ELSEVIER

Physica A 302 (2001) 14–21

PHYSICA A

www.elsevier.com/locate/physa

Statistical mechanics and capacity-approaching error-correcting codes

Nicolas Sourlas

Laboratoire de Physique Théorique de l', UMR 8549, Unité Mixte de Recherche du Centre National de la Recherche Scientifique et de l' Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris, Cedex 05, France

Abstract

I will show that there is a deep relation between error-correction codes and certain mathematical models of spin glasses. In particular, minimum error probability decoding is equivalent to finding the ground state of the corresponding spin system. The most probable value of a symbol is related to the magnetisation at a different temperature. Convolutional codes correspond to one-dimensional spin systems and Viterbi's decoding algorithm to the transfer matrix algorithm of statistical mechanics.

I will also show how the recently discovered (or rediscovered) capacity approaching codes (turbo codes and low-density parity check codes) can be analysed using statistical mechanics. It is possible to show, using statistical mechanics, that these codes allow error-free communication for signal to noise ratio above a certain threshold. This threshold depends on the particular code, and can be computed analytically in many cases. © 2001 Published by Elsevier Science B.V.

It has been known [1–4] that error-correcting codes are mathematically equivalent to some theoretical spin-glass models. As it is explained in Forney's paper in this volume, there have been recently very interesting new developments in error-correcting codes. It is now possible to approach practically very close to Shannon's channel capacity. First came the discovery of turbo codes by Berrou and Glavieux [5] and later the rediscovery of low-density parity check codes [6], first discovered by Gallager [7,8], in his thesis, in 1962. Both turbo codes and low-density parity check (LPDC) codes are based on random structures. It turns out, as I will explain later, that it is possible to use their equivalence with spin glasses, to analyse them using the methods of statistical mechanics.

E-mail address: sourlas@lpt.ens.fr (N. Sourlas).

PII: S0378-4371(01)00439-3

Let me start by fixing the notations. Each information message consists of a sequence of K bits $\vec{u}=\{u_1,\ldots,u_K\}$, $u_i=0$ or 1. The binary vector $\vec{u}$ is called the source-word. Encoding introduces redundancy into the message. One maps $\vec{u} \to \vec{x}$ by encoding. $\vec{u} \to \vec{x}$ has to be a one-to-one map for the code to be meaningful. The binary vector $\vec{x}$ has $N > K$ components. It is called a code-word. The ratio $R=K/N$ which specifies the redundancy of the code, is called the rate of the code. One particularly important family of codes are the so-called linear codes. Linear codes are defined by

$$\vec{x}=G\vec{u}\ .$$

G is a binary (i.e., its elements are zero or one) $(N\times K)$ matrix and the multiplication is modulo two. G is called the generating matrix of the code. Obviously, by construction, all the components x_i of a code-word x are not independent. Of all the 2^N binary vectors only $2^K=2^{NR}$, those corresponding to a vector $\vec{u}$, are code-words. Code-words satisfy the linear constraints (called parity check constraints) $H\vec{x}=0$ (modulo two), where H is a $(K\times N)$ binary matrix, called the parity check matrix. The connection with spin variables is straightforward. $u_i \to \sigma_i=(-1)^{u_i}$, $x_i \to J_i=(-1)^{x_i}$. It follows that $u_i+u_j=\sigma_i\sigma_j$ and

$$J_i=(-1)^{\sum_j G_{ij}u_j}=C^i_{k_1\cdots k_i}\,\sigma_{k_1}\cdots\sigma_{k_i}\ . \tag{1}$$

The previous equation defines the "connectivity matrix" C in terms of the generating matrix of the code G. Similarly, one can write the parity check constraints in the form

$$(-1)^{\sum_j H_{lj}x_j}=1 \to M^l_{k_1\cdots k_l}\,J_{k_1}\cdots J_{k_l}=1\ . \tag{2}$$

This defines the "parity constraint matrix" M in terms of the parity check matrix H of the code.

Code-words are sent through a noisy transmission channel and they get corrupted because of the channel noise. If a $J_i=\pm 1$ is sent, the output will be different, in general, a real number J_i^{out}. Let us call $Q(\vec{J}^{out}|\vec{J})\,\mathrm{d}\vec{J}^{out}$ the probability for the transmission channel's output to be between $\vec{J}^{out}$ and $\vec{J}+\mathrm{d}\vec{J}^{out}$, when the input was $\vec{J}$. The channel "transition matrix" $Q(\vec{J}^{out}|\vec{J})$ is supposed to be known. We will assume that the noise is independent for any pair of bits ("memoryless channel"), i.e.,

$$Q(\vec{J}^{out}|\vec{J})=\prod_i q(J_i^{out}|J_i)\ . \tag{3}$$

Communication is a statistical inference problem. Knowing the noise probability, i.e., $q(J_i^{out}|J_i)$, the code (i.e., in the present case of linear codes knowing the generating matrix G or the parity check matrix H) and the channel output $\vec{J}^{out}$, one has to infer the message that was sent. The quality of inference depends on the choice of the code.

We will now show that there exists a close mathematical relationship between error-correcting codes and theoretical models of disordered systems. To every possible information message (source word) $\vec{\tau}$ we can assign a probability $P^{source}(\vec{\tau}|\vec{J}^{out})$, conditional on the channel output $\vec{J}^{out}$. Or, equivalently, to any code-word $\vec{J}$ we can assign a probability $P^{code}(\vec{J}|\vec{J}^{out})$.

Because of Bayes theorem, the probability for any code-word symbol ("letter") $J_i = \pm 1$, $p(J_i|J_i^{out})$, conditional on the channel output J_i^{out}, is given by

$$\ln p(J_i|J_i^{out}) = c1 + \ln q(J_i^{out}|J_i) = c2 + h_i J_i \,, \tag{4}$$

where $c1$ and $c2$ are constants (nondepending on J_i) and

$$h_i = \frac{1}{2} \ln \frac{q(J_i^{out}| + 1)}{q(J_i^{out}| - 1)} \,. \tag{5}$$

It follows that

$$P^{code}(\vec{J}|\vec{J}^{out}) = c \prod_l \delta(M^l_{k_1 \cdots k_l} J_{k_1} \cdots J_{k_l}; 1) \exp\left(\sum_i h_i J_i\right) \,, \tag{6}$$

where c is a normalising constant. The Kronecker δ's enforce the constraint that $\vec{J}$ obeys the parity check equations (Eq. (2)), i.e., that it is a code-word. The δ's can be replaced by a soft constraint,

$$P^{code}(\vec{J}|\vec{J}^{out}) = const \exp\left[u \sum_l M^l_{k_1 \cdots k_l} J_{k_1} \cdots J_{k_l} + \sum_i h_i J_i\right] \,, \tag{7}$$

where $u \to \infty$. We now define the corresponding spin Hamiltonian by

$$-H^{code}(\vec{J}) = \ln P^{code}(\vec{J}|\vec{J}^{out}) = u \sum_l M^l_{k_1 \cdots k_l} J_{k_1} \cdots J_{k_l} + \sum_i h_i J_i \,. \tag{8}$$

This is a spin system with multispin interactions and an infinite ferromagnetic coupling and a random external magnetic field.

Alternatively, one may proceed by solving the parity check constraints $J_i = C^i_{k_1 \cdots k_i} \sigma_{k_1} \cdots \sigma_{k_i}$ (i.e., by expressing the code-words in terms of the source-words).

$$P^{source}(\vec{\sigma}|\vec{J}^{out}) = const \exp\left(\sum_i h_i C^i_{k_1 \cdots k_i} \sigma_{k_1} \cdots \sigma_{k_i}\right) \,, \tag{9}$$

where the h_i's are given as before. The logarithm of $P^{source}(\vec{\sigma}|\vec{J}^{out})$,

$$H^{source}(\vec{\sigma}) = -\ln P^{source}(\vec{\sigma}|\vec{J}^{out}) = -\sum_i h_i C^i_{k_1 \cdots k_i} \sigma_{k_1} \cdots \sigma_{k_i} \tag{10}$$

has obviously the form of a spin glass Hamiltonian.

We have given two different statistical mechanics formulations of error correcting codes; one in terms of the sourceword probability P^{source} and the other in terms of the code-word probability P^{code}. Because of the one to one correspondence between code-words and source-words, the two formulations are equivalent. In practice, however, it may make a difference. It may be more convenient to work with P^{source} or P^{code}, depending on the case. For the case of turbo codes (see later) it will be more convenient to define another probability, the "register word" probability.

It follows that the most probable sequence ("word MAP decoding") is given by the ground state of this Hamiltonian (H^{code} or H^{source}, depending on the case). Instead of considering the most probable instance, one may only be interested in the most probable value τ_i^p of the ith "bit" τ_i [9–11] ("symbol MAP decoding"). Because $\tau_i = \pm 1$, the probability p_i for $\tau_i = 1$ is simply related to m_i, the average of τ_i, $p_i = (1 + m_i)/2$.

$$m_i = \frac{1}{Z} \sum_{\{\tau_1 \cdots \tau_N\}} \tau_i \exp - H(\vec{\tau}) \qquad Z = \sum_{\{\tau_1 \cdots \tau_N\}} \exp - H(\vec{\tau}) \quad \tau_i^p = \mathrm{sign}(m_i)\,. \quad (11)$$

In the previous equation, m_i is obviously the thermal average at temperature $T = 1$. It is amusing to notice that $T = 1$ corresponds to Nishimori's temperature [12].

When all messages are equally probable and the transmission channel is memoryless and symmetric, i.e., when $q(J_i^{out}|J_i) = q(-J_i^{out}| - J_i)$, the error probability is the same for all input sequences. It is enough to compute it in the case where all input bits are equal to one, i.e., when the transmitted code-word is the all zero's code-word. In this case, the error probability per bit P_e is $P_e = (1 - m^{(d)})/2$, where $m^{(d)} = (1/N)\sum_{i=1}^N \tau_i^{(d)}$ and $\tau_i^{(d)}$ is the symbol sequence produced by the decoding procedure.

This means that it is possible to compute the bit error probability, if one is able to compute the magnetisation in the corresponding spin system.

Let me give a simple example of an $R = \frac{1}{2}$ "convolutional" code. From the N source symbols (letters) u_i's we construct the $2N$ code-word letters x_k^1, x_k^2, $k = 1, \ldots, N$:

$$x_i^1 = u_i + u_{i-1} + u_{i-2}, \qquad x_i^2 = u_i + u_{i-2}\,. \quad (12)$$

It follows that

$$J_k^1 = \sigma_k \sigma_{k-1} \sigma_{k-2}, \qquad J_k^2 = \sigma_k \sigma_{k-2}\,, \quad (13)$$

$$C^{(1,k)}_{i_{k_1} i_{k_2} i_{k_3}} = \delta_{k,i_{k_1}} \delta_{k,i_{k_2}+1} \delta_{k,i_{k_3}+2}, \qquad C^{(2,k)}_{i_{k_1} i_{k_3}} = \delta_{k,i_{k_1}} \delta_{k,i_{k_3}+2}\,. \quad (14)$$

The corresponding spin Hamiltonian is

$$-H = \frac{1}{w^2} \sum_k J_k^{1,out} \tau_k \tau_{k-1} \tau_{k-2} + J_k^{2,out} \tau_k \tau_{k-2}\,. \quad (15)$$

Here I assumed a Gaussian noise. In that case, Eq. (5) reduces to $h_k = J_k^{out}/w^2$, where w^2 is the variance of the noise. This is a one-dimensional spin-glass Hamiltonian. In fact, it is easy to see that convolutional codes correspond to one-dimensional spin systems. Their ground state can be found using the $T = 0$ transfer matrix algorithm. This corresponds to the Viterbi algorithm in coding theory. For symbol MAP (maximum a posteriori probability) decoding, one can use the $T = 1$ transfer matrix algorithm. This in turn is the BCJR algorithm in coding theory [13].

As it is explained in Forney's paper in this volume, the newly discovered (or re-discovered) capacity approaching codes are based on random constructions. Using the equivalence explained above, it has been possible to analyse them using the methods of statistical mechanics.

Gallager's low-density parity check (k, l) codes are defined by choosing at random a sparse parity check matrix H as follows. H has N columns (we consider the case of code-words of length N). Each column of H has k elements equal to one and all other elements equal to zero. Each row has l nonzero elements. It follows that H has Nk/l rows and that the rate of the code is $R = 1 - k/l$. It follows from Eq. (8) that Gallager's k, l codes correspond to diluted spin models with l-spin infinite strength ferromagnetic interactions in an external random field. It turns out that the belief propagation algorithm, used to decode LPDC codes, amounts to an iterative solution of the Thouless Anderson Palmer [14] (TAP) equations for spin glasses. A detailed analysis of these codes is presented in Urbanke's paper in this volume. Low-density parity check codes have been analysed using statistical mechanics methods by Kabashima Kanter and Saad [15,16] in the replica symmetric approximation. More recently, Montanari [17] was able to establish the entire phase diagram of LDPC codes. For $k, l \to \infty$ with k/l fixed, he showed that k, l codes correspond to a random energy model which can be solved without replicas. There is a phase transition in this model, which occurs at a critical value of the noise n_c. n_c separates a zero error phase from a high error phase. n_c in this case equals the value provided by Shannon's channel capacity. For finite k and l, he found an exact one-step replica symmetry breaking solution. The location of the phase transition determines n_c. In this way he computed also for finite values of k and l the critical value of the noise below which error free communication is possible. A different value of n_c, n_c^{bp} had already being computed by Richardson and Urbanke [18] (see Urbanke's paper in this volume). Richardson and Urbanke compute n_c^{bp} by analysing the behaviour of the decoding algorithm, belief propagation in this case. Statistical mechanics provides a threshold n_c which, in principle, is different from n_c^{bp}. n_c is reached by the optimum (but unknown) decoder.

Turbo codes also have been analysed using statistical mechanics [19,20]. Turbo Codes are based on recursive convolutional codes. An example of nonrecursive convolutional code was given in Eq. (12). The corresponding recursive code is given, most conveniently, in terms of the auxiliary bits b_i, defined below. The b_i's are stored in the encoder's memory registers, that's why I call $\vec{b}$ the "register word".

$$x_i^1 = u_i, \quad x_i^2 = b_i + b_{i-2}, \quad b_i = u_i + b_{i-1} + b_{i-2} . \tag{16}$$

It follows that the source letters u_i are given in terms of the auxiliary "register letters" b_i

$$u_i = b_i + b_{i-1} + b_{i-2} . \tag{17}$$

All additions are modulo two.

To construct a turbo code, one artificially considers a second source word $\vec{v}$, by performing a permutation, chosen at random, on the original code-word $\vec{u}$. So one considers $v_i = u_{P(i)}$ where $j = P(i)$ is a (random) permutation of the K indices i and a second "register word" c_i, $c_i = v_i + c_{i-1} + c_{i-2}$. Obviously,

$$v_i = c_i + c_{i-1} + c_{i-2} = u_j = b_j + b_{j-1} + b_{j-2}, \quad j = P(i) . \tag{18}$$

Eq. (18) can be viewed as a constraint on the two register words $\vec{b}$ and $\vec{c}$. Finally in the present example, a rate $R = \frac{1}{3}$ turbo code, one transmits the $3K$ letter code-word $x_i^1 = u_i,\ x_i^2 = b_i + b_{i-2},\ x_i^3 = c_i + c_{i-2},\ i = 1, \dots, K$. Let us call, as before,

$$J_i^\alpha = (-1)^{x_i^\alpha}, \quad \alpha = 1, 2, 3 ,$$

the channel inputs and $J_i^{out,\alpha}$ the channel outputs. In the previous, for reasons of convenience, we formulated convolutional codes using the source-word probability P^{source} and LDPC codes using the code-word probability P^{code}. The statistical mechanics of turbo codes is most conveniently formulated in terms of the "register words" probability $P^{reg}(\vec{\sigma}, \vec{\tau} | \vec{J}^{out})$ conditional on the channel outputs $\vec{J}^{out}$, where $\tau_i = (-1)^{b_i}$ and $\sigma_i = (-1)^{c_i}$. The logarithm of this probability provides the spin Hamiltonian

$$-H = \frac{1}{w^2} \sum_k J_k^{out,1} \tau_k \tau_{k-1} \tau_{k-2} + J_k^{out,2} \tau_k \tau_{k-2} + J_k^{out,3} \sigma_k \sigma_{k-2} . \tag{19}$$

Because of Eq. (18), the two spin chains $\vec{\tau}$ and $\vec{\sigma}$ obey the constraints

$$\sigma_i \sigma_{i-1} \sigma_{i-2} = \tau_j \tau_{j-1} \tau_{j-2}, \quad j = P(i) . \tag{20}$$

(As previously, we have considered the case of a Gaussian noise of variance w^2.) This is an unusual spin Hamiltonian. Two short range one-dimensional chains are coupled through the infinite range, nonlocal constraint, Eq. (20). This constraint is nonlocal because neighbouring i's are not mapped to neighbouring j's under the random permutation. It turns out that this Hamiltonian can be solved by the replica method. One finds a phase transition at a critical value of the noise n_{crit}. For noises less than n_{crit}, the magnetisation equals one, i.e., it is possible to communicate error free. In this respect, turbo codes are similar to Gallager's LDPC codes. The statistical mechanical models, however, are completely different. Let me also mention that, under some reasonable assumptions, the iterative decoding algorithm for turbo codes (turbodecoding algorithm), which I am not explaining here, can be viewed [20] as a time discretisation of the Kolmogorov, Petrovsky and Piscounov equation [21]. This KPP equation has traveling wave solutions. The velocity of the traveling wave, which is analytically computable, corresponds to the convergence rate of the turbodecoding algorithm. The agreement with numerical simulations is excellent.

So the equivalence between linear codes and theoretical models of spin glasses is quite general and we have established the following dictionary of correspondence:

Error-correcting code ⇔ *Spin Hamiltonian*

Signal to noise ⇔ $J_0^2/\Delta J^2$

Maximum likelihood Decoding ⇔ *Find a ground state*

Error probability per bit ⇔ *Ground state magnetisation*

Sequence of most probable symbols ⇔ *magnetisation at temperature* $T = 1$

Convolutional Codes ⇔ *One-dimensional spin-glasses*

$$\textit{Viterbi decoding} \Leftrightarrow T = 0 \textit{ Transfer matrix algorithm}$$

$$\textit{BCJR decoding} \Leftrightarrow T = 1 \textit{ Transfer matrix algorithm}$$

$$\textit{Gallager LDPC codes} \Leftrightarrow \textit{Diluted p-spin ferromagnets in a random field}$$

$$\textit{Turbo Codes} \Leftrightarrow \textit{Coupled spin chains PC}$$

$$\textit{Zero error threshold} \Leftrightarrow \textit{Phase transition point}$$

$$\textit{Belief propagation algorithm} \Leftrightarrow \textit{Iterative solution of TAP equations}$$

I would like to conclude by pointing out some open questions.

What is the order of the phase transition? This question is particularly relevant for turbo codes and has important implications for decoding.

What are the finite size effects? This question is particularly relevant near the zero error noise threshold (i.e., near the phase transition). The answer will depend on the order of the transition.

How does the decoding complexity behave as one approaches the zero error noise threshold? Is there a critical slowing down? As it was said before, the decoding algorithms both for LDPC codes and turbo codes are heuristic and there are not known results as one approaches the phase transition.

Is there a glassy phase in decoding? In other terms, do the heuristic decoding algorithms reach the threshold of optimum decoding, computed by statistical mechanics, or is there a (lower) noise "dynamical" threshold where decoding stops reaching optimal performance?

I hope that at least some of the above questions will be answered in the near future.

References

[1] N. Sourlas, Nature 339 (1989) 693.
[2] N. Sourlas, in: L. Garrido (Ed.), Statistical Mechanics of Neural Networks, Lecture Notes in Physics, Vol. 368, Springer, Berlin, 1990.
[3] N. Sourlas, Ecole Normale Supérieure preprint, April 1993.
[4] N. Sourlas, in: P. Grassberger, J.-P. Nadal (Eds.), From Statistical Physics to Statistical Inference and Back, Kluwer Academic Publishers, Dordrecht, 1994, p. 195.
[5] C. Berrou, A. Glavieux, P. Thitimajshima, Proceedings of the 1993 International Conference on Communication, pp. 1064–1070.
[6] D.J.C. MacKay, R.M. Neal, Electron Lett. 33 (1997) 457.
[7] R.G. Gallager, IRE Trans. Inform. Theory IT-8 (1962) 21.
[8] R.G. Gallager, Low-Density Parity-Check Codes, MIT Press, Cambridge, MA, 1963.
[9] P. Ruján, Phys. Rev. Lett. 70 (1993) 2968.
[10] H. Nishimori, J. Phys. Soc. Japan 62 (1993) 2973.
[11] N. Sourlas, Europhys. Lett. 25 (1994) 169.
[12] H. Nishimori, Progr. Theor. Phys. 66 (1981) 1169.
[13] L. Bahl, J. Cocke, F. Jelinek, J. Raviv, IEEE Trans. Inf. Theory IT-20 (1974) 248.
[14] D.J. Thouless, P.W. Anderson, R.G. Palmer, Philos. Mag. 35 (1977) 593.

[15] I. Kanter, D. Saad, Phys. Rev. Lett. 83 (1999) 2660.
[16] Y. Kabashima, T. Murayama, D. Saad, Phys. Rev. Lett. 84 (2000) 1355.
[17] A. Montanari, cond-mat/0104079.
[18] T.J. Richardson, R.L. Urbanke, IEEE Trans. Inform. Theory 47 (2001) 638.
[19] A. Montanari, N. Sourlas, Eur. Phys. J. B 18 (2000) 107.
[20] A. Montanari, Eur. Phys. J. B 18 (2000) 121.
[21] A. Kolmogorov, I. Petrovsky, N. Piscounov, Moscow Univ. Math. Bull. 1 (1937) 1.