PHYSICAL REVIEW LETTERS

VOLUME 76 | 20 MAY 1996 | NUMBER 21

Entropy of the K-Satisfiability Problem

Rémi Monasson[1,*] and Riccardo Zecchina[2,†]

[1]*Laboratoire de Physique Théorique de l'ENS, 24 rue Lhomond, 75231 Paris cedex 05, France*
[2]*Istituto Nazionale di Fisica Nucleare and Dip. di Fisica, Politecnico di Torino, C.so Duca degli Abruzzi 24, I-10129 Torino, Italy*
(Received 12 January 1996)

The threshold behavior of the K-satisfiability problem is studied in the framework of the statistical mechanics of random diluted systems. We find that at the transition the entropy is finite and hence that the transition itself is due to the abrupt appearance of logical contradictions in all solutions and not to the progressive decreasing of the number of these solutions down to zero. A physical interpretation is given for the different cases $K = 1$, $K = 2$, and $K \geq 3$. [S0031-9007(96)00244-X]

PACS numbers: 05.20.–y, 02.10.–v, 64.60.H+, 89.70.+c

Many famous computational issues concerning optimization problems, discrete structures (such as graphs and networks), and formal logic are classified by complexity theory, according to the running time scaling of their algorithms and memory requirements [1]. In the past few years, it has been realized that the statistical physics of frustrated models could be helpful to acquire a better understanding of complexity, by mapping the optimization problems onto the study of the ground states of disordered models [2]. The tools and concepts of statistical mechanics have therefore opened the way to the analysis of the typical properties of optimization problems as well as to the definition and to successive improvements of search algorithms in which temperature is the main control parameter [3].

More recently, the observation of threshold phenomena in random mathematical and computer science problems, and mainly in one of the most basic [1] of them, the *satisfiability* (SAT) problem [1,4–7], has shown that not only ground state properties but also the critical behavior at glassy phase transitions occurring in disordered systems could be relevant for complexity theory, due to the so-called intractability concentration phenomena [7]. The purpose of this Letter is to provide a stronger support to this statement, by giving an analytical study of the properties of SAT near its transition. In turn, the results will be shown to be of interest for statistical mechanics, due to the very peculiar nature of the glassy transition taking place therein.

The model we shall study is a version of SAT, called K-SAT and defined as follows. Let us consider N Boolean variables $\{x_i = 0, 1\}_{i=1,\ldots,N}$. We first randomly choose K among the N possible indices i and then, for each of them, a literal that is the corresponding x_i or its negation $\bar{x}_i$ with equal probabilities one half. A clause C is the logical OR of the K previously chosen literals; that is, C will be true (or satisfied) if and only if at least one literal is true. Next, we repeat this process to obtain M independently chosen clauses $\{C_\ell\}_{\ell=1,\ldots,M}$ and ask for all of them to be true at the same time (i.e., we take the logical AND of the M clauses). A logical assignment of the $\{x_i\}$'s satisfying all clauses, if any, is called a solution of the K-satisfiability problem.

When the number of clauses becomes of the same order as the number of variables ($M = \alpha N$) and in the large N limit (indeed the case of interest also in the fields of computer science [5,6]), K-SAT exhibits striking threshold phenomena. Numerical simulations show that the probability of finding a solution falls abruptly from one down to zero when α crosses a critical value $\alpha_c(K)$ of the number of clauses per variable. Above $\alpha_c(K)$, all clauses cannot be satisfied any longer and one would rather minimize the number of unsatisfiable clauses, which is the optimization version of K-SAT also referred to as the MAX-K-SAT.

This scenario has been proven to be true in the $K = 2$ case. It has been possible to derive rigorously the threshold value $\alpha_c = 1$ [8], and an explicit 2-SAT algorithm working for $\alpha < \alpha_c$ has been developed, whose running time scales polynomially with N [9]. For $K \geq 3$, much less is known and K-SAT belongs to the class of hard computational problems, the so-called NP-complete class [1], roughly meaning that running times of search algorithms are thought to scale exponentially in N (notice that for $\alpha > 1$ also MAX-2-SAT is NP complete [1]). Some bounds on $\alpha_c(K)$ have been derived [6] and a remarkable application of finite size scaling techniques has recently allowed to find precise numerical values of α_c for $K = 3, 4, 5, 6$ [4]. The situation becomes easier to understand in the large K limit where a simple probabilistic argument gives the asymptotic expression of $\alpha_c(K) \simeq 2^K \ln 2$ [4]. An important rigorous result is the self-averageness taking place in MAX-K-SAT for any K: independently of the particular sample of M clauses, the minimal fraction of violated clauses is narrowly peaked around its mean value when $N \to \infty$ at fixed α [10].

In order to study the K-SAT (and MAX-K-SAT) problem, we map it onto random diluted systems by the introduction of spin variables $S_i = \pm 1$ (a simple shift of the Boolean variables) and a quenched (unbiased) matrix $C_{\ell,i} = 1$ (-1) if x_i $(\bar{x}_i)$ belongs to the clause C_ℓ, 0 otherwise. Then the energy-cost function

$$E[C,S] = \sum_{\ell=1}^{M} \delta\left[\sum_{i=1}^{N} C_{\ell,i}\, S_i; -K\right], \tag{1}$$

where $\delta[i;j]$ denotes the Kronecker symbol, equals the number of violated clauses and therefore its ground state (GS) properties describe the transition from K-SAT ($E_{\rm GS} = 0$) to MAX-K-SAT ($E_{\rm GS} > 0$) (a similar cost function was introduced in [5]).

While previous works on the statistical mechanics of other combinatorial optimization problems—such as traveling salesman, graph partitioning, or matching problems [2,11]—focused mainly on the study of the typical cost of optimal configurations (with no phase transitions in the ground state), the issues arising in K-SAT are of different nature. Below α_c, the ground state energy vanishes and the key quantity to be analyzed is the typical number of existing solutions, i.e., the ground state entropy $S_{\rm GS}$, for which no exact results are available so far. Our main result is that $S_{\rm GS}$ is still extensive at $\alpha = \alpha_c$: The transition is not due to a progressive reduction of the number of solutions but to *the sudden appearance of logical contradictions* in "all" of the exponentially numerous solutions at the threshold.

In order to regularize the model, we compute the partition function

$$Z[C] = \sum_{\{S_i = \pm 1\}} \exp(-\beta E[C,S]), \tag{2}$$

after having introduced a finite "temperature" $1/\beta$. The typical ground state free energy $\overline{\ln Z[C]} = \lim_{n\to 0} \times (\overline{Z[C]^n} - 1)/n$, where $\overline{(\cdots)}$ stands for the average over the random clauses, is then recovered in the limit of zero temperature $\beta \to \infty$. For brevity, we limit ourselves to a general description of the method [2,11] and focus on the discussion of the results (the complete calculation will be presented in [12]).

Once we have introduced n replicas S_i^a, $a = 1, \ldots, n$, of the system, the average over the disorder C couples all replicas together through the overlaps $Q^{a_1,\dot{s},a_{2r}} = (1/N)\sum_{i=1}^{N} S_i^{a_1} \cdots S_i^{a_{2r}}$ and their conjugated Lagrange parameters $\hat{Q}^{a_1,\ldots,a_{2r}}$ $(r = 1, \ldots, n/2)$ [11]. The resulting effective Hamiltonian $N\,\mathcal{H}[\{Q\},\{\hat{Q}\}]$ involves all multireplicas overlaps as in diluted spin glasses [11,13] and is therefore much more complicated than long-range disordered models where only interactions between pairs of replicas appear [2]. The free energy is evaluated by taking the saddle point of $\mathcal{H}$ over all overlaps $Q, \hat{Q}$. This highly difficult task may be simplified by noticing that, due to the indistinguishability of the n replicas, the effective Hamiltonian $\mathcal{H}$ must be invariant under any permutation of the replicas. Therefore, one is allowed to look for a solution such that the overlaps only depend upon the number of coupled replicas: $Q^{a_1,\ldots,a_{2r}} = Q_r$, $\hat{Q}^{a_1,\ldots,a_{2r}} = \hat{Q}_r$ [2,11]. This is the so-called replica symmetric (RS) ansatz we shall use hereafter. Moreover, it results to be convenient to characterize all Q_r by introducing a probability distribution $P(x)$ of the Boolean magnetization $x = \langle S \rangle$, such that $Q_r = \int_{-1}^{1} dx\, P(x) x^{2r}$. Elimination of the Lagrange parameters $\hat{Q}_r$'s leads to the expression

$$\begin{aligned}\frac{1}{N}\overline{\ln Z[C]} = {} & \ln 2 - \frac{1}{2}\int_{-1}^{1} dx\, P(x) \ln(1 - x^2) \\ & + \alpha(1 - K)\int_{-1}^{1} \prod_{\ell=1}^{K} dx_\ell\, P(x_\ell) \ln A_{(K)} \\ & + \frac{\alpha K}{2}\int_{-1}^{1} \prod_{\ell=1}^{K-1} dx_\ell\, P(x_\ell) \ln A_{(K-1)}\,,\end{aligned} \tag{3}$$

with $A_{(J)} \equiv A_{(J)}[\{x_\ell\},\beta] = 1 + (e^{-\beta} - 1)\prod_{\ell=1}^{J}(1 + x_\ell)/2$ for $J = K - 1$ and $J = K$. The measure $P(x)$ is given by the saddle-point integral equation

$$\begin{aligned}P(x) = {} & \frac{1}{1 - x^2}\int_{-\infty}^{\infty} du \cos\left[\frac{u}{2}\ln\left(\frac{1 + x}{1 - x}\right)\right] \\ & \times \exp\left[-\alpha K + \alpha K \int_{-1}^{1} \prod_{\ell=1}^{K-1} dx_\ell\, P(x_\ell) \right. \\ & \left. \times \cos\left(\frac{u}{2}\ln A_{(K-1)}\right)\right].\end{aligned} \tag{4}$$

A toy version of the K-SAT problem is obtained when $K = 1$. The typical free energy may then be computed either by a simple combinatorial analysis as well as within our approach, showing that the RS ansatz is exact for all β and α when $K = 1$. The saddle-point equation for $P(x)$ can be explicitly solved at any temperature $1/\beta$ and the solution read

$$P(x) = \sum_{\ell=-\infty}^{\infty} e^{-\alpha} I_\ell(\alpha)\, \delta\left(x - \tanh\left(\frac{\beta\ell}{2}\right)\right). \quad (5)$$

One finds that in the limit of zero temperature the energy and the entropy of the ground state read $E_{\rm GS}(\alpha) = \alpha[1 - e^{-\alpha}I_0(\alpha) - e^{-\alpha}I_1(\alpha)]/2$ and $S_{\rm GS}(\alpha) = e^{-\alpha}I_0(\alpha)\ln 2$, respectively, where I_ℓ is the ℓth modified Bessel function. Therefore, as soon as $\alpha > \alpha_c(1) = 0$, the clauses cannot be satisfied all together but there is an exponentially large number $\exp[NS_{\rm GS}(\alpha)]$ of different spins configurations giving the same minimum fraction $E_{\rm GS}(\alpha)/\alpha$ of unsatisfiable clauses. At zero temperature, $P(x)$ reduces to a sum of three Dirac peaks in $x = \pm 1$ and 0 with weights $[1 - e^{-\alpha}I_0(\alpha)]/2$ and $e^{-\alpha}I_0(\alpha)$, respectively. It clearly appears that the finite value of the ground state entropy is due to the presence of unfrozen spins [12]. Another relevant mechanism is the accumulation of magnetizations $\overline{\langle S\rangle} = \pm[1 - O(e^{-|z|\beta})]$, $z = O(1)$, around the limit values $x = \pm 1$, as can be seen from (5). The occurrence of such peaks means that a finite fraction of spins are frozen, and hence that a new clause presented to the systems would not be satisfiable with a finite probability.

This scenario remains valid for any K. The fraction of violated clauses at temperature $1/\beta$ can indeed be computed through $E = -(1/N)\partial\overline{\ln Z}/\partial\beta$. The ground state energy will clearly depend only upon the magnetizations of order $\pm[1 - O(e^{-|z|\beta})]$, if any. These contributions can be picked up by the new function $R(z) = \lim_{\beta\to\infty}\beta P[\tanh(\beta z/2)]/2\cosh^2(\beta z/2)$ satisfying the saddle-point equation

$$R(z) = \int_{-\infty}^{\infty}\frac{du}{2\pi}\cos(uz)\exp\left[-\frac{\alpha K}{2^{K-1}} + \alpha K \times \int_0^\infty \prod_{\ell=1}^{K-1} dz_\ell\, R(z_\ell)\cos[u\min(1,z_1,\ldots,z_{K-1})]\right]. \quad (6)$$

Remarkably, it is possible to find analytically an exact solution to this functional equation for any K and α:

$$R(z) = \sum_{\ell=-\infty}^{\infty} e^{-\gamma}I_\ell(\gamma)\delta(z - \ell), \quad (7)$$

where γ is solution of the implicit equation

$$\gamma = \alpha K\left[\frac{1 - e^{-\gamma}I_0(\gamma)}{2}\right]^{K-1}. \quad (8)$$

The corresponding cost function equals $E_{\rm GS}(\alpha) = \gamma[1 - e^{-\gamma}I_0(\gamma) - Ke^{-\gamma}I_1(\gamma)]/2K$. We shall now analyze the physical structure of this solution and show how the predictions it leads to for $K = 2$ qualitatively differ with respect to the case $K \geq 3$.

The self-consistency equation (8) admits the solution $\gamma = 0$ for any α (if $K > 1$). When $K = 2$, there is another solution $\gamma(\alpha) > 0$ above $\alpha = 1$. This new solution maximizes $E_{\rm GS}$ (and then the free energy) and must be preferred [2]. Therefore, our RS theory predicts that $E_{\rm GS} = 0$ for $\alpha \leq 1$ and increases continuously when $\alpha > 1$, giving back the rigorous result $\alpha_c(2) = 1$. The transition taking place at α_c is of second order with respect to the order parameter: the value of γ does not show any jump, and two Dirac peaks for $P(x)$ progressively appears in $x = \pm 1$ with amplitude $[1 - e^{-\gamma}I_0(\gamma)]/2$ each. For large α, the RS ground state energy scales as $E_{\rm GS} \simeq \alpha/4$ which is known to be exact [5]. As far as 2-SAT is concerned, the value of the threshold is correctly predicted ($\alpha_c = 1$) and the RS solution appears to be correct for any α (even $\alpha > 1$) [12].

For $\alpha > \alpha_c$, there do not exist any more sets of S_i's such that the energy (1) remains nonextensive. The vanishing of the exponentially large number of solutions that were present below the threshold is surprisingly abrupt. We have indeed studied the number of such solutions as a function of the number of clauses per spin in the range $0 \leq \alpha \leq \alpha_c$. Their logarithm (divided by N), that is the entropy of the ground state $S_{\rm GS}(\alpha)$, is given by (3) when $\beta \to \infty$. We have resorted to an exact expansion of $P(x)$ (4) in powers of α, starting from $P(x)|_{\alpha=0} = \delta(x)$, and injected the resulting probability function into (3) to obtain the expansion of $S_{\rm GS}(\alpha)$. At the 7th order (which implies an uncertainty less than 1%), we have found that $S_{\rm GS}(\alpha_c) \simeq 0.38$, which is still very high as compared to $S_{\rm GS}(0) = \ln 2$ (see Fig. 1). The transition is therefore due to the abrupt appearance of contradictory logical loops in "all" solutions at $\alpha = \alpha_c$ and not to the progressive decreasing of the number of these solutions down to zero at the threshold.

Let us turn now to the $K \geq 3$ case. Resolution of the implicit equation (8) leads to the following picture. For $\alpha < \alpha_m(K)$, there exists the solution $\gamma = 0$ only. At $\alpha_m(K)$, a nonzero solution $\gamma(\alpha)$ discontinuously appears. The corresponding ground state energy is negative in the range $\alpha_m(K) \leq \alpha < \alpha_s(K)$, meaning that the new solution is metastable and that $E_{\rm GS} = 0$ up to $\alpha_s(K)$. For $\alpha > \alpha_s(K)$ the $\gamma(\alpha) \neq 0$ solution becomes thermodynamically stable, leading to the conclusion that $\alpha_s(K)$ corresponds to the desired threshold $\alpha_c(K)$. However, this prediction is wrong as can be immediately seen for $K = 3$, since the experimental value $\alpha_c(3) = 4.17 \pm 0.05$ [4] is lower than $\alpha_m(3) \simeq 4.667$ and $\alpha_s(3) \simeq 5.181$. In addition, large K evaluations give $\alpha_m(K) \sim K2^K/16\pi$ and $\alpha_s(K) \sim K2^K/4\pi$, which grow faster than the asymptotic value $\alpha_c(K) \sim 2^K \ln 2$ [14].

This situation may be understood by an inspection of the RS ground state entropy. To do so, we have expanded $S_{\rm GS}$ to the ℓth order in α using the same method as for $K = 2$ and denoted by $\alpha_{ze}^{(\ell)}(K)$, the point where it

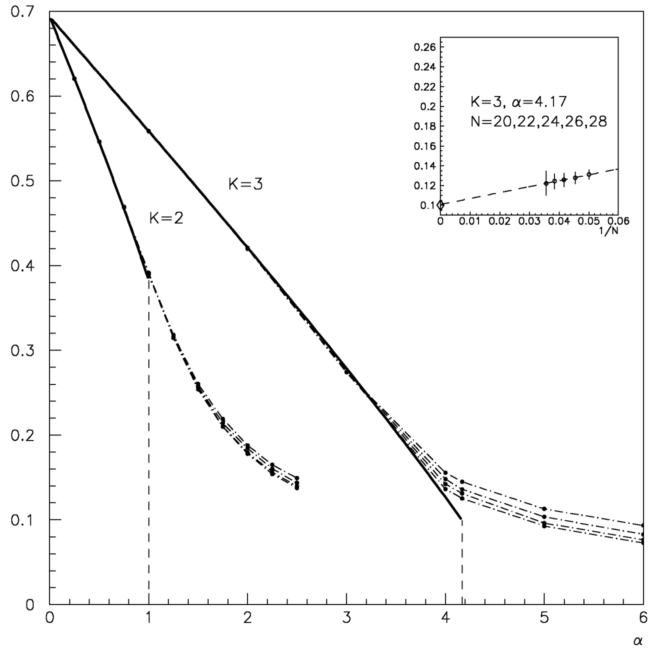

FIG. 1. Entropy vs α for $K = 2$ and 3. The analytical solutions (solid lines) are compared with numerical exhaustive simulations for $N = 12, 16, 20, 24$ and 30 000, 15 000, 7500, 2500 samples, respectively (for $K = 2$ we stop at $\alpha = 2.5$, whereas for $K = 3$ at $\alpha = 6$). Error bars are within 10% and thus not reported explicitly. Inset: $1/N$ entropy extrapolation for $\alpha = 4.17$ (in average for each N), $N = 20, 22, 24, 26, 28$ (with 16 500, 11 500, 7500, 4000, 3000 samples, respectively) and $K = 3$.

vanishes. Note that $\alpha_{ze}^{(1)}(K)$ corresponds to the annealed theory while $\alpha_{ze}^{(\ell)}(K)$ converges to $\alpha_{ze}(K)$ when $\ell \to \infty$, that is, the exact value of α at which S_{GS} goes to zero. For $K = 3$, we have performed the expansion up to $\ell = 8$ and found $\alpha_{ze}^{(1)} = 5.1909$, $\alpha_{ze}^{(2)} = 5.0144$, $\alpha_{ze}^{(3)} = 4.9189$, $\alpha_{ze}^{(4)} = 4.8589$, $\alpha_{ze}^{(5)} = 4.8187$, $\alpha_{ze}^{(6)} = 4.7893$, $\alpha_{ze}^{(7)} = 4.7677$, $\alpha_{ze}^{(8)} = 4.7504$, indicating that α_{ze} is definitively larger that $\alpha_c \simeq 4.17$. Repeating the calculation for $K = 4, 5, 6$, we have obtained qualitatively similar results which show an even quicker convergence towards a zero entropy point such that $\alpha_c(K) < \alpha_{ze}(K) < \alpha_s(K)$. Finally, in the large K limit, $\alpha_{ze}(K)$ asymptotically reaches the threshold $\alpha_c(K)$ from above. As K grows, fluctuations get weaker and weaker and Gaussian RS theory becomes exact. Solving Eq. (4), we find for $K \gg 1$ and $\alpha \le \alpha_c(K)$

$$P(x) \simeq \frac{1}{\sqrt{2\pi u(\alpha)}\,(1 - x^2)} \times \exp\left[-\frac{1}{8u(\alpha)} \ln^2\left(\frac{1 + x}{1 - x}\right)\right], \qquad (9)$$

where $u(\alpha) = \alpha K/4^{K-1}$. As a consequence, $P(x) \to \delta(x)$ when $K \to \infty$ tells us that the annealed theory becomes exact in the large K limit.

Therefore, the situation is as follows for (finite) $K \ge 3$. Above α_{ze}, the RS entropy is negative, whereas it has to be the logarithm of an integer number. The RS ansatz is clearly unphysical in this range, explaining why α_s and α_c do not coincide. At the threshold α_c (which is experimentally known to be lower than α_{ze}), the RS entropy is still extensive. The crucial question now arises whether this result is exact or is affected by replica symmetry breaking (RSB) effects. To clear up this dilemma, an analysis of RSB effects would be required. Because of the general complexity of such an approach in diluted models [13] and the technical difficulty of the K-SAT problem, the preliminary attempts we have done in this direction have not been successful yet [12]. We have then resorted to exhaustive numerical simulation in the range $N = 12, \ldots, 28$ and compared the corresponding ground state entropies $S_{\mathrm{GS}}^{(N)}(\alpha)$ to our RS theory for $K = 3$. As reported in Fig. 1, for $\alpha < \alpha_c$, our analytical solution agrees very well with the numerical results. This confirms that the entropy of the ground state is finite at the threshold. The comparison may be made more precise by a careful extrapolation of the entropy $S_{\mathrm{GS}}^{(N)}(\alpha \simeq 4.17)$ in $1/N$ (see the inset of Fig. 1). The extrapolated value appears to be in perfect agreement with the RS prediction $S_{\mathrm{GS}} \simeq 0.1$. Therefore, RSB corrections to the RS theory seem to be absent below α_c, which leads us to think that RSB could occur at α_c exactly. For $\alpha > \alpha_c$, the numerical results of Fig. 1 correspond to the MAX-K-SAT typical entropy and could be recovered analytically by extending our solutions beyond the critical points (within the RS framework for $K = 2$ and with one step of RSB for $K = 3$).

To conclude, let us say that one should, however, not deduce from the above remark that the structure of the solution space is simple. It might well happen that the solution space could have a nontrivial structure which is not reflected by the magnetization distribution $P(x)$ only [15]. It would be very interesting to understand if such a phenomenon could take place in the K-SAT problem and what information the hidden structure of the solution space could give us about its algorithmic complexity near phase transition.

LPTENS is a unité propre du CNRS, associée à l'Ecole Normale Supérieure et à l'Université de Paris-Sud.

*Electronic address: monasson@physique.ens.fr
†Electronic address: zecchina@to.infn.it

[1] M. Garey and D.S. Johnson, *Computers and Intractability. A Guide to the Theory of NP-Completeness* (W.H. Freeman and Co., San Francisco, 1979).
[2] M. Mézard, G. Parisi, and M.A. Virasoro *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[3] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, Science **220**, 339 (1983).
[4] S. Kirkpatrick and B. Selman, Science **264**, 1297 (1994).

[5] S. Kirkpatrick, G. Gyorgyi, N. Tishby and L. Troyansky, in *Advances in Neural Information Processing Systems,* edited by J. Cowan, G. Tesauro, and J. Alspector (Morgan Kaufmann, San Mateo, CA, 1994), Vol. 6.

[6] A. Kamath, R. Motwani, K. Palem, and P. Spirakis, Random Structures and Algorithms **7**, 59 (1995); O. Dubois, Theor. Comput. Sci. **81**, 49 (1991).

[7] Issue 1-2, Artificial Intelligence **81**, edited by T. Hogg, B. A. Huberman and C. Williams (1996).

[8] A. Goerdt, in *Proceedings of the 17th International Symposium on Mathematical Foundations of Computer Science,* Prague, Czechoslovakia (1992), p. 264; V. Chvàtal and B. Reed, in *Proceedings of the 33rd IEEE Symposium on Foundations of Computer Science* (IEEE Society Press, Los Alamitos, CA, 1992), p. 620.

[9] B. Aspvall, M. F. Plass, and R. E. Tarjan, Inf. Process. Lett. **8**, 121 (1979).

[10] A. Z. Broder, A. M. Frieze, and E. Upfal, in *Proceedings of the 4th Annual ACM-SIAM Symposium on Discrete Algorithms* (Association for Computing Machinery, New York, 1993), p. 322.

[11] M. Mézard and G. Parisi, J. Phys. Lett. **46**, L771 (1985); J. Phys. (Paris) **47**, 1285 (1986).

[12] R. Monasson and R. Zecchina (to be published).

[13] C. De Dominicis and P. Mottishaw, J. Phys. A **20**, L1267 (1987); K. Y. M. Wong and D. Sherrington, J. Phys. A **21**, L459 (1988).

[14] Though the scaling of $\alpha_c(K)$ for large K is wrong within the RS ansatz, the asymptotic value for large α (and any K) of the ground state energy for MAX-K-SAT is correctly predicted : $E_{GS}(\alpha) \sim \alpha/2^K$ [10].

[15] G. Parisi and M. Virasoro, J. Phys. (Paris) **50**, 3317 (1986); T. R. Kirkpatrick and D. Thirumalai, Phys. Rev. B **36**, 5388 (1987).