PHYSICAL REVIEW LETTERS

# Clustering of Solutions in the Random Satisfiability Problem

M. Mézard,[1] T. Mora,[1] and R. Zecchina[2]

[1]*Laboratoire de Physique Théorique et Modèles Statistiques, bâtiment 100, Université Paris-Sud, F-91405 Orsay, France.*
[2]*Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34100 Trieste, Italy*

Using elementary rigorous methods we prove the existence of a clustered phase in the random $K$-SAT problem, for $K \geq 8$. In this phase the solutions are grouped into clusters which are far away from each other. The results are in agreement with previous predictions of the cavity method and give a rigorous confirmation to one of its main building blocks. It can be generalized to other systems of both physical and computational interest.

Constraint satisfaction problems (CSPs) provide one of the main building blocks for complex systems studied in computer science, information theory, and statistical physics, and may even turn out to be important in the statistical studies of biological networks. Typically, they involve a large number of discrete variables, each one taking a finite number of values, and a set of constraints: each constraint involves a few variables, and forbids some of their joint assignments. A simple example is the $q$ coloring of a graph, where one should assign to each vertex of the graph a color in $\{1, \ldots, q\}$, in such a way that two vertices related by an edge have different colors. In the case $q = 2$, this is nothing but the zero temperature limit of an antiferromagnetic problem, which is known to display a spin glass behavior when the graph is frustrated and disordered. CSPs also appear naturally in the studies of structural glasses [1] and rigidity percolation [2].

Given an instance of a CSP, one wants to know whether there exists a solution, that is, an assignment of the variables which satisfies all the constraints (e.g., a proper coloring). When it exists the instance is called satisfiable, and one wants to find a solution. Most of the interesting CSPs are NP-complete problems: in the worst case the number of operations needed to decide whether an instance is SAT or not is expected to grow exponentially with the number of variables. But recent years have seen an upsurge of interest in the theory of typical case complexity, where one tries to identify random ensembles of CSPs which are hard to solve, and the reason for this difficulty. Random ensembles of CSPs are also of great theoretical and practical importance in communication theory: some of the best error correcting codes (the so-called low density parity check codes) are based on such constructions [3,4].

The archetypical example of CSP is the satisfiability problem (SAT). This is a core problem in computational complexity: it is the first one to have been shown to be an NP-complete problem [5], and since then thousands of problems have been shown to be computationally equivalent to it. Yet it is not so easy to find difficult instances. The main ensemble which has been used for this goal is the

random $K$-satisfiability ($K$-SAT) ensemble. The variables are $N$ binary variables—Ising spins—$\vec{\sigma} = \{\sigma_i\} \in \{-1, 1\}^N$. The constraints are called $K$-clauses. Each of them involves $K$ distinct spin variables, randomly chosen with uniform distribution, and it forbids one configuration of these spins, randomly chosen among the $2^K$ possible ones. A set of $M$ clauses defines the problem. This corresponds to generating a random logical formula in conjunctive normal form, which is a very generic problem appearing in logic. $K$-SAT can also be written as the problem of minimizing a spin glasslike energy function which counts the number of violated clauses and in this respect random $K$-SAT is seen as a prototypical diluted spin glass [6]. Here we shall keep to the most interesting case $K \geq 3$ (for $K = 2$ the problem is polynomial).

In the recent years random $K$-SAT has attracted much interest in computer science and in statistical physics [7–10]. The interesting limit is the thermodynamic limit $N \rightarrow \infty$, $M \rightarrow \infty$ at fixed clause density $\alpha = M/N$. Its most striking feature is certainly its sharp threshold. It is strongly believed that there exists a phase transition for this problem: Numerical and heuristic analytical arguments are in support of the so-called *satisfiability threshold conjecture: There exists $\alpha_c(K)$ such that, with high probability, if $\alpha < \alpha_c(K)$, a random instance is satisfiable; if $\alpha > \alpha_c(K)$, a random instance is unsatisfiable.* Throughout this Letter "with high probability" (w.h.p.) means with a probability going to one in the $N \rightarrow \infty$ limit. Although this conjecture remains unproven, Friedgut has come close to it by establishing the existence of a nonuniform sharp threshold [11]. A lot of effort has been devoted to understanding this phase transition. This is interesting not only from the physics point of view but also from the computer science one, because the random instances with $\alpha$ close to $\alpha_c$ are the hardest to solve. The most important rigorous results so far are bounds for the threshold $\alpha_c(K)$. The best upper bounds were derived using first moment methods [12,13]. Lower bounds can be found by analyzing some algorithms which find SAT assignments [14,15], but recently a new method, based on second moment methods, has found better and

197205-1

algorithm-independent lower bounds [16,17]. Using these bounds, it was shown that $\alpha_c(K)$ scales as $2^K \ln(2)$ when $K \to \infty$.

On the other hand, some claim that the cavity method, which is a powerful tool from the statistical physics of disordered systems [18], can be used to compute the exact value of the threshold [19–21], giving for instance $\alpha_c(3) \simeq 4.2667$. It is a nonrigorous method but the self-consistency of its results have been checked by a "stability analysis" [21–23], and it also led to the development of a new algorithmic strategy, "survey propagation," which can solve very large instances at clause densities which are very close to the threshold (e.g., $N = 10^6$ and $\alpha = 4.25$).

The main hypothesis on which the cavity analysis of random $K$-satisfiability relies is the existence, in a region of clause density $[\alpha_d, \alpha_c]$ close to the threshold, of an intermediate phase called the "hard-SAT" phase. In this phase the set $S$ of solutions (a subset of the vertices in the $N$-dimensional hypercube) is supposed to split into many disconnected *clusters* $S = S_1 \cup S_2 \cup \ldots$. If one considers two solutions $X, Y$ in the same cluster $S_j$, it is possible to walk from $X$ to $Y$ (staying in $S$) by flipping at each step a finite numbers of spins. If on the other hand $X$ and $Y$ are in different clusters, in order to walk from $X$ to $Y$ (staying in $S$), at least one step will involve an extensive number (i.e., $\propto N$) of spin flips. This clustered phase is held responsible for trapping many local search algorithms into nonoptimal metastable states [24]. This phenomenon is not exclusive to random $K$-SAT. It is also predicted to appear in many other hard satisfiability and optimization problems such as coloring [25,26] or the multi-index matching problem [27], and corresponds to a "one step replica symmetry breaking" phase in the language of statistical physics. It is also a crucial limiting feature for decoding algorithms in some error correcting codes [28]. So far, the only CSP for which the existence of the clustering phase has been established rigorously is the simple polynomial problem of random exclusive-OR–SAT (XOR-SAT) [29,30]. In other cases it is an hypothesis, the self-consistency of which is checked by the cavity method.

In this Letter we provide rigorous arguments which show the existence of the clustering phenomenon in random $K$-SAT, for large enough $K$, in some region of $\alpha$ included in the interval $[\alpha_d(K), \alpha_c(K)]$ predicted by the statistical physics analysis. Our result is not able to confirm all the details of this analysis but it provides strong evidence in favor of its validity.

Given an instance $F$ of random $K$-satisfiability, we define a SAT-$x$-pair as a pair of assignments $(\vec{\sigma}, \vec{\tau}) \in \{-1, 1\}^{2N}$, which both satisfy $F$, and which are at a Hamming distance $d_{\sigma\tau} \equiv \Sigma_{i=1}^{N}(1 - \sigma_i \tau_i)/2$ specified by $x$ as follows:

$$d_{\sigma\tau} \in [Nx - \epsilon(N), Nx + \epsilon(N)] \tag{1}$$

Here $x$ is the normalized distance between the two con-

figurations, which we keep fixed as $N$ and $d$ go to infinity. The resolution $\epsilon(N)$ must be such that $\lim_{N\to\infty} \epsilon(N)/N = 0$, but its precise form is unimportant for our large $N$ analysis. One can choose for instance $\epsilon(N) = \sqrt{N}$.

We call *x-satisfiable* a formula for which such a pair of solutions exists. Our study mimics the usual steps which are taken in rigorous studies of $K$-SAT, but taking pairs of assignments at a fixed distance instead of single assignments.

We first formulate the *x-satisfiability threshold conjecture: For all $K \geq 2$ and for all $x$, $0 < x < 1$, there exists an $\alpha_c(K, x)$ such that, w.h.p., if $\alpha < \alpha_c(K, x)$, a random $K$-CNF is $x$-satisfiable; if $\alpha > \alpha_c(K, x)$, a random $K$-CNF is $x$-unsatisfiable*, which generalizes the usual satisfiability threshold conjecture (obtained for $x = 0$). We shall find explicitly below two functions, $\alpha_{LB}(K, x)$ and $\alpha_{UB}(K, x)$, which give lower and upper bounds for $\alpha$ for $x$-satisfiability at a given value of $K$. Numerical computations of these bounds show that $\alpha(K, x)$ is *nonmonotonous* as a function of $x$ for $K \geq 8$, as illustrated in Fig. 1. This in turn shows that, for $K$ large enough and in some well chosen interval of $\alpha$ below the satisfiability threshold, SAT-$x$-pairs exist for $x$ close to 0 ($\vec{\sigma}$ and $\vec{\tau}$ in the same cluster) and $x$ close to 0.5 ($\vec{\sigma}$ and $\vec{\tau}$ in different clusters),
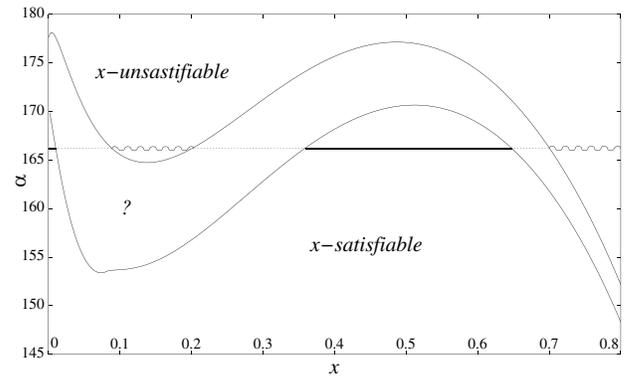


FIG. 1. Lower and upper Bounds for the $x$-satisfiability threshold $\alpha_c(K = 8, x)$. The upper curve is obtained by the first moment method. Above this curve there exists no SAT-$x$-pair, w.h.p. The lower curve is obtained by the second moment method. Below this curve there exists a SAT-$x$-pair, w.h.p. For values of $\alpha$ lying between 164.735 and 170.657, these bounds guarantee the existence of a clustering phenomenon. The horizontal line gives an example of this phenomenon for $\alpha = 166.1$. We exhibit the successive phases as one varies $x$: $x$-satisfiable regions are represented by a thick solid line, $x$-unsatisfiable regions by a wavy line, and "donot know" regions by a dotted line. The $x$-satisfiable region near $x = 0$ corresponds to intracluster pairs, whereas the $x$-satisfiable region around $x = 0.5$ corresponds to intercluster pairs. In this example, the intermediate $x$-unsatisfiable region around $x \sim 0.13$ shows the existence of a "gap" between clusters. We recall that the best refined lower and upper bounds for the satisfiability threshold $\alpha_c(K = 8)$ from [13,17] are, respectively, 173.253 and 176.596. The cavity prediction is 176.543 [21].

but there is an intermediate $x$ region where they do not exist. Figure 1 shows an explicit example of this scenario for a particular value of $\alpha$.

In what follows we first establish a rigorous and explicit upper bound using a simple first moment method. Subsequently, we provide a (numerical) lower bound using a second moment method [16,17]. Both results are based on elementary probabilistic techniques which could be generalized to other physical systems or random combinatorial problems.

*Upper bound: the first moment method.*—We use the fact that, when $Z$ is a non-negative random variable,

$$\mathbf{P}(Z \geq 1) \leq \mathbf{E}(Z). \tag{2}$$

Given a formula $F$, we take $Z(F)$ to be the number of pairs of solutions at fixed distance [with resolution $\epsilon(N)$]:

$$Z(F) = \sum_{\vec{\sigma}, \vec{\tau}} \delta\left(\frac{d_{\sigma\tau}}{N} \simeq x\right) \delta(\vec{\sigma}, \vec{\tau} \in S(F)), \tag{3}$$

where $S(F)$ is the set of solutions to $F$. Throughout this Letter $\delta(A)$ is an indicator function, equal to one if the statement $A$ is true, and to 0 otherwise. Since $Z(F) \geq 1$ is equivalent to "$F$ is $x$-satisfiable," (3) gives an upper bound for the probability of $x$-satisfiability. The expected value of the double sum over the choice of a random $F$ is

$$\mathbf{E}(Z(F)) = 2^N \binom{N}{Nx} \mathbf{E}[\delta(\vec{\sigma}, \vec{\tau} \in S(c))]^M. \tag{4}$$

We have used $\delta(\vec{\sigma}, \vec{\tau} \in S(F)) = \prod_c \delta(\vec{\sigma}, \vec{\tau} \in S(c))$, where $c$ denotes the clauses, and the fact that clauses are drawn independently. The expectation $\mathbf{E}[\delta(\vec{\sigma}, \vec{\tau} \in S(c))]$ is equal to $1 - 2^{1-K} + 2^{-K}(1-x)^K$ (there are only two realizations of the clause among $2^K$ that do not satisfy $c$ unless the two configurations overlap exactly on the domain of $c$).

In the thermodynamic limit, $\ln\mathbf{E}(Z(F))/N \to \Phi_1(x, \alpha)$, where

$$\Phi_1(x, \alpha) = \ln 2 + H_2(x) + \alpha \ln\{1 - 2^{-K}[2 - (1-x)^K]\},$$

where $H_2(x) = -x\ln x - (1-x)\ln(1-x)$ is the two-state entropy function. This gives the upper bound

$$\alpha_{\text{UB}}(K, x) = -\frac{\ln 2 + H_2(x)}{\ln[1 - 2^{1-K} + 2^{-K}(1-x)^K]}. \tag{5}$$

*Lower bound: the second moment method.*—We use the fact that, when $Z$ is a non-negative random variable,

$$\mathbf{P}(Z > 0) \geq \frac{\mathbf{E}(Z)^2}{\mathbf{E}(Z^2)}. \tag{6}$$

However, using this formula with $Z$ equal to the number of solutions fails, and one must instead use a weighted sum [16]. We follow the strategy recently developed in [17], which we generalize to SAT-$x$-pairs by taking

$$Z(F) = \sum_{\vec{\sigma}, \vec{\tau}} \delta\left(\frac{d_{\sigma\tau}}{N} \simeq x\right) \prod_c W(\vec{\sigma}, \vec{\tau}, c). \tag{7}$$

$W(\vec{\sigma}, \vec{\tau}, c)$ is a weight associated with the clause $c$, given the couple $(\vec{\sigma}, \vec{\tau})$, and is defined as follows: Suppose that $c$ is satisfied by $n_\sigma$ among the $K$ $\vec{\sigma}$ variables involved in $c$, and by $n_\tau$ among the $K$ $\vec{\tau}$ variables. Call $n_0$ the number of common values between the $\vec{\sigma}$ and $\vec{\tau}$ variables involved in $c$. Then define

$$W(\vec{\sigma}, \vec{\tau}, c) = \begin{cases} \lambda^{n_\sigma + n_\tau} \nu^{n_0} & \text{if } n_\sigma > 0 \quad \text{and} \quad n_\tau > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Note that with this definition of $Z$ the choice $\lambda = 1$, $\nu = 1$ simply yields the number of solutions (3).

Let us now compute the first two moments of $Z$ ([31]):

$$\mathbf{E}(Z) = 2^N \binom{N}{Nx} \left[f_1^{(\lambda, \nu)}(x)\right]^M, \tag{9}$$

where $f_1^{(\lambda, \nu)}(x) = \mathbf{E}(W(\vec{\sigma}, \vec{\tau}, c))$ can be calculated by simple combinatorics (via multinomial sums). To compute $\mathbf{E}(Z^2)$, we sum over four spin configurations $\vec{\sigma}, \vec{\tau}, \vec{\sigma}', \vec{\tau}'$. Symmetry allows us to fix $\sigma_i = 1$. Let $Na(t, s, t')$ be the number of sites $i$ such that $\tau_i = t$, $\sigma_i' = s'$, and $\tau_i' = t'$ (where $t, s, t' \in \{\pm 1\}$). It turns out that the term of the sum depends only on these eight numbers $a(\pm 1, \pm 1, \pm 1)$. We collect them into a vector $\mathbf{a}$ and get

$$\mathbf{E}(Z^2) = 2^N \int_V d\mathbf{a} \frac{N!}{\prod_{t,s',t'} (Na(t, s', t'))!} [f_2^{(\lambda, \nu)}(\mathbf{a})]^M, \tag{10}$$

where $f_2^{(\lambda, \nu)}(\mathbf{a}) = \mathbf{E}(W(\vec{\sigma}, \vec{\tau}, c)W(\vec{\sigma}', \vec{\tau}', c))$ can be calculated by simple combinatorics in the same way as $f_1$. The integration set $V$ is a 5-dimensional simplex taking into account the normalization $\Sigma_{t,s',t'} a(t, s', t') = 1$ and the two constraints: $d_{\sigma\tau}/N \simeq x$, $d_{\sigma'\tau'}/N \simeq x$.

A saddle point evaluation of Eq. (10) gives, for $N \to \infty$,

$$\frac{\mathbf{E}(Z)^2}{\mathbf{E}(Z^2)} \geq C_0 \exp\left(-N\max_{\mathbf{a} \in V} \Phi_2(\mathbf{a})\right), \tag{11}$$

where $C_0$ is a constant depending on $K$ and $x$, and

$$\Phi(\mathbf{a}) = H_8(\mathbf{a}) - \ln 2 - 2H_2(x) + \alpha \ln f_2^{(\lambda, \nu)}(\mathbf{a})$$
$$- 2\alpha \ln f_1^{(\lambda, \nu)}(x), \tag{12}$$

with $H_8(\mathbf{a}) = -\Sigma_{t,s',t'} a(t, s', t') \ln a(t, s', t')$. In general $\max_{\mathbf{a} \in V} \Phi(\mathbf{a})$ is non-negative and one must choose appropriate weights $W(\vec{\sigma}, \vec{\tau}, c)$ in such a way that $\max_{\mathbf{a} \in V} \Phi(\mathbf{a}) = 0$. We notice that at the particular point $\mathbf{a}^*$ where $(\vec{\sigma}, \vec{\tau})$ is uncorrelated with $(\vec{\sigma}', \vec{\tau}')$, we have $\Phi(\mathbf{a}^*) = 0$. We fix the parameters $\lambda$ and $\mu$ defining the weights (8) in such a way that $\mathbf{a}^*$ is a local maximum of $\Phi$. This gives two algebraic equations in $\lambda$ and $\nu$ which have a unique solution $\lambda > 0$, $\nu > 0$. Fixing $\lambda$ and $\nu$ to these

values, $\alpha_{LB}$ is the largest value of $\alpha$ such that the local maximum at $\mathbf{a}^*$ is a *global* maximum, i.e., such that there exists no $\mathbf{a} \in V$ with $\Phi(\mathbf{a}) > 0$:

$$\alpha_{LB}(K, x) = \inf_{\mathbf{a} \in V} \frac{\ln 2 + 2H_2(x) - H_8(\mathbf{a})}{\ln f_2^{(\lambda, \nu)}(\mathbf{a}) - 2\ln f_1^{(\lambda, \nu)}(x)}. \quad (13)$$

We devised several numerical strategies to evaluate $\alpha_{LB}(K, x)$. The implementation of Powell's method starting from each point of a grid of size $\mathcal{N}^5$ ($\mathcal{N} = 10, 15, 20$) on $V$ turned out to be the most efficient and reliable. The results are given by Fig. 1 for $K = 8$, the smallest $K$ such that the clustering conjecture is confirmed. We found a clustering phenomenon for all the values of $K \geq 8$ that we checked, and in fact the relative difference $[\alpha_{UB}(K, x) - \alpha_{LB}(K, x)]/\alpha_{LB}(K, x)$ seems to go to zero at large $K$.

We have shown a simple probabilistic argument which shows rigorously the existence of a clustered hard-SAT phase. The prediction from the cavity method is in fact a weaker statement. It can be stated in terms of the overlap distribution function $P(x)$, which is the probability, when two SAT assignments are taken randomly (with uniform distribution), that their distance is given by $x$. The cavity method finds that this distribution has a support concentrated on two values: a small value $x_1$, close to zero, gives the characteristic "radius" of a cluster; a larger value $x_0$ gives the characteristic distance between clusters. This does not imply that there exists no pair of solutions for values of $x$ distinct from $x_0, x_1$; it just means that such pairs are exponentially less numerous than the typical ones. Our rigorous result shows that in fact there exists a true gap in $x$, with no SAT-$x$-pairs, at least for $K \geq 8$. More sophisticated moment computations might allow to get some results for smaller values of $K$. Still the conceptual simplicity of our computation makes it a useful tool for proving similar phenomena in other systems of physical or computational interests, like for instance the graph-coloring (antiferromagnetic Potts) problem.

[1] M. Sellitto, G. Biroli, and C. Toninelli, Europhys. Lett. **69**, 496 (2005).

[2] J. Barré *et al.*, cond-mat/0408385 [Phys. Rev. Lett. (to be published)].

[3] Robert G. Gallagher, *Information Theory and Reliable Communication* (Wiley, New York, 1968).

[4] David J.C. MacKay, *Information Theory, Inference & Learning Algorithms* (Cambridge University Press, Cambridge, 2002).

[5] Stephen Cook, in *Proceedings of the Third Annual ACM Symposium on Theory of Computing, Shaker Heights, Ohio, United States*, (ACM Press, New York, 1971), p. 151.

[6] R. Monasson and R. Zecchina, Phys. Rev. E **56**, 1357 (1997).

[7] Special issue on Frontiers in Problem Solving: Phase Transitions and Complexity, edited by T. Hogg, B.A. Huberman, and C. Williams [Artif. Intell. **81**, 1 (1996).

[8] Special Issue on NP-hardness and Phase transitionls, edited by O. Dubois, R. Monasson, B. Selman, and R. Zecchina [Theor. Comput. Sci. **265** 1 (2001)].

[9] S. Kirkpatrick and B. Selman, Science **264**, 1297 (1994).

[10] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyanski, Nature (London) **400**, 133 (1999).

[11] E. Friedgut, J. Am. Math. Soc. **12**, 1017 (1999).

[12] L.M. Kirousis, E. Kranakis, and D. Krizanc, School of Computer Science-Carleton University Technical Report No. TR-96-09, 1996 (unpublished).

[13] O. Dubois and Y. Boufkhad, Journal of Algorithms **24**, 395 (1997).

[14] M.-T. Chao and J. Franco, Information Sciences (NY) **51**, 289 (1990).

[15] A.M. Frieze and S. Suen, Journal of Algorithms **20**, 312 (1996).

[16] D. Achlioptas and C. Moore, *Proc. Foundations of Computer Science* (2002).

[17] D. Achlioptas, Y. Peres, J. Am. Math. Soc. **17**, 947 (2004).

[18] M. Mézard and G. Parisi, J. Stat. Phys. **111**, 1 (2003).

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

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

[21] S. Mertens, M. Mézard, and R. Zecchina, "Threshold values of Random $K$-SAT from the cavity method" (to be published).

[22] A. Montanari and F. Ricci-Tersenghi, Eur. Phys. J. B **B33**, 339 (2003).

[23] A. Montanari, G. Parisi, and F. Ricci-Tersenghi, J. Phys. A **37**, 2073 (2004).

[24] G. Semerjian and R. Monasson, *Proceedings of the SAT 2003 Conference*, edited by E. Giunchiglia and A. Tacchella, Lect. Notes Comput. Sci. Vol. 2919 (Springer), New York, New York, 2004), p. 120.

[25] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. Lett. **89**, 268701 (2002).

[26] A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. E **68**, 036702 (2003).

[27] O.C. Martin, M. Mézard, and O. Rivoire, Phys. Rev. Lett. **93**, 217205 (2004).

[28] A. Montanari, Eur. Phys. J. B **B23**, 121 (2001).

[29] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina, J. Stat. Phys. **111**, 505 (2003).

[30] S. Cocco, O. Dubois, J. Mandler, and R. Monasson, Phys. Rev. Lett. **90**, 047205 (2003).

[31] M. Mézard, T. Mora, and R. Zecchina (to be published).