Finite Size Scaling and Metastable States of Good Codes*

Andrea Montanari

Laboratoire de Physique Théorique
de l'Ecole Normale Supérieure†
24, rue Lhomond, 75231 Paris CEDEX 05, FRANCE
Internet: Andrea.Montanari@lpt.ens.fr

Abstract

I discuss some recent results on Gallager codes obtained through the statistical mechanics approach. The results concern: A. The behavior of finite length codes: I propose a characterization which is complementary to the one given by the error exponent; B. The performance of *local* decoding algorithms: I show that they fail above the threshold for belief propagation decoding. Point B lead me to conjecture that any linear-time decoder fail above the threshold for belief propagation decoding.

1 Introduction

In this paper I will present some recent investigations of Gallager codes [1] within the statistical mechanics approach [2]. Although I will always refer to Gallager codes, the results presented in this paper can be probably generalized to much larger classes of codes. The basic requirement is that the code ensemble must have vanishing block error probability in the infinite length limit. This is why I referred to *good* codes in the title of this paper.

Before continuing it can be helpful to define the basic notations. I will denote by N the block length of the code. For sake of simplicity I will refer to two particular models for the noisy channel: the binary symmetric channel, mainly in Sec. 2, and the binary erasure channel in Sec. 3. In both cases will denote the noise level by p. The precise meaning (erasure probability or flipping probability) will be determined, when necessary, by the context.

I will discuss the following topics:

A. In Section 2 I will discuss the *finite size scaling* of Gallager codes. Characterizing the behavior of finite length codes is a problem of outmost practical relevance. Coding theory focus on the $N \to \infty$ limit at fixed noise level p and characterizes it through the error exponent $E(p)$. More explicitly the block error probability vanishes as $P_{\rm bl} \sim e^{-NE(p)}$. The error exponent vanishes at some critical noise p_*.

*This work was supported through a European Community Marie Curie Fellowship.

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

near to the the threshold p_*, this description is not accurate.

Finite size scaling describes a different asymptotics: namely $p \to p_*$ and $N \to \infty$ at the same time,while a product of the form $(p - p_*)N^{1/\nu}$ is kept fixed. We will show, through numerical simulations, that this type of analysis can be pertinent for Gallager codes.

B. In Section 3 I will describe the role of *metastable states* in the decoding dynamics of Gallager codes. We will work in a simplified setting: the binary erasure channel [11].

In order to introduce the concept of metastable state it is convenient to use the language of combinatorial optimization [3]. Decoding can be regarded as a combinatorial optimization problem. We receive a message $\underline{x}$ containing some erasures and we want to find a bit sequence $\underline{z}$ (compatible with the channel output $\underline{x}$) which maximize the number of satisfied parity checks. Metastable states are *locally* optimal states of this optimization problem.

Recall that, for Gallager codes, the critical noise level under belief propagation decoding (we shall call it p_d) is, in general, different from the threshold under maximum likelihood decoding (p_c).

The concept of metastable state gives us an intrinsic (i.e. algorithm-independent) characterization of p_d. We will show that, at p_d, an exponential (in N) number of metastable states appears. This leads us to formulate the following conjecture: *Gallager codes cannot be decoded in linear time between p_d and p_c.*

Apparently there is no connection between the two topics outlined above. However the point B give us an intrinsic characterization of decoding failures. This could be the first step for an analytical computation of finite-length effects in Gallager codes (point A).

The results presented in this paper have been obtained in the context of two collaborations. The first one [4] concerns the point A and involves F. Rosati (Rome) and N. Sourlas (Paris). The second one [5] focus on point B and involves S. Franz, M. Leone, F. Ricci-Tersenghi and R. Zecchina (Trieste).

2 Finite size scaling[1]

Finite size scaling is an ubiquitous phenomenon both in combinatorics and in statistical physics systems.

In order to introduce it, we shall consider a very simple example, the link percolation transition on the complete graph [6]. Suppose you have N sites $i = 1, \ldots, N$. Any of the $N(N-1)/2$ pairs of can be either connected or not by a link. Let us say that a link is present with probability p/N and that two links between different pairs of sites are independent. This defines our model.

The most striking phenomenon in this model is the percolation transition. Let us define a cluster as a set of sites connected by links, and its size as the number of sites which belong to it. For $p < p_c = 1$ the clusters have a size at most of order $\log N$. For $p > p_c$ a single giant (i.e. with size of order N) cluster appears. Let us define its (average) size to be $G(p, N)N$. In the $N \to \infty$ limit the size is given by the following equation

$$1 - G = \exp\{-p\, G\}\,. \tag{1}$$

[1] In collaboration with F. Rosati and N. Sourlas.

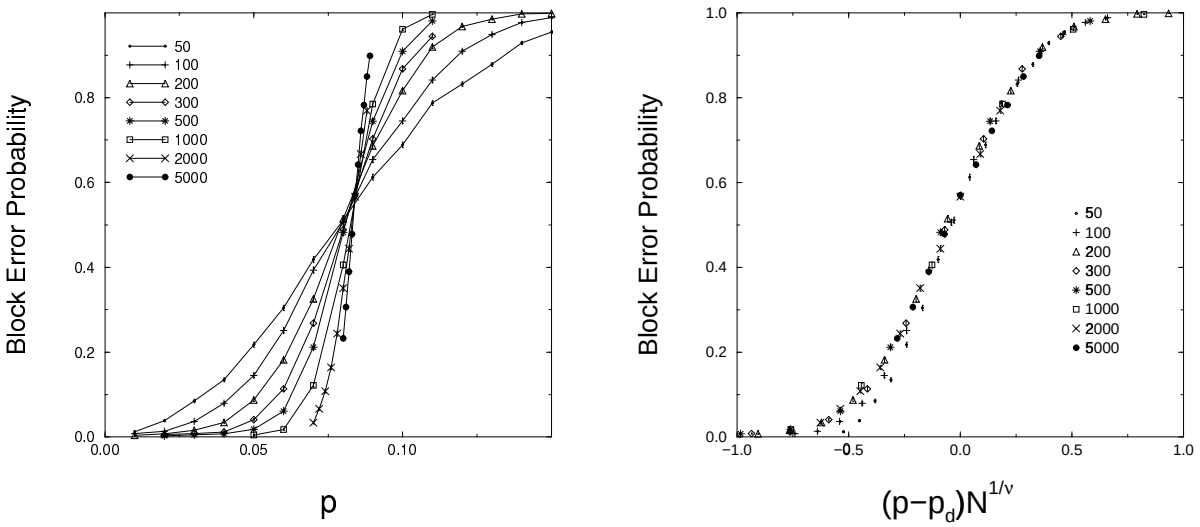

Figure 1: The block error probability for a $(6, 3)$ regular Gallager code for several block lengths (see the legend). On the right the *scaling plot* (see text) for the same quantity.

In particular the size $G(p, \infty)$ of the giant cluster is given, near to the transition, by $G(p, \infty) \sim 2(p - p_c)$. This non-analytic behavior cannot hold for finite N. In fact it is smoothened as follows:

$$G(p, N) \approx N^{-\beta/\nu} \mathcal{G}[(p - p_c) N^{1/\nu}] \,. \tag{2}$$

This equality is understood to hold asymptotically in the limit $p \to p_c$, $N \to \infty$ with the product $(p - p_c)N^{1/\nu}$ kept fixed. The *critical exponents* $\beta = 1$ and $\nu = 3$ have been exactly computed [7]. The function $\mathcal{G}$ is called a *scaling function*.

There are at least two reasons for thinking that the same scenario must hold for Gallager codes. In the statistical mechanics approach [2], the transition between between error-free communication and the high noise region is described by a phase transition. This is exactly the context to which finite size scaling theory applies [8]. Moreover Gallager codes share some similarities with the percolation example presented above. For instance, on the binary erasure channel the transition at p_d can be described as a particular percolation transition on the hypergraph defining the code [9].

In order to check this hypothesis we simulated belief propagation decoding for the $(6, 3)$ regular Gallager code on the binary symmetric channel. We averaged the results over 5000 codes of the ensemble and realizations of the channel noise. We varied the number of iterations of belief propagation algorithm between 50 and 1000, checking that the dependence upon this parameter was negligeable.

The results for the block error probability are shown in Fig. 1, left frame. It can be noticed that, as N increases, the transition at p_d becomes sharper and sharper. Indeed the finite size scaling form (2) predicts the width of the transition to be of order $N^{-1/\nu}$. More explicitly we can guess that, in the present case, the correct finite-size form is

$$P_{\rm bl}(p, N) \approx \mathcal{F}_{\rm bl}[(p - p_d) N^{1/\nu}] \,. \tag{3}$$

Notice that in this case there is no prefactor as in Eq. (2). This is due to fact that, in the $N \to \infty$ limit, $P_{\rm bl} \to 1$ above p_d, and $P_{\rm bl} \to 0$ below p_d. In order to check the

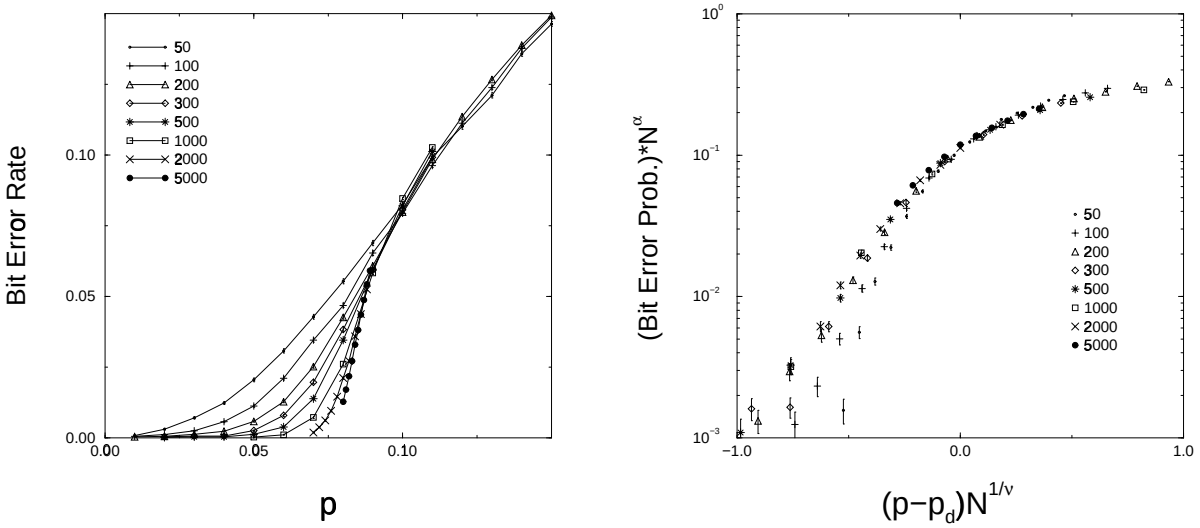

Figure 2: The bit error probability for a $(6, 3)$ regular Gallager code (on the left) and the corresponding *scaling plot* (on the right).

Ansatz (3) we plot the same data versus $(p-p_d)N^{1/\nu}$ in the left frame of Fig. 1. Here we used $p_d \approx 0.0842$ (which can be obtained either through density evolution [10] or in the statistical mechanics approach [5]) and $\nu = 2$. The good collapse of the data supports the relevance of the Ansatz (3).

We can go further and study the bit error probability. In Fig. 2 we present the corresponding data. The finite size scaling form is, in this case,

$$P_{\rm bit}(p, N) \approx N^{-\alpha}\mathcal{F}_{\rm bit}[(p - p_d)N^{1/\nu}]\,. \tag{4}$$

In order to verify this prediction we plot on Fig. 2, right frame, $N^{\alpha}P_{\rm bit}(p, N)$ versus $(p - p_d)N^{1/\nu}$. We used $p_d \approx 0.0842$ and $\nu = 2$ as in Fig. 1 and $\alpha = 0.15$. Once again there is a good data collapse.

3 Metastable states[2]

In this Section we shall focus on the binary erasure channel. We shall treat decoding as a combinatorial optimization problem within the space of bit sequences of length N_E (the number of erased bits, the others being fixed by the received message). The function to be minimized is the *energy density*

$$\epsilon = \frac{2}{N} \cdot (\text{number of violated parity checks})\,, \tag{5}$$

where we introduced the normalizing factor for future convenience. In Section 1 we described metastable states as *locally* optimal states (bit sequences) of the decoding problem. Now we must say what do we mean by *locally*.

Since we deal with a combinatorial optimization problem on bit sequences of length N_E, there is a natural notion of distance to use: the Hamming distance. Let us denote

[2]In collaboration with S. Franz, M. Leone, F. Ricci-Tersenghi and R. Zecchina.

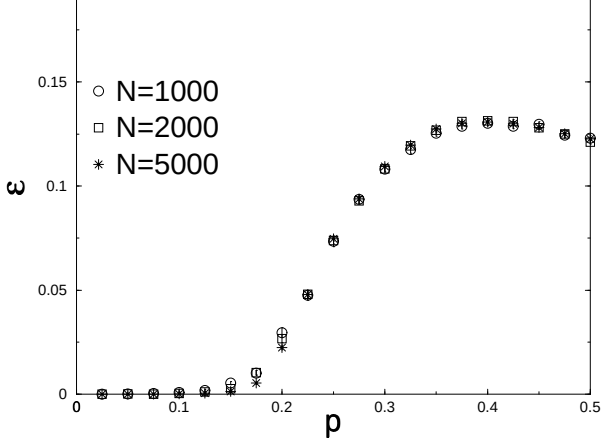

Figure 3: The $(6,3)$ Gallager code decoded by local search. We plot the number of violated parity checks (multiplied by $2/N$) as a function of the erasure probability p.

as $d_{\rm H}(\underline{x}_1,\underline{x}_2)$ the Hamming distance between the two sequences $\underline{x}_1$ and $\underline{x}_2$. We can define the k-neighborhood of a given sequence $\underline{x}$ as the set of sequences $\underline{z}$ such that $d_{\rm H}(\underline{x},\underline{z}) \le k$. Finally we call k-stable states the bit sequences which are optima of the decoding problem within their k-neighborhood.

One can easily invent local search algorithms [3] for the decoding problem which use the k-neighborhoods. The algorithm start from a random sequence and, at each step, optimize it within its k-neighborhood. This algorithm is clearly suboptimal and halts on k-stable states. Let us consider, for instance, a $(6,3)$ regular code and decode it by local search in 1-neighborhoods. We recall that such a code has $p_d \approx 0.429440$ and $p_c \approx 0.488151$ [11]. In Fig. 3 we report the resulting energy density ϵ after the local search algorithm halts, as a function of the erasure probability p. We averaged over 100 different realizations of the noise and codes in the ensemble.

Let us now turn to *metastable* states. This are k-stable states for any $k = o(N)$[3]. Notice that this definition is slightly ambiguous: we do not know how to compare k-stable states for different values of N. A possible approach for avoiding this ambiguity is the following: work with k-stable states, take the $N \to \infty$ limit and, at the end, the $k \to \infty$ limit. It is quite clear that no local search algorithm can escape these states.

The replica method [12] allows to compute the number of metastable states with a given energy density ϵ [13, 14]. In particular it yields

$$\mathcal{N}_{MS}(\epsilon) \sim \exp\{N\Sigma(\epsilon)\}\,. \tag{6}$$

The function $\Sigma(\epsilon)$ is called the complexity. The computation of $\Sigma(\epsilon)$ can be done within a variational (approximate) scheme. We report in Fig. 4 the result of this computation for three different values of the erasure probability p. As before we consider the $(6,3)$ code. The general picture is as follows. Below p_d there is no metastable state, excepting the one corresponding to the correct codeword. Between p_d and p_c there is an exponential number of metastable states with energy density belonging to the interval $\epsilon_G S < \epsilon < \epsilon_D$. Above p_c, $\epsilon_{GS} = 0$. The maximum of $\Sigma(\epsilon)$ is always at ϵ_D.

The above picture tell us that any local algorithm will run into difficulties above p_d. In order to confirm this picture, we made some numerical computations using simulated

[3] I use the standard notation: $f_N = o(N)$ if $\lim_{N\to\infty} f_N/N = 0$.

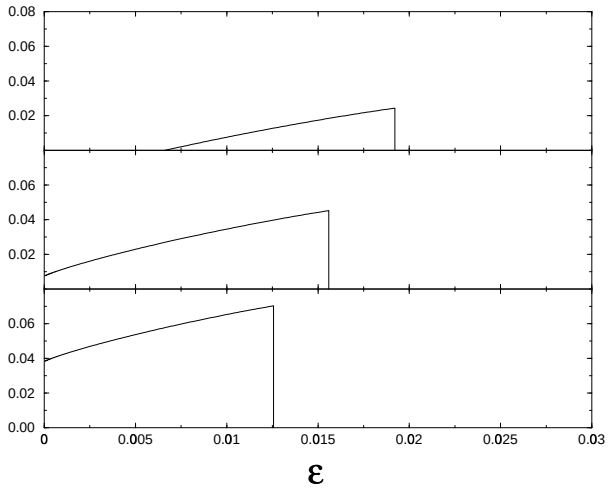

Figure 4: The complexity $\Sigma(\epsilon)$ for (from top to bottom) $p = 0.45$ (below p_c), $p = 0.5$, and $p = 0.55$ (above p_c).

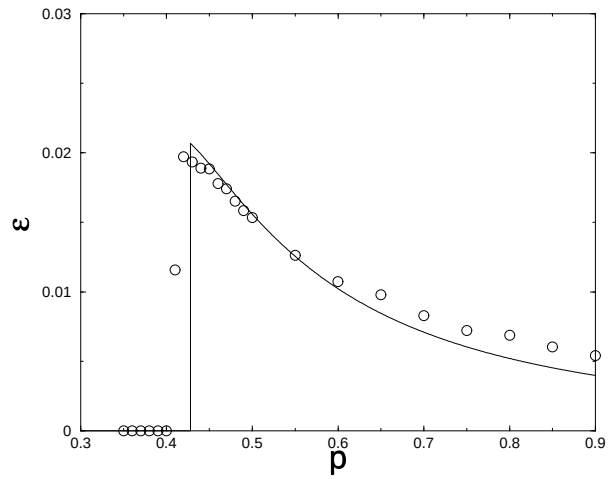

Figure 5: The $(6, 3)$ Gallager code decoded by simulated annealing. The circles give the number of violated checks in the resulting sequence. The continuous line is the analytical result for the typical energy density of metastable states.

together with the theoretical prediction for ϵ_D. The good agreement confirm our picture: the algorithm gets stucked in metastable states, which have, in the great majority of cases, energy density ϵ_D.

Both message passing and local search algorithms fail decode correctly between p_d and p_c. This leads us to formulate the conjecture already mentioned in the Introduction: no linear time algorithm can decode in this regime of noise. The (typical case) computational complexity change from being linear below p_d to superlinear above p_d. In the case of the binary erasure channel it remains polynomial between p_d and p_c. However it is plausible that for a general channel it becomes non-polynomial.

References

[1] R. G. Gallager. *Low Density Parity Check Codes*, Research Monograph Series Vol. 21 (MIT, Cambridge, MA., 1963).

[2] See N. Sourlas contribution to this conference and references therein.

[3] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization* (Prentice-Hall,Inc., Englewood Cliffs, New Jersey, 1982).

[4] A. Montanari, F. Rosati and N. Sourlas, in preparation.

[5] S. Franz, M. Leone, A. Montanari, F. Ricci-Tersenghi and R. Zecchina, in preparation.

[6] B. Bollobas, *Random Graphs* (Academic Press, London, 1985).

[7] T. Łuczak, Rand. Struc. Alg. **1** (1990), 287.

[8] M. Barber in *Phase Transitions and Critical Phenomena*, vol. 8 (Academic Press, London, 1983).

[9] C. Di, D. Proietti, T. Richardson, E. Telatar and R. Urbanke, *Finite length analysis of low-density parity-check codes*, submitted to IEEE Trans. on Inf. Theory.

[10] T. Richardson, and R. Urbanke, in *Codes, Systems, and Graphical Models*, pp 1–37 (Springer, New York, 2001).

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

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

[13] R. Monasson, Phys. Rev. Lett. 75, 2847-2850 (1995)

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