0}) = 1$. In contrast, in the low-noise regime, $p < (1-R)/(1+R)$, the error is most probably due to the noise driving the received string into a “ferromagnetic phase” where an isolated incorrect codeword happens to be closer than the correct codeword; this is reflected by the fact that $\text{P}(\exists \mathbf{x} \neq \mathbf{0} \text{ such that } \tilde{A}^E \mathbf{x} = \mathbf{0})$ differs from 1 only by an exponentially small term in N , as seen from Eq. (B4).

Average error exponent over the BSC

With the binary symmetric channel, starting again from the transmitted codeword is $(0, \dots, 0)$, the received string \mathbf{y} cannot be decoded if there exists $\mathbf{x} \neq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{0}$ and

$|\mathbf{x} - \mathbf{y}| < |\mathbf{y}|$. Denoting $P_e(\mathbf{y})$ the probability of this event, the probability of error is

$$\mathbb{E}_C[\text{P}_N^{(B)}(\mathbf{0})] = \sum_{d=0}^N \binom{N}{d} p^d (1-p)^{N-d} P_e(\mathbf{y}^{(d)}), \quad (\text{B9})$$

where $\mathbf{y}^{(d)}$ is a generic string of weight d —e.g., $y_i = 1$ if $i \leq d$, $y_i = 0$ if $i > d$. If $d/N > \delta_{GV}(R)$, $P_e(\mathbf{y}^{(d)})$ goes to 1 in the infinite block-length limit. Although no published proof is available in the literature, it is reported as proved [7] that, when $d/N < \delta_{GV}(R)$, $P_e(\mathbf{y}^{(d)})$ is asymptotically equivalent to its union bound approximation (see the following appendix)—i.e.,

$$P_e(\mathbf{y}^{(d)}) \sim \mathbb{E}_C \left[\sum_{\mathbf{x} \neq \mathbf{0}} \theta(d - |\mathbf{x} - \mathbf{y}^{(d)}|) \delta(A\mathbf{x}, \mathbf{0}) \right] \quad (\text{B10})$$

$$\sim \sum_{i=0}^d \mathbb{E}_C[\mathcal{N}_C(i, \mathbf{y}^{(d)})] \quad (\text{B11})$$

$$\sim \mathbb{E}_C[\mathcal{N}_C(d, \mathbf{y}^{(d)})], \quad (\text{B12})$$

where $\mathcal{N}_C(i, \mathbf{y}^{(d)})$ is the number of codewords at distance i from $\mathbf{y}^{(d)}$ and $\theta(x) = 1$ if $x > 0$ and 0 otherwise. Straightforward combinatorics shows that the asymptotic behavior of $\mathbb{E}_C \mathcal{N}_C(i, \mathbf{y}^{(d)})$ is given by the standard weight enumerator exponent $\Sigma(R, i/N)$. In the limit $N \rightarrow \infty$ where $\delta = d/N$ is kept fixed, a saddle-point evaluation leads to the following expression of the average error exponent:

$$E_1(\text{RLM}) = - \max_{\delta < \delta_{GV}} [\Sigma(R, \delta) - D(\delta \parallel p)] \quad (\text{B13})$$

$$= \begin{cases} (1-R)\ln 2 - \ln[1 + 2\sqrt{p(1-p)}] & \text{if } \frac{\sqrt{p}}{\sqrt{p} + \sqrt{1-p}} < \delta_{GV}(R), \\ D(\delta_{GV}(R) \parallel p) & \text{otherwise.} \end{cases} \quad (\text{B14})$$

This result with two distinct regime is very similar to that obtained previously for the BEC.

APPENDIX C: UNION BOUNDS

The so-called *union bound exponent* is a rigorous lower bound of the average error exponent in the expurgated ensemble. We show in this appendix how the average weight enumerator exponent of (regular) LDPC codes can be used to derive this union bound exponent, for both the BEC and BSC. We will thus recover results first established by Gallager in [4,39]. In a nutshell, the idea of the union bound is to upper-bound the probability that at least one (bad) codeword causes an error by the sum of the probabilities that each does. Remarkably, this union bound turns out to be tight for the RLM ensemble.

Weight enumerator function

The weight enumerator function [see Eq. (B1) for the definition] of regular LDPC codes with $k=6$ and $\ell=3$ was computed in [4] and reads

$$\begin{aligned} \mathbb{E}_C[\mathcal{N}_C(d)] &= \sum_{\mathbf{x}} \delta(|\mathbf{x}|, d) \mathbb{E}_C[\delta(A\mathbf{x} = \mathbf{0})] \\ &= \binom{N}{d} \mathbb{E}_C[\delta(A\mathbf{x}^{(d)} = \mathbf{0})] \end{aligned} \quad (C1)$$

$$\mathbb{E}_C[\mathcal{N}_C(d = \delta N)] \asymp e^{N\Sigma(k, \ell, \delta)}, \quad (C2)$$

with

$$\Sigma(k, \ell, \delta) = \min_{\mu} \left(2\mu \ell \delta + (1 - \ell)H(\delta) + \frac{\ell}{k} \ln C(\mu) \right), \quad (C3)$$

and

$$C(\mu) = \frac{1}{2} [(1 + e^{-2\mu})^k + (1 - e^{-2\mu})^k]. \quad (C4)$$

We introduce δ_m , the smallest δ such that $\Sigma(k, \ell, \delta) \geq 0$. By construction, the average enumerator exponent in the expurgated ensemble is

$$\Sigma_{\text{exp}}(k, \ell, \delta) = \begin{cases} \Sigma(k, \ell, \delta) & \text{if } \Sigma(k, \ell, \delta) > 0 \text{ (i.e., if } \delta > \delta_m), \\ -\infty & \text{otherwise.} \end{cases} \quad (C5)$$

This expurgated *average* enumerator exponent $\Sigma_{\text{exp}}(k, \ell, \delta)$ is believed to coincide with the *typical* enumerator exponent [40,41].

Union bound for the BEC

Given the set E of erased bits, we want to estimate the probability $P_e(d)$ that the CSP-decoding problem has at least two solutions, when a code C is drawn at random from its ensemble. We call A the matrix characterizing C , \tilde{A}^E the

submatrix induced by A on E , and d the number of erased bits. The union bound consists in the following inequality:

$$P_e(d) = \mathbb{P}(\exists \tilde{\mathbf{x}} \in \{0, 1\}^d \neq \mathbf{0} \text{ such that } \tilde{A}^E \tilde{\mathbf{x}} = \mathbf{0}) \quad (C6)$$

$$\leq \min \left[\sum_{\tilde{\mathbf{x}} \neq \mathbf{0}} \mathbb{P}(\tilde{A}^E \tilde{\mathbf{x}} = \mathbf{0}), 1 \right]. \quad (C7)$$

Let $w = |\tilde{\mathbf{x}}|$ and \mathbf{x} be constructed from $\tilde{\mathbf{x}}$ by setting $x_i = \tilde{x}_i$ for $i \in E$, $x_i = 0$ otherwise: $\tilde{\mathbf{x}}$ belongs to the kernel of \tilde{A} if and only if \mathbf{x} belongs to the kernel of A . The probability of the latter event reads

$$\mathbb{E}_C[\mathcal{N}_C(w)] \binom{N}{w}^{-1}. \quad (C8)$$

The error probability is consequently bounded by

$$\mathbb{E}_C[\mathbb{P}_N^{(B)}] = \sum_{d=0}^N \binom{N}{d} p^d (1-p)^{N-d} P_e(d) \quad (C9)$$

$$\begin{aligned} &\leq \sum_{d=0}^N \binom{N}{d} p^d (1-p)^{N-d} \\ &\quad \times \min \left[\sum_{w=0}^d \binom{d}{w} \mathbb{E}_C[\mathcal{N}_C(w)] \binom{N}{w}^{-1}, 1 \right]. \end{aligned} \quad (C10)$$

In the infinite block-length limit, a saddle-point estimate yields, as upper bound for the expurgated average error exponent, the exponent

$$\begin{aligned} E_{\text{exp}}(k, \ell) &\geq E_{UB} \\ &= -\max_{\delta} \left\{ -D(\delta \| p) \right. \\ &\quad \left. + \min \left[\max_{\omega} \left(\Sigma(\omega) + \delta H\left(\frac{\omega}{\delta}\right) - H(\omega) \right), 0 \right] \right\} \\ &= -\max_{\delta < \delta_{UB}} \left\{ -D(\delta \| p) + \max_{\omega > \delta_m} \min_{\mu} \left[\delta H\left(\frac{\omega}{\delta}\right) \right. \right. \\ &\quad \left. \left. + 2\mu \ell \omega - \ell H(\omega) + \frac{\ell}{k} \ln C(\mu) \right] \right\}, \end{aligned} \quad (C11)$$

where $\delta = d/N$, $\omega = w/N$, and δ_{UB} is the largest δ such that $\max_{\omega} (\Sigma(\omega) + \delta H(\frac{\omega}{\delta}) - H(\omega))$ is nonpositive.

As p is varied, three regimes can be distinguished. For small p , the maximum over ω is reached on the boundary δ_m , meaning that errors are dominated by the nearest codewords. For large p instead, the maximum over δ is reached at δ_{UB} , in which case the union bound is simply replaced by 1, physically corresponding to a large number of bad codewords arising from the large amplitude of the noise. Finally, in the intermediate region of p , the extremum is reached in the interior of the (δ, ω) domain. Note that this last regime is not always present when k and ℓ are too small (for $k=6$ and $\ell=3$ in particular). These three regimes are given in the limit $k, \ell \rightarrow \infty$ by

$$E_0(\text{RLM}) = \begin{cases} -\delta_{GV}(R)\ln p & \text{if } p < p_y, \\ (1-R)\ln 2 - \ln(1+p) & \text{if } p_y < p < \frac{1-R}{1+R}, \\ D(1-R\|p) & \text{if } \frac{1-R}{1+R} < p < 1-R, \end{cases} \quad (\text{C12})$$

with p_y defined as in Eq. (69). Union bounds for the BEC are plotted in Fig. 12 for several regular ensembles.

Union bound for the BSC

The union bound for the BSC is derived following the same steps than for the BEC. The counterpart of Eq. (C6) reads

$$P_e(d) = \mathbb{P}(\exists \mathbf{x} \neq 0 \text{ such that } |\mathbf{x} - \mathbf{y}^{(d)}| < d \text{ and } A\mathbf{x} = 0), \quad (\text{C13})$$

where $\mathbf{y}^{(d)}$ is a generic string of weight d . Let \mathbf{x} be a string a weight w and $Q(w, d, g)$ be the probability for $\mathbf{y}^{(d)}$ to be at distance g from \mathbf{x} , conditioned on $|\mathbf{y}^{(d)}| = d$:

$$Q(w, d, g) = \binom{w}{(d-g+w)/2} \binom{N-w}{(d+g-w)/2} \binom{N}{d}^{-1}. \quad (\text{C14})$$

The probability for $\mathbf{y}^{(d)}$ to be at distance g from any codeword \mathbf{x} is upper-bounded by

$$\sum_w \mathbb{E}_C[\mathcal{N}_C(w)] Q(w, d, g), \quad (\text{C15})$$

and we can write

$$P_e(d) \leq \min \left[\sum_{w,g} \mathbb{E}_C[\mathcal{N}_C(w)] Q_C(w, d, g), 1 \right] \\ \asymp \min \left[\sum_w \mathbb{E}_C[\mathcal{N}_C(w)] Q_C(w, d, d), 1 \right]. \quad (\text{C16})$$

From this inequality and Eq. (C9), we obtain the union bound for the error exponent via the saddle-point method:

$$E_{\text{exp}}(k, l) \geq E_{UB} = -\max_{\delta} \left\{ -D(\delta\|p) + \min_{\omega} [\max(\Sigma(\omega) + L(\omega, \delta, \delta), 0)] \right\} \\ = -\max_{\delta < \delta_{UB}} \left\{ -D(\delta\|p) + \max_{\omega > \delta_m} \min_{\mu} \left[2\mu \ell \omega + (1 - \ell)H(\omega) + \frac{\ell}{k} \ln C(\mu) + L(\omega, \delta, \delta) \right] \right\},$$

$$L(\omega, \delta, \gamma) = \omega H\left(\frac{\delta - \gamma + \omega}{2\omega}\right) + (1 - \omega) H\left(\frac{\delta + \gamma - \omega}{2(1 - \omega)}\right) - H(\delta). \quad (\text{C17})$$

As for the BEC, three regimes can be distinguished, according to the value of p . In the limit $k, \ell \rightarrow \infty$, these three regimes are

$$E_0(\text{RLM}) = \begin{cases} -\delta_{GV}(R)\ln[2\sqrt{p(1-p)}] & \text{if } p < p_y, \\ (1-R)\ln 2 - \ln[1 + 2\sqrt{p(1-p)}] & \text{if } p_y < p < p_e, \\ D(\delta_{GV}(R)\|p) & \text{if } p_e < p < \delta_{GV}(R), \end{cases} \quad (\text{C18})$$

where p_y and p_e are given by Eq. (91) and (85)

Union bounds for the BSC are plotted in Fig. 12.

APPENDIX D: IRREGULAR CODES

Definition of the ensemble

In this appendix we discuss the generalization to irregular graphs. We shall only treat the entropic large deviations with the BEC, but our arguments can easily be generalized to the other cases. With irregular codes, it is necessary to specify more precisely the definition of the ensemble. The usual definition is via the degree distributions v_ℓ and c_k . It is, however, possible to define different ensembles having same distribution and sharing the same typical properties, but differing at the level of atypical properties, including error exponents (see also [15] for similar nonequivalences in an other context).

The simplest construction takes all factor graphs with exactly $v_\ell N$ checks of degree ℓ , $c_k M$ variables of degree k , and pick them with uniform probability. Such ensembles are used to build actual codes, and we shall therefore analyze them with some details.

Average error exponent

We revisit the arguments of Sec. III B and emphasize the differences with the regular case.

A crucial modification is the introduction of Lagrange multipliers enforcing the number of nodes of each degree. Call N_ℓ the number of variables of degree ℓ and M_k the number of checks of degree k . Denote $n_\ell = N_\ell/N$ and $m_k = M_k/N$. The rate L_1 is now a function of the n_ℓ and m_k . Its multiple Legendre transform is defined as

$$\phi(x, \{\lambda_\ell\}, \{\nu_k\}) \doteq xs + \sum_{\ell} \lambda_\ell n_\ell + \sum_k \nu_k m_k - L_1, \quad (\text{D1})$$

with

$$x = \partial_s L_1, \quad \lambda_\ell = \partial_{n_\ell} L_1, \quad \nu_k = \partial_{m_k} L_1.$$

Let us consider the addition of a new bit. ℓ checks are added along with it, where ℓ is drawn with probability v_ℓ . Each of these checks, in turn, is connected to $k_a - 1$ old bits ($a=1, \dots, \ell$), where k_a is drawn with probability $k_a c_{k_a} / \langle k \rangle$. Equation (31) is modified in the following way:

$$\begin{aligned} \phi(x, \{\lambda_\ell\}, \{v_k\}) &= \ln \sum_\ell v_\ell \sum_{\{k_1, \dots, k_\ell\}} \prod_{a=1}^{\ell} \frac{k_a c_{k_a}}{\langle k \rangle} \\ &\times \int d\Delta S P_{\square}^{(\ell, k_1, \dots, k_\ell)}(\Delta S) \exp \left[x\Delta S + \sum_{a=1}^{\ell} [(k_a \right. \\ &\left. - 1)z_{k_a} + v_{k_a}] + \lambda_\ell \right]. \end{aligned} \quad (\text{D2})$$

The addition of a variable of degree ℓ is reflected by a factor e^{λ_ℓ} and the addition of a check of degree k by a factor e^{μ_k} . Call the k degree the degree of a variable with respect to checks of degree k . Here z_k is related to the increase of k degrees in the ensemble. Let us consider for a moment a more general setting, where the ensemble is determined by the k -degree distributions, denoted by $v_\ell^{(k)}$ [42]. Then z_k is defined by

$$z_k = \sum_\ell \delta v_\ell^{(k)} \frac{\partial L_1(s, \{v_\ell^{(k)}\})}{\partial v_\ell^{(k)}}, \quad (\text{D3})$$

where $\delta v_\ell^{(k)} = v_{\ell-1}^{(k)} - v_\ell^{(k)}$. z_k is obtained in a very similar way as z in Eq. (37):

$$z_k = -\frac{1}{k} \ln \int d\Delta S P_{\square}^{(k)}(\Delta S) e^{x\Delta S + v_k}, \quad (\text{D4})$$

where $P_{\square}^{(k)}(\Delta S)$ now depends on the degree k .

The cavity equation (24) is modified in a very similar way as the expression of ϕ_1 in Eq. (D2). The inversion of the Legendre transformation allows one to recover the relevant quantities:

$$s = \partial_x \phi, \quad n_\ell = \partial_{\lambda_\ell} \phi, \quad m_k = \partial_{v_k} \phi. \quad (\text{D5})$$

Replacing $P_{\square}^{(\ell, k_1, \dots, k_\ell)}(\Delta S)$ and $P_{\square}^{(k)}(\Delta S)$ by their values, we obtain

$$\phi_1 = xs - L_1 = \ln[v(A) + p(2^x - 1)v(B)], \quad (\text{D6})$$

with

$$A = e^{\lambda_\ell} \sum_k \frac{kc_k}{k} e^{(k-1)z_k + v_k} [2^{-x} + (1 - 2^{-x})(1 - v)^k],$$

$$B = 2^{-x} e^{\lambda_\ell} \sum_k \frac{kc_k}{k} e^{(k-1)z_k + v_k} [1 - (1 - v)^{k-1}],$$

$$z_k = -\frac{1}{k} \ln[2^{-x} + (1 - 2^{-x})(1 - v)^k] - \frac{v_k}{k},$$

$$v = \frac{p2^x v'(B)}{v'(A) + p(2^x - 1)v'(B)}.$$

To evaluate L_1 as a function of s , we simply need to tune the parameters λ_ℓ and m_k such that the conditions $n_\ell = v_\ell$ and $m_k = \alpha c_k$ are satisfied.

In Fig. 13, we represent the error exponent for the irregular ensemble with $v(x) = (1/2)x^3 + (1/2)x^4$ and $c(x) = (1/2)x^6 + (1/2)x^8$.

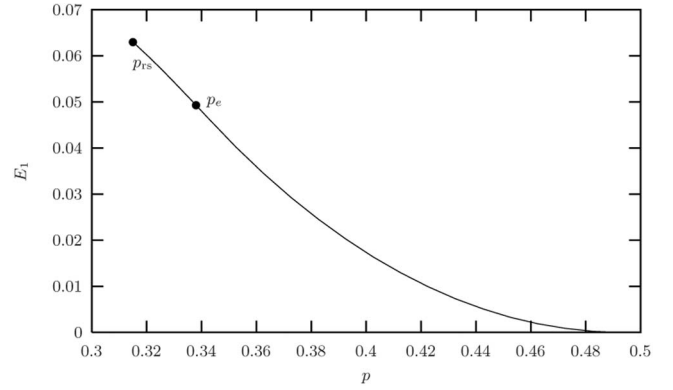


FIG. 13. Average error exponent of a given code as a function of the noise level p for irregular codes with $c_k = (1/2)(\delta_{k,6} + \delta_{k,8})$ and $v_\ell = (1/2)(\delta_{\ell,3} + \delta_{\ell,4})$ through the BEC.

APPENDIX E: CALCULATIONS IN THE BSC

Belief propagation and the Bethe approximation

In this section we write down the BP equations for a given code over the BSC or, equivalently, the cavity equations at the RS level. The expression of the free energy is also given. The cavity equations read

$$p_{\tau_i}^{(i \rightarrow a)} \propto \prod_{b \in i-a} q_{\tau_i}^{(b \rightarrow i)} e^{-\beta h_i \tau_i},$$

$$q_{\tau_i}^{(b \rightarrow i)} = \sum_{\tau_{b-i}} \prod_{j \in b-i} p_{\tau_j}^{(j \rightarrow b)} \delta[\tau_b = 1]. \quad (\text{E1})$$

$p_{\tau_i}^{(i \rightarrow a)}$ is the probability that the variable i takes the value τ_i in the absence of a , and $q_{\tau_i}^{(b \rightarrow i)}$ is proportional to the probability that the variable i takes the value τ_i when connected to b only.

Denoting $p_{\tau_i}^{(i \rightarrow a)} = e^{\beta h_{i \rightarrow a} \tau_i} / \cosh \beta h_{i \rightarrow a}$ and $q_{\tau_i}^{(b \rightarrow i)} = e^{\beta u_{b \rightarrow i} \tau_i} / \cosh \beta u_{b \rightarrow i}$, the cavity equations simplify to

$$h_{i \rightarrow a} = \hat{h}(h_i, \{u_{b \rightarrow i}\}) \equiv h_i + \sum_{b \in i-a} u_{b \rightarrow i},$$

$$u_{b \rightarrow i} = \hat{u}(\{h_{j \rightarrow b}\}) \equiv \frac{1}{\beta} \operatorname{arctanh} \left(\prod_{j \in b-i} \tanh \beta h_{j \rightarrow b} \right). \quad (\text{E2})$$

The local magnetization is given by $\langle \sigma_i \rangle = \tanh \beta H_i$, with $H_i = h_i + \sum_{a \in i} u_{a \rightarrow i}$. The Bethe approximation to the free energy reads

$$F_{\text{RS}}(\beta) = \sum_i \Delta F_i - \sum_a (k_a - 1) \Delta F_a, \quad (\text{E3})$$

with

$$\begin{aligned} \Delta F_i &= \Delta F_{\square}(\{u_{a \rightarrow i}\}) \equiv \frac{1}{\beta} \sum_{a \in i} \ln[2 \cosh(\beta u_{a \rightarrow i})] \\ &- \frac{1}{\beta} \ln \left[2 \cosh \left(\beta h_i + \beta \sum_{a \in i} u_{a \rightarrow i} \right) \right], \end{aligned}$$

$$\Delta F_a = \Delta F_{\square}(\{h_{i \rightarrow a}\}) \equiv -\frac{1}{\beta} \ln \left(\frac{1 + \prod_{i \in a} \tanh \beta h_{i \rightarrow a}}{2} \right). \quad (\text{E4})$$

Define

$$P(h) = \frac{1}{N\langle \ell \rangle} \mathbb{E}_C \left[\sum_{(i,a)} \delta(h - h_{i \rightarrow a}) \right],$$

$$Q(u) = \frac{1}{N\langle \ell \rangle} \mathbb{E}_C \left[\sum_{(i,a)} \delta(u - u_{a \rightarrow i}) \right]. \quad (\text{E5})$$

Averaging (E1) over the codes, the noise, and the edges, we obtain the self-consistency equations

$$P(h) = \sum_{\ell} \frac{\ell v_{\ell}}{\langle \ell \rangle} \int \prod_{a=1}^{\ell-1} du_a Q(u_a) \langle \delta[h - \hat{h}(h_{\xi}, \{u_a\})] \rangle_{h_{\xi}}, \quad (\text{E6})$$

$$Q(u) = \sum_k \frac{k c_k}{\langle k \rangle} \int \prod_{i=1}^{k-1} P(h_i) \delta[u - \hat{u}(\{h_i\})], \quad (\text{E7})$$

where $h_{\xi} = h_0$ with probability $1-p$ and $-h_0$ with probability p . The RS free energy reads

$$f_{RS}(\beta) = \sum_{\ell} v_{\ell} \int \prod_{a=1}^{\ell} du_a Q(u_a) \langle \Delta F_{\square}(\{h_{\xi}, \{u_a\}\}) \rangle_{h_{\xi}}$$

$$- \sum_k c_k (k-1) \int \prod_{i=1}^k dh_i P(h_i) \Delta F_{\square}(\{h_i\}). \quad (\text{E8})$$

Large deviations

As in the BEC, we study the statistics of BP over the codes, under the measure $\propto \exp[-x_f \beta_f F_{\text{cor}}(\beta_f) - x_e \beta_e F_{\text{RS}}(\beta_e)]$. The large deviation cavity equations read, for a regular code,

$$P(h) \propto \int \prod_{a=1}^{\ell-1} du_a Q(u_a) \frac{\langle \delta(h - h_{\xi} - \sum_{a=1}^{\ell-1} u_a) e^{\beta_f x_f h_{\xi}} \{2 \cosh[\beta_e (h_{\xi} + \sum_{a=1}^{\ell-1} u_a)]\}^{x_e} \rangle_{h_{\xi}}}{\prod_{a=1}^{\ell-1} [2 \cosh(\beta_e u_a)]^{x_e}},$$

$$Q(u) = \int \prod_{i=1}^{k-1} dh_i P(h_i) \delta \left[u - \frac{1}{\beta} \operatorname{arctanh} \left(\prod_{i=1}^{k-1} \tanh(\beta_p h_i) \right) \right], \quad (\text{E9})$$

and the potential

$$\phi(\beta_f, \beta_e, x_f, x_e) = \ln \int \prod_{a=1}^{\ell} du_a Q(u_a) \frac{\langle e^{\beta_f x_f h_{\xi}} \{2 \cosh[\beta_e (h_{\xi} + \sum_{a=1}^{\ell} u_a)]\}^{x_e} \rangle_{h_{\xi}}}{\prod_{a=1}^{\ell} [2 \cosh(\beta_e u_a)]^{x_e}}$$

$$- \frac{\ell}{k} (k-1) \ln \int \prod_{i=1}^k dh_i P(h_i) \left[\frac{1 + \prod_{i=1}^k \tanh(\beta_e h_i)}{2} \right]^{x_e}. \quad (\text{E10})$$

The solution to (E9) is obtained numerically. In the limit $k, \ell \rightarrow \infty$, this solution simplifies:

$$Q(u) = \delta(u), \quad P(h) = (1-p)\delta(h-h_0) + p\delta(h+h_0), \quad (\text{E11})$$

yielding the error exponent (83).

Another solution, called ‘‘type I’’ in [10], also exists:

$$Q(u) = \eta \delta_{+\infty}(u) + (1-\eta) \delta_{-\infty}(u),$$

$$P(h) = \nu \delta_{+\infty}(h) + (1-\nu) \delta_{-\infty}(h), \quad (\text{E12})$$

with

$$\nu = \frac{\eta^{\ell-1}}{\eta^{\ell-1} + (1-\eta)^{\ell-1} \langle e^{-2y h_0 \sigma} \rangle_{\sigma}}, \quad \eta = \frac{1}{2} [1 + (2\nu-1)^{k-1}]. \quad (\text{E13})$$

We automatically have $s_p=0$, and the condition $f_p=f_f$ implies $m=\beta_e x_e=1/2$. Then the rate function reads

$$L_1(f_p=f_f) = -\phi = -\ln[\eta^{\ell} + (1-\eta)^{\ell} \langle e^{-h_0 \sigma} \rangle_{\sigma}]$$

$$- \frac{\ell}{k} (k-1) \ln \left[\frac{1}{2} [1 + (2\nu-1)^k] \right]. \quad (\text{E14})$$

This solution (E12) is numerically unstable, and the rate function thus obtained is clearly unphysical. However, for $k, \ell \rightarrow \infty$, $\ell/k=1-R$, we have $\eta=\nu=1/2$ and the resulting rate function

$$L_1(f_p=f_f) = -\ln \frac{1}{2} [1 + 2\sqrt{p(1-p)}]$$

$$- R \ln 2 = \ln 2 [R_0(p) - R] \quad (\text{E15})$$

coincides with the error exponent of the RLM in the low- p regime (B14).

Two-step large deviations

The potential $\psi(\beta_e, m, y)$ defined in Eq. (87) is obtained by extremizing the following expression with respect to $P(h)$ and $Q(u)$:

$$\begin{aligned} \psi(\beta_e, m, y) = & \ln \int \prod_{a=1}^{\ell} du_a Q(u_a) \\ & \times \left[\frac{\langle e^{-mh\xi} \{ 2 \cosh[\beta_e(h\xi + \sum_{a=1}^{\ell} u_a)] \}^{m/\beta_e} \rangle_{h\xi}}{\prod_{a=1}^{\ell} [2 \cosh(\beta_e u_a)]^{m/\beta_e}} \right]^y \\ & - \frac{\ell}{k} (k-1) \ln \int \prod_{i=1}^k dh_i P(h_i) \\ & \times \left[\frac{1 + \prod_{i=1}^k \tanh(\beta_e h_i)}{2} \right]^{ym/\beta_e}. \end{aligned} \tag{E16}$$

We can only handle this calculation in the $k, \ell \rightarrow \infty$ limit. Equations (E11) are still a solution in this case and yield

$$\psi(\beta_e, m, y) = y \hat{\phi}(\beta_e, m), \tag{E17}$$

where $\hat{\phi}(\beta_e, m)$ is obtained from the average case. Therefore, the typical exponent is the same as the average error exponent in the high- p regime.

There also exists a counterpart of solution (E12), which gives

$$\begin{aligned} \psi(\beta_e, m, y) = & (R-1) \ln 2 + \ln \{ 1 + [(1-p)^{1-m} p^m \\ & + p^{1-m} (1-p)^m]^y \}. \end{aligned} \tag{E18}$$

The condition $\partial_m \psi = 0$ is again enforced by setting $m = 1/2$. Thus we get

$$\psi(y) = -yL - \mathcal{L} = (R-1) \ln 2 + \ln \{ 1 + [2\sqrt{p(1-p)}]^y \}. \tag{E19}$$

This expression yields the rate function $\mathcal{L}(L)$ by inverse Legendre transformation.

[1] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948); **27**, 623 (1948).
 [2] C. Berrou, A. Glavieux, and P. Thitimajshima, in *Proceedings of the IEEE International Conference on Communications, Geneva, 1993* (IEEE, New York, 1993), pp. 1064–1070.
 [3] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, England, 2003).
 [4] R. G. Gallager, *IRE Trans. Inf. Theory* **IT-8**, 21 (1962).
 [5] S. Verdú, *IEEE Trans. Inf. Theory* **44**, 2057 (1998).
 [6] E. R. Berlekamp, *Not. Am. Math. Soc.* **49**, 17 (2002).
 [7] A. Barg and G. D. Forney, Jr., *IEEE Trans. Inf. Theory* **48**, 2568 (2002).
 [8] C. Di, D. Proietti, I. E. Telatar, R. L. Urbanke, and T. J. Richardson, *IEEE Trans. Inf. Theory* **48**, 1570 (2002).
 [9] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, e-print cs.IT/0406060.
 [10] N. S. Skantzos, J. van Mourik, D. Saad, and Y. Kabashima, *J. Phys. A* **36**, 11131 (2003).
 [11] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, Oxford, 2001).
 [12] N. Surlas, *Nature (London)* **339**, 693 (1989).
 [13] F. den Hollander, *Large Deviations, Fields Institute Monographs No. 14* (American Mathematical Society, Providence, RI, 2000).
 [14] M. Mézard and G. Parisi, *Eur. Phys. J. B* **20**, 217 (2001).
 [15] O. Rivoire, *J. Stat. Mech.: Theory Exp.* 2005, P07004.
 [16] T. Mora and O. Rivoire, e-print cs.IT/0605130.
 [17] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
 [18] R. M. Tanner, *IEEE Trans. Inf. Theory* **27**, 533 (1981).
 [19] B. Bollobás, *Random Graphs*, 2nd ed. (Cambridge University Press, Cambridge, England, 2001).
 [20] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass Theory and Beyond, Vol. 9 of Lecture Notes in Physics* (World Scientific, Singapore, 1987).
 [21] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization, Algorithms and Complexity* (Prentice-Hall, Englewood Cliffs, NJ, 1982).
 [22] F. Ricci-Tersenghi, M. Weigt, and R. Zecchina, *Phys. Rev. E* **63**, 026702 (2001).
 [23] S. Cocco, O. Dubois, J. Mandler, and R. Monasson, *Phys. Rev. Lett.* **90**, 047205 (2003).
 [24] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina, *J. Stat. Phys.* **111**, 505 (2003).
 [25] M. Mézard and G. Parisi, *J. Stat. Phys.* **111**, 1 (2003).
 [26] Y. Kabashima and D. Saad, *J. Phys. A* **37**, R1 (2004).
 [27] A. Montanari, *Eur. Phys. J. B* **23**, 121 (2001).
 [28] S. Franz, M. Leone, A. Montanari, and F. Ricci-Tersenghi, *Phys. Rev. E* **66**, 046120 (2002).
 [29] M. Mézard, M. Palassini, and O. Rivoire, *Phys. Rev. Lett.* **95**, 200202 (2005).
 [30] A. Montanari and F. Ricci-Tersenghi, *Eur. Phys. J. B* **33**, 339 (2003).
 [31] Contrary to what indicates the last equations of [15], the nature of the order parameter is unchanged when additional levels of disorder are taken into account. The reason is that the cavity method encodes in a unique spatial distribution both the statistics over the nodes of a single graph and the statistics over the graphs in a ensemble. The discrimination between the two levels is done only through the unequal weighting attributed to the different nodes, as controlled by the two independent temperatures x and y .
 [32] T. Nattermann, in edited by A. P. Young *Spin Glasses and Random Fields* (World Scientific, Singapore, 1998).

- [33] A. Montanari and G. Semerjian, *Phys. Rev. Lett.* **94**, 247201 (2005).
- [34] A. Montanari and G. Semerjian, *J. Stat. Phys.* **124**, 103 (2006).
- [35] O. C. Martin, M. Mézard, and O. Rivoire, *J. Stat. Phys.* **P09006**, 2005.
- [36] C. Measson, A. Montanari, T. Richardson, and R. Urbanke, e-print cs.IT/0410028.
- [37] O. Rivoire and J. Barré, *Phys. Rev. Lett.* **97**, 148701 (2006).
- [38] J. van Mourik and Y. Kabashima, e-print cond-mat/0310177.
- [39] R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York, 1968).
- [40] S. Condamin, <http://www.inference.phy.cam.ac.uk/condamin/report.ps>
- [41] C. Di, A. Montanari, and R. Urbanke, in *Proceedings of the International Symposium on Information Theory, 2004* (IEEE, New York, 2004), p. 102.
- [42] In our case $v_{\ell}^{(k)} = \sum_{\ell' \geq \ell} v_{\ell'} \binom{\ell'}{\ell} c_k^{\ell} (1 - c_k)^{\ell' - \ell}$.