Message passing algorithms for non-linear nodes and data compression

Stefano Ciliberti[*] and Marc Mézard[†]

Laboratoire de Physique Théorique et Modèles Statistiques,

Université de Paris-Sud, bâtiment 100, 91405, Orsay Cedex, France

Riccardo Zecchina[‡]

ICTP, Strada Costiera 11, I-34100 Trieste, Italy

Abstract

The use of parity-check gates in information theory has proved to be very efficient. In particular, error correcting codes based on parity checks over low-density graphs show excellent performances. Another basic issue of information theory, namely data compression, can be addressed in a similar way by a kind of dual approach. The theoretical performance of such a Parity Source Coder can attain the optimal limit predicted by the general rate-distortion theory. However, in order to turn this approach into an efficient compression code (with fast encoding/decoding algorithms) one must depart from parity checks and use some general random gates. By taking advantage of analytical approaches from the statistical physics of disordered systems and SP-like message passing algorithms, we construct a compressor based on low-density non-linear gates with a very good theoretical *and* practical performance.

---

[*]Electronic address: ciliberti@roma1.infn.it

[†]Electronic address: mezard@lptms.u-psud.fr

[‡]Electronic address: zecchina@ictp.trieste.it

I. INTRODUCTION

Information theory and statistical physics have common roots. In the recent years, this interconnection has found further confirmation in the analysis of modern error correcting codes, known as Low Density Parity Check (LDPC) codes [1, 2, 3], by statistical physics methods [4, 5, 6]. In such codes, the choice of the encoding schemes maps into a graphical model: the way LDPC codes exploit redundancy is by adding bits which are bound to satisfy some random sparse linear set of equations – the so called parity checks. Such equations are used in the decoding phase to reconstruct the original codeword (and hence the message) from the corrupted signal. The parity checks can be represented by a graph indicating which variables participate in each check. The randomness and the sparsity of the parity checks is reflected into the characteristics of the associated graphs, a fact that makes mean-field type statistical physics methods directly applicable. Thanks to the duality between channel coding and source coding [3], similar constructions can be used to perform both lossless and lossy data compression [7, 8]. It must be mentioned that while there exist some practical algorithms for the lossless source coding which are very efficient [9], much less work has been done so far for the lossy case [10]. In this work we present a new coding technique which consists in a generalization of parity-check codes for lossy data compression to non-linear codes [11]. This issue has been addressed recently using methods from the statistical physics of disordered systems: a non-linear perceptron [12] or a satisfiability problem [13] have been used to develop practical source coding schemes. Other recent advances in the field can be found in [14, 15, 16].

We consider the following problem. We have a random string of uncorrelated bits $x_1, \ldots x_M$ ($\mathrm{prob}(x_a = 0) = \mathrm{prob}(x_a = 1) = 1/2$) and we want to compress it to the shorter coded sequence $\sigma_1, \ldots \sigma_N$ (encoding). The rate R of the process is $R = N/M$. Once we recover the message (decoding), we may have done some errors and thus we are left in principle with a different sequence $x_1^*, \ldots x_M^*$. The number of different bits normalized by the length of the string is the distortion,

$$D = \frac{1}{M} \sum_{a=1}^{M} (1 - \delta(x_a, x_a^*)) \ .$$ (1)

The Shannon theorem states that the minimum rate at which we can achieve a given distortion D is $R_{min} = 1 - H_2(D)$, where $H_2(D)$ is the binary entropy of the probability

distribution used to generate the x_a's. This R_{min} is the value one has to compare to in order to measure the performance of a compression algorithm. This problem is a simple version of the lossy data compression problem. In real applications one is often more interested in compression through quantization of a signal with letters in a larger alphabet (or continuous signal [17]). Here we shall keep to the memoryless case of uncorrelated unbiased binary sources, and we hope to be able to generalize the present approach to more realistic cases in the near future.

A. Constraint Satisfaction Problems

The underlining structure of the protocol that we are going to introduce is provided by a constraint satisfaction problem (CSP). A CSP with N variables is defined in general by a cost function

$$E[\sigma] = \sum_{a=1}^{M} \varepsilon_a \left(\{ \sigma_i \in a \} \right) \ , \tag{2}$$

where the *function node* $\varepsilon_a(\sigma_1, \ldots \sigma_{K_a})$ involves a subset of K_a variables randomly chosen among the whole set. The problem is fully defined by choosing some ε_a and an instance is specified by the actual variables involved in each constraint. A useful representation for a CSP is the factor graph (Fig. 1, left). The two kinds of nodes in this graph are the variables and the constraints. For the sake of simplicity we shall take the functions ε_a to have values in $\{0, 2\}$ (the clause is resp. SAT or UNSAT) and the variables σ_i will be Ising spins ($\sigma_i = \pm 1$). A somewhat special case of CSP which has been studied extensively occurs when $K_a = K$ for any a, and we shall restrict ourselves to this case in what follows. One can then show that the degree of a given variable is Poisson distributed with mean value KM/N and that the typical length of a loop is $\mathcal{O}(\log N / \log(KM/N))$. We are interested in the limit $M, N \to \infty$, where $\alpha \equiv M/N$ is fixed and plays the role of a control parameter. Once a problem is given in terms of its graph representation, an instance corresponds to a given graph chosen among all the legal ones with uniform probability. We will go back to this problem in section II. We first show how a CSP can be used as a tool for compressing data.
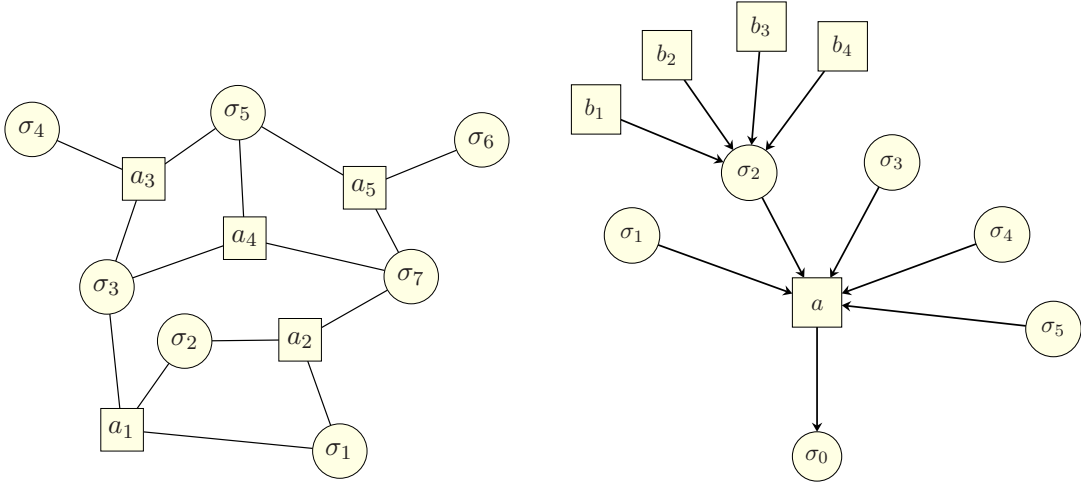
3

FIG. 1: Left: Factor graph for a constraint satisfaction problem with $N = 7$ variables (circles) and $M = 5$ constraints (squares). In this example the connectivity of the function nodes is kept fixed, that is $K_a = K$ in the notations of the text. Right: Message passing procedure ($K = 6$): The probability $q_{a \to 0}(u_{a \to 0})$ depends on all the probabilities $q_{b_i \to i}(u_{b_i \to i})$, $i = 1, \ldots 5$ (see text).

B. Encoding

We use the initial word $x_1, x_2, \ldots x_M$ to construct a set of M constraints between N boolean variables $\sigma_1, \sigma_2, \ldots \sigma_N$. The factor graph is supposed to be given. Each constraint a involves K_a variables and is defined by two complementary subsets of configurations $\mathcal{S}_0^a$ and $\mathcal{S}_1^a$. The value of x_a controls the role of the subsets as follows: If $x_a = 0$ then all the configurations $\{\sigma_1 \ldots \sigma_{K^a}\} \in \mathcal{S}_0^a$ satisfy the constraint, i.e. $\varepsilon_a = 0$; all the configurations $\{\sigma_1 \ldots \sigma_{K^a}\} \in \mathcal{S}_1^a$ do not satisfy the constraint and thus $\varepsilon_a = 2$. If $x_a = 1$ the role of the two subsets is exchanged. This defines the CSP completely and we can look for a configuration of σ_i that minimizes the number of violated constraints. This is the encoded word and the rate of the process is $R = 1/\alpha$.

C. Decoding

Given the configuration $\sigma_1, \sigma_2, \ldots \sigma_N$, we compute for each node a whether the variables $\sigma_{i_1^a}, \ldots \sigma_{i_{K_a}^a}$ are in $\mathcal{S}_0$ (in this case we set $x_a^* = 0$) or in $\mathcal{S}_1$ (leading to $x_a^* = 1$). Since a cost
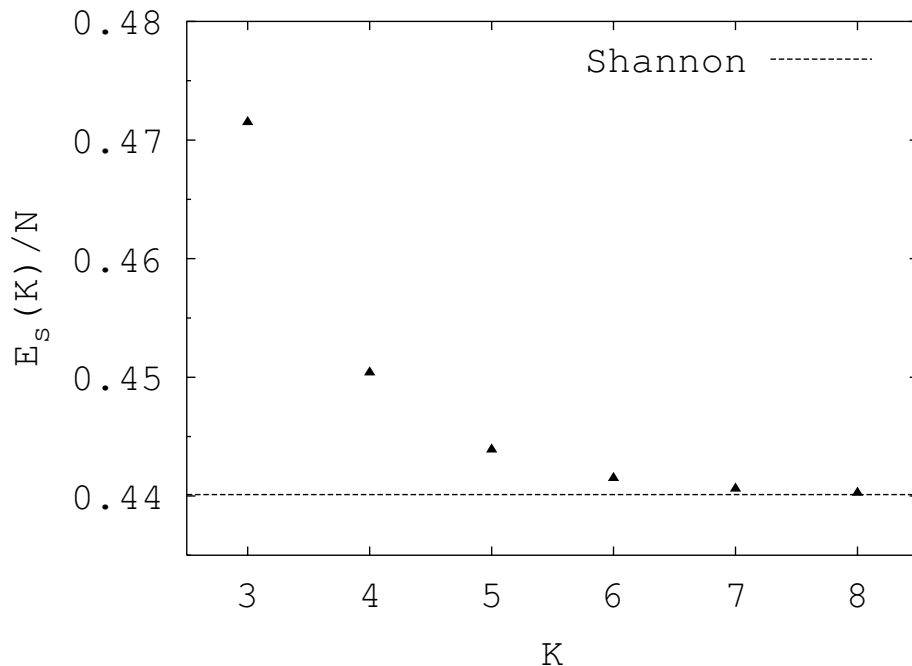
FIG. 2: The ground state energy density for the XORSAT problem at $\alpha = 2.0$ versus K. We also show the corresponding Shannon bound as computed from rate-distortion theory, $E_{Sh} = 2\alpha D_{Sh}$, where D_{Sh} is such that $1 - H_2(D_{Sh}) = 1/\alpha$.

$\varepsilon = 2$ is paid for each UNSAT clause, the energy is related to the distortion by

$$D = \frac{E}{2M} = \frac{E}{2\alpha N} = R\frac{E}{2N} \ . \tag{3}$$

D. The Parity Source Coder

In order to give an example, we apply these ideas to the case where the CSP is a parity check (the optimization problem is then called XORSAT). In other words, the cost of the constraint a is written as

$$\varepsilon_a = 1 + (-1)^{x_a} \prod_{i \in a} \sigma_i \ . \tag{4}$$

This Parity Source Coder (PSC) is the counterpart of the LDPC codes in error-correcting codes and then it seems natural to expect a good performance from it.

In Fig. 2 we show that the ground state energy of the XORSAT problem, that is the theoretical capacity of the PSC, quickly approaches the Shannon bound as K increases [18]. A PSC with checks of degree $K \gtrsim 6$ has then a theoretical capacity close to the optimal one. The problem here is that there does not exist any fast algorithm that can encode a

string (for example, the SID algorithm that we discuss in the next section is known not to converge).

We will thus investigate a new kind of constraint satisfaction problem, using non-linear nodes, with (almost) the same good theoretical performance as the XORSAT problem, but with a fast encoding algorithm. Before we do this, we introduce in the next section the general formalism used in the study of non-linear nodes.

II. THE GENERAL FORMALISM

Given a CSP at some α, we are interested in knowing whether a *typical* instance of the problem is satisfiable (i.e. all the constraints can be satisfied by one global configuration) as the number of variables goes to infinity. Generally speaking, there will be a phase transition between a SAT regime (at low α the problem has a small number of constraints and thus is solvable – at least in principle) and an UNSAT regime where there are too many constraints and one cannot find a zero-energy configuration. In this UNSAT regime, one wants to minimize the number of violated constraints. This kind of problem can be approached by message passing algorithms.

Due to the large length of typical loops, the local structure of a typical graph is equivalent to a tree, and this is crucial in what follows. We introduce the *cavity bias* $u_{a \to i} \in \{-1, 0, +1\}$ sent from a node a to a variable i. A non-zero message means that the variable i is requested to assume the actual value of $u_{a \to i}$ in order to satisfy the clause a. If $u_{a \to i} = 0$ the variable i is free to assume any value. It is clear that this message sent to i should encode the information that a receives from all the other variables attached to it. In order to clarify this point, we refer to Fig. 1 (right), that is we focus on a small portion of the graph. As the graph is locally tree-like, the variables σ_i, $i = 1, 2, ..., K-1$, are only connected through clause a if N is large enough. If a is absent, the total energy of the system can be written as $E^N(\sigma_1, ...\sigma_{K-1}) = A - \sum_{i=1}^{K-1} h_i \sigma_i$. We have used here the assumption that the probability of these σ_i factorizes. This is again motivated by the local tree structure, but we shall go back to this point. After clause a is added, $E^{N+1}(\sigma_0, \sigma_1, ...\sigma_{K-1}) = E^N(\sigma_1, ...\sigma_{K-1}) + \varepsilon_a(\sigma_0, \sigma_1, ...\sigma_{K-1})$ and the variables rearrange in order to minimize the total cost. The minimization then defines $\tilde{\varepsilon}(\sigma_0)$

from

$$A - \sum_{i=1}^{K-1} |h_i| + \tilde{\varepsilon}(\sigma_0) = \min_{\sigma_1, \dots \sigma_{K-1}} E^{N+1}(\sigma_0, \sigma_1, \dots \sigma_{K-1}) . \tag{5}$$

This $\tilde{\varepsilon}(\sigma_0)$ is then the cost to be paid for adding one variable with a fixed value σ_0. Without losing generality, it can be written as

$$\tilde{\varepsilon}(\sigma_0) \equiv \Delta_{a \to 0} - \sigma_0 u_{a \to 0} , \tag{6}$$

where $u_{a \to 0}$ is the cavity bias acting on the new variable and $\Delta_{a \to 0}$ is related to the actual energy shift by

$$\Delta E \equiv \min_{\sigma_0, \dots \sigma_{K-1}} \left[E^{N+1}(\sigma_0, \sigma_1, \dots \sigma_{K-1}) - E^N(\sigma_1, \dots \sigma_{K-1}) \right] = \Delta_{a \to 0} - |u_{a \to 0}| . \tag{7}$$

Given that ε_a can be 0 or 2, depending on the set of fields $h_1, \dots h_{K-1}$ one has four possibilities:

$$\tilde{\varepsilon}(+1) = 0 \quad \text{and} \quad \tilde{\varepsilon}(-1) = 0 \Rightarrow u = 0 , \quad \Delta = 0 \tag{8}$$

$$\tilde{\varepsilon}(+1) = 0 \quad \text{and} \quad \tilde{\varepsilon}(-1) = 2 \Rightarrow u = +1 , \Delta = 1 \tag{9}$$

$$\tilde{\varepsilon}(+1) = 2 \quad \text{and} \quad \tilde{\varepsilon}(-1) = 0 \Rightarrow u = -1 , \Delta = 1 \tag{10}$$

$$\tilde{\varepsilon}(+1) = 2 \quad \text{and} \quad \tilde{\varepsilon}(-1) = 2 \Rightarrow u = 0 , \quad \Delta = 2 . \tag{11}$$

In other words, a non-zero message u is sent from clause a to variable σ_0 only if the satisfiability of clause a depends on σ_0. A null message ($u = 0$) can occur in the two distinct cases (8) and (11).

The main hypothesis we have done so far consisted in assuming that the two variables are uncorrelated if they are distant (the energy is linear in the σ_i's). It turns out that, in a large region of the parameter space [19, 20], including the regime we are interested in, this is not correct. This is due to the fact that the space of solutions breaks into many disconnected components if α is greater than a critical value. In order to deal with this case, one has to introduce, for each directed link, a probability distribution of the cavity biases, namely $\mathsf{q}(u) \equiv \eta^+ \delta_{u,+1} + \eta^- \delta_{u,-1} + (1 - \eta^+ - \eta^-)\delta_{u,0}$. The hypothesis of no correlation holds if the phase space is restricted to one component. The interpretation of $\mathsf{q}_{a \to i}(u_{a \to i})$ is the probability that a cavity bias $u_{a \to i}$ is sent from clause a to variable i when one component is picked at random [21]. According to the rules (8,9,10,11), and with the topology of Fig. 1

(right) as a reference for notations, the Survey Propagation (SP, [22]) equations are then

$$\eta_{a\to 0}^+ \propto \text{Prob}\left[\left\{(u_{b\to i_1})_{b\in i_1\backslash a}, \dots (u_{b\to i_{K-1}})_{b\in i_{K-1}\backslash a}\right\}\middle| (\tilde{\varepsilon}_a(+1) < \tilde{\varepsilon}_a(-1))\right] e^{-y\Delta E} \ , \ (12)$$

$$\eta_{a\to 0}^- \propto \text{Prob}\left[\left\{(u_{b\to i_1})_{b\in i_1\backslash a}, \dots (u_{b\to i_{K-1}})_{b\in i_{K-1}\backslash a}\right\}\middle| (\tilde{\varepsilon}_a(+1) > \tilde{\varepsilon}_a(-1))\right] e^{-y\Delta E} \ , \ (13)$$

$$\eta_{a\to 0}^0 \propto \text{Prob}\left[\left\{(u_{b\to i_1})_{b\in i_1\backslash a}, \dots (u_{b\to i_{K-1}})_{b\in i_{K-1}\backslash a}\right\}\middle| (\tilde{\varepsilon}_a(+1) = \tilde{\varepsilon}_a(+1))\right] e^{-y\Delta E} \ , \ (14)$$

where the energy shift ΔE is given in eq. (7) and is non-zero only when the constraint is UNSAT for any value of σ_0 (cf. eq. (11)). The crucial reweighting term $\exp(-y\Delta E)$ thus acts as a "penalty" factor each time a clause can not be satisfied. This term is necessary in the UNSAT regime which we explore here (while simpler equations with $y = \infty$ are enough to study the SAT phase).

A. Encoding: SP and decimation

For each fixed value of y, the iterative solution for the SP equations can be implemented on a single sample [22], *i.e.* on a given graph where we know all the function nodes involved. The cavity probability distributions $\mathsf{q}(u)$'s are updated by picking up one edge at random and using (8,9,10,11). This procedure is iterated until convergence. This yields a set of messages $\{\eta_{a\to i}^-, \eta_{a\to i}^0, \eta_{a\to i}^+\}$ on each edge of the factor graph which is the solution of the SP equations. This solution provides very useful information about the single instance that can be used for decimation. As explained in [23], the finite value of y used for this purpose must be properly chosen. Given the solution for this y, one can compute the distribution $P(H_i)$ of the total bias $H_i = \sum_{a\in i} u_{a\to i}$ on each variable. One can then fix the most biased variables, *i.e.* the one with the largest $|P(H_i = +1) - P(H_i = -1)|$, to the value suggested by the $P(H_i)$ itself. This leads to a reduced problem with $N - 1$ variables. After solving again the SP equations for the reduced problem, the new most-biased variable is fixed and one goes on until the problem is reduced to an "easy" instance. This can be finally solved by some conventional heuristic (*e.g.* walksat or simulated annealing). A significant improvement of the decimation performance can be obtained by using a backtracking procedure [23, 24]: At each step we also rank the fixed variables with a strongly opposed bias and unfix the most "unstable" variable with finite probability. The algorithm described here is called Survey Inspired Decimation (SID) and its peculiar versions have been shown to be very useful in

many CSP problems recently. Unfortunately, the basic version of SID does not work for the XORSAT problem because of the symmetric character of the function nodes that is reflected in a large number of unbiased variables (some improvements appear to be possible [25]).

B. Statistical analysis: Theoretical performance

One can perform a statistical analysis of the solutions of the SP equations by population dynamics [21]. The knowledge of the function node ε_a allows to build up a table of values of u for each configuration of the local fields $h_1, \ldots, h_{K-1}$ (this is done according to the minimization procedure described above). We then start with some initial (random) population of $\vec{\eta}_i \equiv \{\eta_i^+, \eta_i^0, \eta_i^-\}$, for $i = 1, \ldots N$. We extract a Poisson number p of neighbors and p probabilities $\vec{\eta}_{i_1}, \ldots \vec{\eta}_{i_p}$. According to these weights, p biases $u_1, \ldots u_p$ are generated and their sum computed, $h_i = \sum_{a=1}^{p} u_a$. Once we have $K - 1$ of these fields we perform the minimization in (5) and compute the new probability for u_a according to the rules (8,9,10,11). The whole process is then iterated until a stationary distribution of the cavity biases is reached. This method for solving the SP equations is very flexible with respect to the change of the choice of the node, because this choice just enters in the calculation once at the beginning of the algorithm in order to initialize some tables. One can also study problems with many different types of nodes in a given problem: in this case, one among them is randomly chosen each time the updating is performed. We finally stress that once the probability distribution of the cavity biases is known the ground state energy of the problem can be computed according to the formalism introduced in [21]. The expression for the energy in term of the probability distributions as

$$E(\alpha) = \max_{y} \Phi(\alpha, y) , \tag{15}$$

$$\Phi(\alpha, y) = \frac{(-1)}{y} \left\{ [1 + (K-1)\alpha] \overline{\log A_p(y)} - (K-1)\alpha \overline{\log A_{p+1}(y)} \right\} , \tag{16}$$

$$A_p(y) \equiv \sum_{u_1, \ldots u_p} \mathsf{q}_1(u_1) \cdots \mathsf{q}_p(u_p) \exp\left(y \left| \sum_{a=1}^{p} u_a \right| - y \sum_{a=1}^{p} |u_a| \right) , \tag{17}$$

the overline standing for an average over the Poisson distribution and the choice of the distributions $\mathsf{q}_1, \ldots \mathsf{q}_p$ in the population. Recalling that $R = 1/\alpha$, the average distortion (*i.e.*, the theoretical capacity) of a compressor based on this CSP is computed through equations (3) and (15).
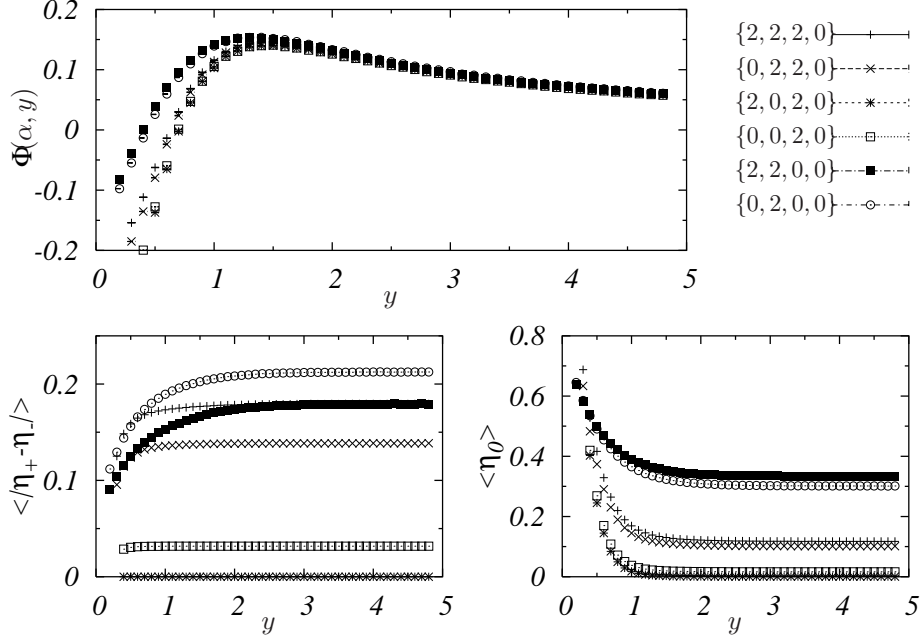
9

FIG. 3: Top: The free energy $\Phi(\alpha, y)$ of a number of symmetric nodes with $K = 7$ at $\alpha = 1.4$, classified according to the rule given in the text. Bottom: The asymmetry $|\eta_+ - \eta_-|$ of these nodes (left) is measured as well as the "paramagnetic degree" $\eta_0 = 1 - \eta_- - \eta_+$ (right). Note that the usual 7-XOR node corresponds to the $\{2, 0, 2, 0\}$ case in this notation.

Let us study here as an example a family of function nodes whose energy is fully invariant under permutations of the arguments. These nodes can be classified according to the energy of the node for a given value of the "magnetization" $m \equiv \sum_{i=1}^{K} \sigma_i$. We keep to odd K and label a particular node by the sequence of values $\{\varepsilon(m = K), \varepsilon(m = K - 2), \ldots, \varepsilon(m = 1)\}$. In Fig. 3 we report the results of the population dynamics algorithm for some types of nodes at $K = 7$. According to eq. (15), the ground state energy $E(\alpha)$ corresponds to the maximum over y of the free energy $\Phi(\alpha, y)$ represented in the top plot. The theoretical performance of these nodes are quite close to the PSC case (the XOR node, characterized by the sequence $\{2, 0, 2, 0\}$).

III. NON-LINEAR NODES

We consider in this section the function nodes that give the best performance for compression, both form the theoretical point of view (theoretical capacity close the parity-check nodes) and from the algorithmic aspect (the SID algorithm at finite y is found to converge in
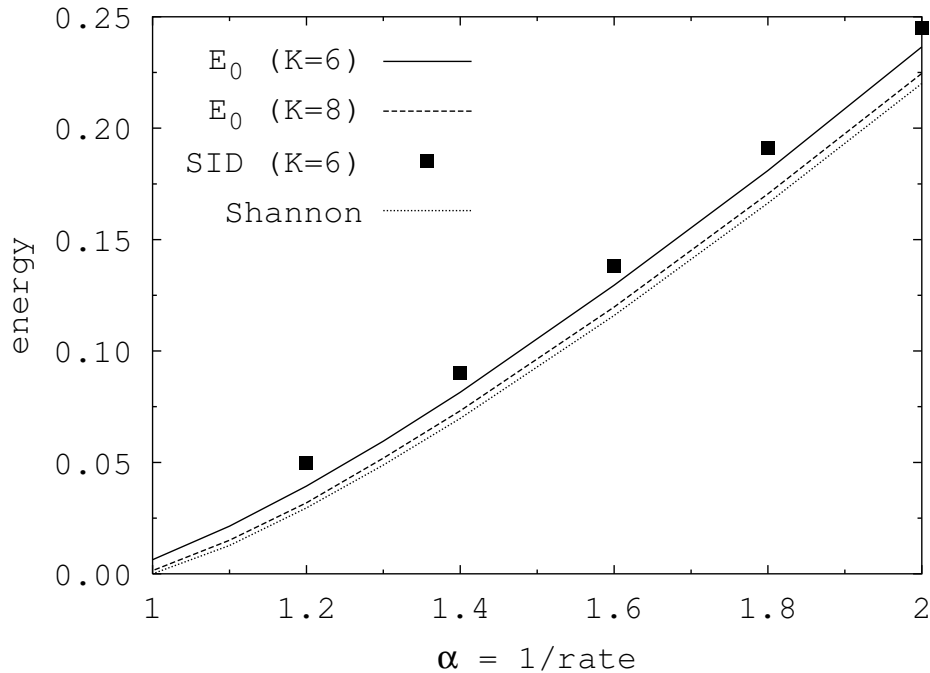
FIG. 4: The ground state energy for a system with 30 non-linear nodes. We also plot the Shannon bound and the performance of the SID algorithm for the $K = 6$ case.

the UNSAT regime, thus giving an explicit encoding algorithm). These non-linear function nodes are defined as follows. We recall that the output ε_{XOR} of a K-XORSAT node is given by twice the sum modulo 2 of the input bits. We label each configuration $\vec{\sigma} \in \{0, 1\}^K$ with an integer $l \in [1, 2^K]$ and consider a random permutation π of the vector $\{1, 2, \ldots, 2^K\}$. Then, we can associate the output of the *random node* by letting $\varepsilon^{(\pi)}(l) = \varepsilon_{XOR}(\pi(l))$. In this way we are left with a random but balanced output which can be different from XOR. Also, it is clear that they are not more defined by a linear formula over the boolean variables.

We can take advantage of the formalism introduced in the previous section in order to study the theoretical performance of these new function nodes. In particular, we have used 10 to 30 different random nodes to build the factor graphs. In order to improve the performance, we have also forbidden "fully-canalizing" nodes, that is nodes whose SAT character depends on just one variable.

In Fig. 4 we show our results. The ground state energy is shown to quickly approach the Shannon bound as K increases and for any α. As an example, at $\alpha = 2$ (corresponding to a compression rate $R = 1/2$) the difference between the $K = 8$ value and the theoretical limit is $\simeq 2\%$. This looks very promising from the point of view of data compression.

Furthermore, the results obtained from the SID (same plot) show that in this case the algorithm does converge and its performance is very good. It should be noted that when K becomes large the difference between SP and the Belief Propagation (BP) becomes small (this can be seen for instance in the analysis of [18]), so in fact BP does also provide a good encoding algorithm for $K \gtrsim 6$. At fixed K, the time needed to solve the SP equations is $\mathcal{O}(N \log N)$. The actual computational time required by the decimation process is of the same order and it slightly depends on the details of the SID algorithm (*e.g.* the number of variables fixed at each iteration, whether or not a backtracking procedure is used, which kind of heuristic is adopted to solve an "easy" reduced instance, the proper definition of the latter, etc...). The dependence on K is exponential. Thus, even if increasing K is good from the theoretical point of view, it turns out to be very difficult to work at high K. In practice, it takes a few hours to compress a string of $N = 1000$ bits at $K = 6$ by using our general purpose software. We think that some more specified code would lead to a better performance.

To conclude, we have shown how the methods of statistical mechanics, properly adapted to deal with a new class of constraint satisfaction problems, allow to implement a new protocol for data compression. The new tool introduced here, the non-linear gates, looks very promising for other practical applications in information theory.

Acknowledgments

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

[2] D. A. Spielman, in Lecture Notes in Computer Science **1279,** pp. 67-84 (1997).

[3] D. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, (2003).

[4] N. Sourlas, Nature **339,** 693-694 (1989).

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

[6] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing : An Introduction*, (Oxford University Press, 2001).

[7] F. Jelinek, IEEE Trans. Inform. Theory, **15,** No. 6, 584 (1969).

[8] E. Martinian, J. S. Yedidia, MERL report 2003.

[9] J. Ziv, A. Lempel, IEEE Trans. Inform. Theory **23,** 337 (1977); J. Cleary, I. Witten, IEEE Trans. Commun. **32,** 396 (1984); F. Willems, Y. Shtarkov, T. Tjalkens, IEEE Trans. Inform. Theory **41,** 653 (1995).

[10] T. Berger, J.D. Gibson, IEEE Trans. Inform. Theory, **44,** No. 6, 2693 (1998).

[11] Some of the results presented here have appeared in S. Ciliberti, M. Mézard, R. Zecchina, Phys. Rev. Lett. **95,** 038701 (2005).

[12] T. Hosaka, Y. Kabashima, H. Nishimori, Phys. Rev. E **66,** 066126 (2002).

[13] D. Battaglia, A. Braunstein, J. Chavas, R. Zecchina, Phys. Rev. E **72,** 015103 (2005).

[14] J. Garcia-Frias, Y. Zhao, IEEE Commun. Lett., 394 (Sept 2002).

[15] G. Caire, S. Shamai, S. Verdù, in Proc. of IEEE Inform. Theory Workshop, 291, Paris (2003);

[16] Y. Matsunaga, H. Yamamoto, in Proc. of IEEE Intern. Symp. on Inform. Theory, 461, Lausanne (2002).

[17] R. M. Gray, D. L. Neuhoff, IEEE Trans. Inform. Theory, **44,** 2325 (1998).

[18] S. Ciliberti, M. Mézard, `cond-mat/0506652.`

[19] M. Mézard, G. Parisi, R. Zecchina, Science **297,** 812 (2002).

[20] G. Biroli, R. Monasson, M. Weigt. Eur. Phys. J. **B14,** 551 (2000).

[21] M. Mézard, R. Zecchina, Phys. Rev. E **66,** 056126 (2002).

[22] A. Braunstein, M. Mézard, R. Zecchina, in *Random Structures and Algorithms* **27**, 201-226 (2005)

[23] D. Battaglia, M. Kolàr, R. Zecchina Phys. Rev. E **70,** 036107 (2004).

[24] G. Parisi, `cond-mat/0308510.`

[25] M. Wainwright, private communication.