Eur. Phys. J. B **14**, 551–568 (2000)

THE EUROPEAN PHYSICAL JOURNAL B

EDP Sciences

A variational description of the ground state structure in random satisfiability problems

G. Biroli[a], R. Monasson, and M. Weigt[b]

Laboratoire de Physique Théorique de l'ENS[c], 24 rue Lhomond, 75231 Paris Cedex 05, France

Received 12 July 1999

Abstract. A variational approach to finite connectivity spin-glass-like models is developed and applied to describe the structure of optimal solutions in random satisfiability problems. Our variational scheme accurately reproduces the known replica symmetric results and also allows for the inclusion of replica symmetry breaking effects. For the 3-SAT problem, we find two transitions as the ratio α of logical clauses per Boolean variables increases. At the first one $\alpha_{\rm s} \simeq 3.96$, a non-trivial organization of the solution space in geometrically separated clusters emerges. The multiplicity of these clusters as well as the typical distances between different solutions are calculated. At the second threshold $\alpha_{\rm c} \simeq 4.48$, satisfying assignments disappear and a finite fraction $B_0 \simeq 0.13$ of variables are overconstrained and take the same values in all optimal (though unsatisfying) assignments. These values have to be compared to $\alpha_{\rm c} \simeq 4.27$, $B_0 \simeq 0.4$ obtained from numerical experiments on small instances. Within the present variational approach, the SAT-UNSAT transition naturally appears as a mixture of a first and a second order transition. For the mixed $2+p$-SAT with $p < 2/5$, the behavior is as expected much simpler: a unique smooth transition from SAT to UNSAT takes place at $\alpha_{\rm c} = 1/(1-p)$.

PACS. 05.20.-y Classical statistical mechanics – 64.60.-i General studies of phase transitions – 89.90.+n Other topics of general interest to physicists

1 Introduction

Over the last few years, the computer science community has become increasingly aware of the occurrence of phase transitions in hard combinatorial problems [1]. When some control parameters are tuned, many problems of practical importance indeed exhibit drastic changes of their behavior. The interest in such threshold phenomena has been enhanced by the observation that instances located at phase boundaries are the most difficult ones to solve. Even NP-complete problems [2,3] (whose solving times are thought to grow exponentially with their sizes) do not behave so badly far from the threshold. As a consequence, the results of *worst-case* complexity theory do not seem to be much relevant in practice and the need for a *typical-case* complexity theory has clearly emerged. Recently, the use of techniques and concepts of the statistical physics of disordered systems combined with numerical investigations have suggested that the nature of the transition taking place could be related to the upsurge of complexity at the threshold [4]. This conjecture can be best exemplified on the paradigm of combinatorial problems showing a phase transition behavior, that is the random K-Satisfiability (K-SAT) problem.

K-SAT is defined as follows. Consider N Boolean variables $\{x_i = 0,1\}_{i=1,\ldots,N}$. Choose randomly K among the N possible indices i and then, for each of them, choose 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, 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). For large instances ($M, N \to \infty$), numerical simulations and mathematical analysis indicate that the probability of finding a logical assignment of the $\{x_i\}$'s satisfying all the clauses falls abruptly from one down to zero when $\alpha = M/N$ crosses a critical value $\alpha_{\rm c}(K)$. Above $\alpha_{\rm c}(K)$, not all clauses can be satisfied simultaneously. This scenario is rigorously established for 2-SAT which is a polynomial problem and whose threshold $\alpha_{\rm c}(2)$ equals 1 [5]. When $K \geq 3$, K-SAT is NP-complete. Some upper and lower bounds on $\alpha_{\rm c}(K)$ have been derived and numerical simulations have recently allowed to find estimates of $\alpha_{\rm c}$, *e.g.* $\alpha_{\rm c}(3) \simeq 4.25$–$4.30$ [4,6,7].

[a] e-mail: biroli@physique.ens.fr

[b] *Present address*: Institut für Theoretische Physik, Bunsenstrasse 9, 37073 Göttingen, Germany

[c] Unité Mixte de Recherche du Centre National de la Recherche Scientifique et de l'École Normale Supérieure

When combining pM of clauses of length 3 with $(1-p)M$ clauses of length 2, one obtains the so-called $2+p$-SAT model, which smoothly interpolates between the instances of the easy-polynomial (2-SAT when $p=0$) and the hard-exponential (3-SAT when $p=1$) classes [4]. Statistical mechanics and replica theory show that there is a tricritical value $p_0 \simeq 0.4$ separating second-order SAT-UNSAT phase transitions for $p < p_0$ from random first-order SAT-UNSAT phase transitions for $p > p_0$. The change of the nature of the transition results from a change of the structure of the optimal Boolean assignments (satisfying all clauses when $\alpha < \alpha_c(2+p)$ or minimizing the number of violated clauses for $\alpha > \alpha_c(2+p)$) when crossing the threshold. As shown in [4], the SAT-UNSAT transition results from the freezing of a finite fraction of Boolean variables which acquire a constant value in all optimal assignments. The emergence of such over-constrained variables at $\alpha_c(2+p)$ appears to be continuous when $p < p_0$ and becomes strongly discontinuous above p_0. The existence of a $O(N)$ backbone of over-constrained variables at the threshold above the tricritical point has deep consequences. Indeed, a common search algorithm such as the Davis-Putnam procedure will fail with finite probability to correctly assert the first variable and will waste much time in exploring empty branches of the search tree before backtracking and correcting the early mistake. Numerical experiments strongly support this feeling: at the threshold $\alpha_c(2+p)$, the running time to solve an instance of the $2+p$-SAT problem behaves polynomially with N for $p \leq 0.4$ and exponentially for $p \geq 0.6$ [4].

A further understanding of the SAT-UNSAT transition undoubtedly requires a deeper knowledge of the organization of the optimal assignments. Information about the mutual (Hamming) distance between solutions, the size of the backbone, etc. is indeed of high relevance to understand and hopefully improve the efficiency of algorithms. From a statistical physics point of view, the main difficulty stems from the fact that K-SAT is naturally mapped onto a disordered spin model with finite connectivity. Although the lack of geometrical correlation in the clauses makes this model mean-field, the finite number of neighbours to each spin results in much stronger local field fluctuations and the theory is not as simple as its infinite-connectivity counterpart. Previous studies have shown that even at the simplest replica symmetric (RS) level, the order parameter describing finite-connectivity spin-glasses turns out to be a full distribution of effective fields [13]. Its determination requires to solve a functional self-consistent equation and is far from being easy. The situation becomes even worse and apparently mathematically intractable (except in some very peculiar cases [8]) when replica symmetry breaking (RSB) effects are taken into account.

To circumvent the difficulty of solving the RS or RSB self-consistent equations, we propose in this article a different strategy. Our claim is that a variational approach is of high efficiency to provide very precise results at a bearable calculation cost. Using some elementary information about the gross physical features of the K-SAT model, *e.g.* the existence of a backbone, we show that a RS variational calculation is able to recover all known results and to predict new ones (under certain assumptions) such as the value of the tricritical point $p_0 = 2/5$. In addition, we present some new results obtained from RSB variational calculations in both SAT and UNSAT regimes that unveil the structure of optimal assignments in the K-SAT problem. This paper is organised as follows. In Section 2, we recall the main steps of the statistical mechanics approach to the K-SAT problem [9,10]. We then explain the variational procedure to be followed depending on the particular phase, SAT or UNSAT, under investigation. Section 3 is devoted to the analysis of the structure of optimal configurations in the SAT phase. The SAT-UNSAT transition is studied in Section 4. For both Sections 3 and 4, we first focus on the replica symmetric variational solution and then expose the additional features corresponding to replica symmetry breaking effects. Finally, the emerging picture of the space of solutions is summed up and some perspectives may be found in Section 5.

2 Statistical mechanics and variational approach

2.1 Replica formalism and free-energy functional

In this section, we shall give an overview of the statistical mechanics approach to random K-satisfiability problems, see [9,11] for the original works on this subject. We adopt the Ising-spin notion, so a true Boolean variable is mapped onto $S_i = +1$, whereas a false variable gives $S_i = -1$. A logical assignment $\{S\}$ is a set of N spins S_i out of all 2^N possible configurations. We denote the (random) set of clauses by $\{C\}$. We choose the energy-cost function $\mathcal{H}[\{C\},\{S\}]$ to be the number of clauses violated by the configuration $\{S\}$ [9]. If the ground state energy is zero (respectively strictly positive), the logical clauses are satisfiable (resp. unsatisfiable). The free-energy density f of the resulting spin system at a formal temperature T is given by the logarithm of the partition function

$$Z[\{C\}] = \sum_{\{S\}} \exp(-\mathcal{H}[\{C\},\{S\}]/T) \,, \tag{1}$$

and is assumed to be self-averaging [12] as the size N of the instance of the K-SAT problem goes to infinity. In order to calculate the disorder average, the replica trick is used:

$$\overline{\ln Z} = \lim_{n\to 0} \partial_n \overline{Z^n} \tag{2}$$

where at first a positive integer number n is considered, and the replica limit $n \to 0$ is achieved by some kind of analytical continuation in n. Introducing the 2^n order parameters $c(\boldsymbol{\sigma})$ equal to the fractions of "sites" i such that $\sigma^a = S_i^a, \forall a = 1,\ldots,n$ [11], the thermodynamic limit of the free-energy density can be calculated by the saddle

point method from

$$f(K,\alpha,\beta) = \lim_{n\to 0} \frac{1}{\beta n} \sum_{\boldsymbol{\sigma}} c(\boldsymbol{\sigma}) \ln c(\boldsymbol{\sigma}) - \frac{\alpha}{\beta n} \times \ln \left[\sum_{\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_K} c(\boldsymbol{\sigma}_1)\ldots c(\boldsymbol{\sigma}_K) \prod_{a=1}^{n} \left(1 + \left(\mathrm{e}^{-\beta} - 1\right) \prod_{l=1}^{K} \delta_{\sigma_l^a,1} \right) \right] \tag{3}$$

through a maximization over all normalized – $\sum_{\boldsymbol{\sigma}} c(\boldsymbol{\sigma}) = 1$ – and even – $c(-\boldsymbol{\sigma}) = c(\boldsymbol{\sigma})$ – order parameters [13]. Eventually, the ground state properties are obtained as the temperature $T = 1/\beta$ is sent to zero in (3). Following [11], the first (respectively second) term on the r.h.s. of (3) will be hereafter called the effective entropy (resp. effective energy) contribution to the free-energy.

2.2 Simplest order parameter and replica symmetry

Finding the saddle-point $c(\boldsymbol{\sigma})$ of (3) is in general a very hard task. Since the functional (3) is invariant under permutations of the n replicas, it is possible to restrict the variational problem to the subspace of $c(\boldsymbol{\sigma})$ with the same permutation symmetry. In this so-called replica symmetric (RS) subspace, $c(\boldsymbol{\sigma})$ depends on the argument $\boldsymbol{\sigma}$ only through $\sum_{a=1}^{n} \sigma_a$. This allows the introduction of the generating function $P(h)$:

$$c(\boldsymbol{\sigma}) = \int_{-\infty}^{+\infty} \mathrm{d}h\ P(h) \prod_{a=1}^{n} \left(\frac{\mathrm{e}^{\beta h \sigma_a}}{\mathrm{e}^{\beta h} + \mathrm{e}^{-\beta h}} \right) \cdot \tag{4}$$

The normalization of $c(\boldsymbol{\sigma})$ implies the normalization of the generating function, $\int_{-\infty}^{+\infty} \mathrm{d}h\ P(h) = 1$. Plugging this form into (3), one can easily obtain the analytical continuation in n, finally getting [11]

$$\begin{aligned} f_{\mathrm{rs}}(\beta) = & -\frac{1}{\beta} \int \frac{\mathrm{d}h\ \mathrm{d}\nu}{2\pi} \mathrm{e}^{\mathrm{i}\nu h} P_{\mathrm{ft}}(\nu) \\ & \times \left[1 - \ln P_{\mathrm{ft}}(\nu)\right] \ln(2\cosh \beta h) \\ & - \frac{\alpha}{\beta} \int \prod_{l=1}^{K} \mathrm{d}h_l\ P(h_l) \\ & \times \ln \left[1 + (\mathrm{e}^{-\beta} - 1) \prod_{l=1}^{K} \left(\frac{\mathrm{e}^{-\beta h_l}}{2\cosh \beta h_l} \right) \right] \end{aligned} \tag{5}$$

where $P_{\mathrm{ft}}(\nu) = \int \mathrm{d}h \mathrm{e}^{-\mathrm{i}\nu h} P(h)$ denotes the Fourier transform of the generating function $P(h)$. The free energy (5) now has to be optimized with respect to $P(h)$.

To understand the physics hidden in this approach, it is useful to consider the Boolean magnetizations $m_i = \langle\langle S_i \rangle\rangle$, where $\langle\langle . \rangle\rangle$ denotes the Gibbs average with Hamiltonian $\mathcal{H}$ at fixed disorder $\{C\}$, see Section 2.1. An effective field h_i is associated to each local magnetization m_i through the relation $m_i = \tanh(\beta h_i)$. Within the RS framework, the order parameter $P(h)$ is simply the histogram of the effective fields,

$$P(h) = \frac{1}{N} \sum_{i=1}^{N} \overline{\delta(h - h_i)}, \tag{6}$$

where the overbar denotes the average over the random choices of clauses $\{C\}$. At very low temperature, effective fields are related to elementary excitations around ground state configurations.

2.3 More sophisticated order parameters and replica symmetry breaking

A corollary of ansatz (4) is that the Hamming distance d between any two assignments (*i.e.* the number of variables which are different in the two configurations) weighted with the Gibbs measure almost surely equals

$$d_{\mathrm{rs}} = \frac{1}{2} - \frac{1}{2} \int \mathrm{d}h\ P(h) \left(\tanh \beta h\right)^2, \tag{7}$$

once divided by N. In other words, on the N-dimensional hypercube whose vertices are the Boolean configurations, all relevant assignments belong to a single *cluster* of typical diameter $d_{\mathrm{rs}} N$. On general grounds, there is no *a priori* reason to trust this simple picture. At zero temperature for instance, there could well exist a non trivial geometrical organization of the space of solutions to the SAT or MAX-SAT problem which would give rise to a non-trivial (*i.e.* not fully concentrated) probability distribution for d. The simplest and immediate extension of (7) corresponds to a bimodal distribution for d, with two peaks in d_0 and $d_1 (< d_0)$. The corresponding picture on the hypercube of configurations is that solutions are now gathered into different clusters having average internal diameter $d_1 N$ and being separated by a typical distance $d_0 N$.

Let us label these clusters, also called pure states [13], by a new index Γ. In a given cluster Γ, the magnetizations m_i^Γ can be calculated as the average values of the spins S_i over the $\mathcal{N}_\Gamma$ assignments belonging to Γ. As before, it is convenient to consider the effective fields h_i^Γ through the relations $m_i^\Gamma = \tanh(\beta h_i^\Gamma)$. These effective fields fluctuate

- from state to state: for a given site i, the effective fields h_i^Γ depend on the cluster Γ. We introduce the histogram $\rho_i(h) = \sum_\Gamma \mathcal{N}_\Gamma\ \delta(h - h_i^\Gamma) / \sum_\Gamma \mathcal{N}_\Gamma$ to take these fluctuations into account;
- from spin to spin: in turn, $\rho_i(h)$ explicitly depends upon the index i of the variable it is related to. This multiplicity of field histograms is encoded in a functional probability distribution $\mathcal{P}[\rho] = \sum_{i=1}^{N} \delta[\rho(h) - \rho_i(h)]/N$ over the set of possible $\rho(h)$.

Within the replica formalism exposed in Section 2.1, the above picture corresponds to the first step of Parisi's hierarchical replica symmetry breaking (RSB)

scheme [13]. The RSB order parameter $c(\boldsymbol{\sigma})$ reads [11]

$$c(\boldsymbol{\sigma}) = \int \mathcal{D}\rho\mathcal{P}[\rho] \prod_{b=1}^{n/m} \left\{ \int_{-\infty}^{+\infty} \mathrm{d}h\ \rho(h) \times \prod_{a=1+(b-1)m}^{b\,m} \left(\frac{\mathrm{e}^{\beta h \sigma_a}}{\mathrm{e}^{\beta h} + \mathrm{e}^{-\beta h}} \right) \right\}. \quad (8)$$

With the above ansatz, the analytical continuation $n \to 0$ can be performed and the resulting free-energy (3) has to be optimized over $P[\rho(h)]$, see [11]. The parameter m in (8) determines the relative weights of d_0 and d_1 [11,13].

2.4 Variational approach

The direct way to complete the calculation of free energy (3) within the replica symmetric or the one-step broken approximation would be the following: a variation of (3) with respect to the order parameters yields functional equations for $P(h)$ or $\mathcal{P}[\rho(h)]$. In the replica symmetric case, this equation could be solved in [10] by a class of distributions consisting of a larger and larger number of Dirac peaks. In the replica symmetry broken case, only the very simplest possible solution could be obtained in [11]. The evaluation of any more involved solution seems a hopeless task due to the complexity of the saddle point equations [14].

In this paper a different route will be chosen [15]. Based on physical grounds, some simple trial functions for $P(h)$ or $\mathcal{P}[\rho(h)]$ will be proposed. These functions only depend on a small number of parameters; this fact significantly reduces the complexity of the problem. In the replica symmetric case, the exact results of [10] can be reproduced within a precision of less then one percent. In the replica symmetry broken case, new results can be obtained which are far beyond the solution given in [11].

2.5 Zero temperature limit and scaling of the effective fields

The phase transition in K-SAT separates a low α regime in which all variables are typically under-constrained (SAT regime) from a high α regime in which a finite fraction of variables is typically over-constrained (UNSAT regime).

Variables can be under-constrained when they do not appear in any clauses, or more generally when the minimal number of violated clauses is independent of their possible assignments (true or false). In the language of statistical physics, such under-constrained variables correspond at low temperature T to spins S_i submitted to effective fields h_i vanishing linearly with T: $h_i = Tz_i$. This way, their magnetizations $m_i = \tanh(\beta h_i) = \tanh z_i$ are different from ± 1 in the ground state. These unfrozen spins do not contribute to the energy when $T \to 0$, but only to the entropy. In the SAT phase, effective fields are expected to show this behavior for low temperatures.

Conversely, over-constrained variables correspond to spins S_i seeing effective fields h_i that remain finite, *i.e.* of the order of one in the zero temperature limit. The excitation energy to flip any of these spins S_i is finite and the spins are frozen in up or down directions depending on the signs of the associated fields h_i. In the zero temperature limit the only contribution to the energy comes from these frozen spins. To study the SAT-UNSAT transition, one therefore has to focus onto the probability distribution of effective fields on the scale of $O(1)$. Note that on this scale, the effective fields corresponding to unfrozen spins vanish and give rise to a Dirac peak centered at zero; the weight of this $h = 0$ peak is precisely the fraction of under-constrained spins.

3 The satisfiable phase

3.1 Replica symmetric approximation

As already discussed in the previous section, the interesting quantity to be calculated in the satisfiable phase ($\alpha < \alpha_c$) is the ground state entropy density $s = -\lim_{\beta\to\infty} \beta f$. In the replica symmetric approximation, the entropy s reads, according to equation (5),

$$s_{\mathrm{rs}} = \alpha \int \prod_{l=1}^{K} \mathrm{d}z_l\ \tilde{P}(z_l) \ln \left[1 - \prod_{l=1}^{K} \frac{\mathrm{e}^{-z_l}}{2\cosh z_l} \right] + \int \mathrm{d}z\ \tilde{P}(z) \ln[2\cosh z] - \int \frac{\mathrm{d}z\ \mathrm{d}\nu}{2\pi} \mathrm{e}^{\mathrm{i}z\nu} \tilde{P}_{\mathrm{ft}}(\nu) \ln[\tilde{P}_{\mathrm{ft}}(\nu)] \ln[2\cosh z]\ , \quad (9)$$

where $z = \beta h$ is the rescaled effective field of order one, see Section 2.5. As a consequence, the distribution $\tilde{P}(z)$ has a finite variance in the limit $\beta \to 0$. As in the previous section, $\tilde{P}_{\mathrm{ft}}(\nu)$ denotes the Fourier transform of $\tilde{P}$.

We start with a simple Gaussian ansatz for the rescaled field distribution,

$$\tilde{P}(z) = G_{\Delta}(z), \quad (10)$$

where $G_{\Delta}(z)$ denotes a Gaussian distribution with zero mean and variance Δ. Note that $\tilde{P}(z)$ is expected to be even due to the symmetry of the disorder distribution. In the case of infinite connectivity spin glasses equation (10) would give the exact distribution, *cf.* [13], but due to the finite connectivities in K-SAT effective fields are not necessarily Gaussianly distributed. We find the expression

$$s_{\mathrm{rs}} = \alpha \int \prod_{l=1}^{K} \mathrm{D}z_l \ln \left[1 - \prod_{l=1}^{K} \frac{\mathrm{e}^{-\sqrt{\Delta} z_l}}{2\cosh \sqrt{\Delta} z_l} \right] + \int \mathrm{D}z \left(\frac{3 - z^2}{2} \right) \ln[2\cosh \sqrt{\Delta} z] \quad (11)$$

which has to be optimized numerically with respect to the variational parameter Δ. Hereafter, $\mathrm{D}z = G_1(z)\mathrm{d}z$ denotes the Gaussian measure with zero mean and variance

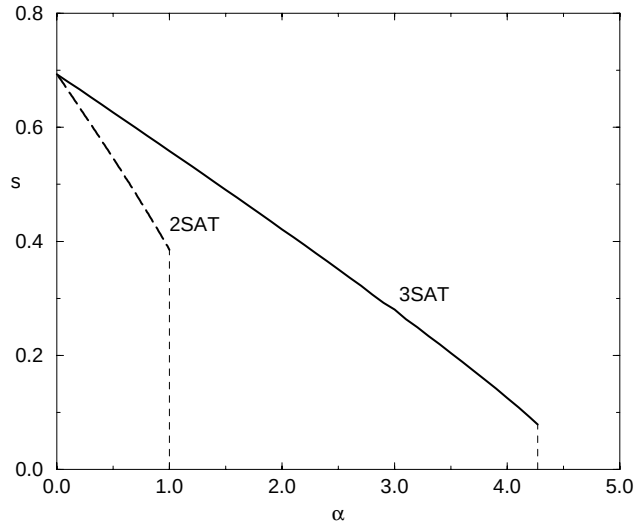

Fig. 1. Variational entropy s of the solutions for the 2-SAT (bold dashed line) and the 3-SAT (full line) problems as functions of the ratio of clauses per variable α. The curves are practically indistinguishable from the full RS results in [9]. The vertical dashed lines indicate the threshold $\alpha_c(2) = 1$ and $\alpha_c(3) \simeq 4.27$ [4].

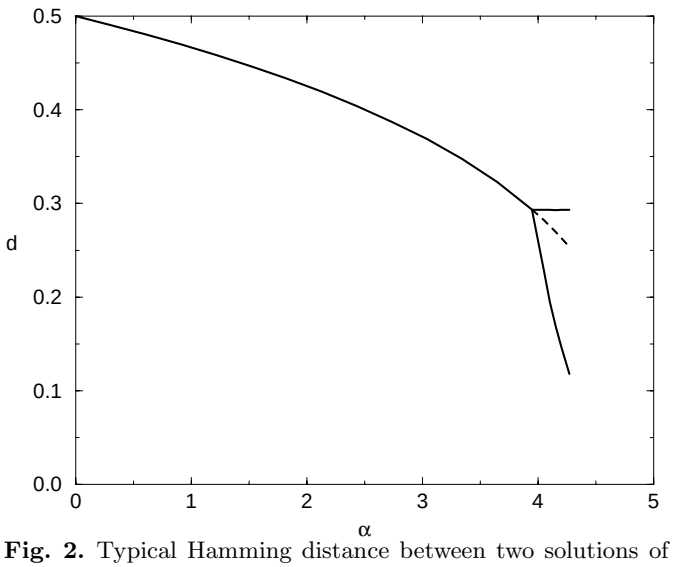

Fig. 2. Typical Hamming distance between two solutions of a random 3-SAT problem. Whereas there is only one distance d in the replica symmetric phase $0 \leq \alpha < \alpha_s \simeq 3.96$ which is monotonously decreasing with the number of clauses per variables α, we find two characteristic lengths in the replica symmetry broken case $\alpha_s < \alpha < \alpha_c \simeq 4.25$–$4.30$ [4,6]. The Hamming distance d_0 between two clusters of solutions (upper line) remains almost constant with α, whereas the entropy loss is mostly due to a shrinking of the typical size d_1 of the clusters themselves (lower curve). The dashed line denotes the continuation of the replica symmetric result.

one. For $\alpha = 0$ the variational parameter is found to be $\Delta = 0$, and the entropy follows to be $s_{rs} = \ln 2$: there are no clauses and the solution space coincides with the full phase space of the model. For increasing α, the entropy diminishes due to the growing number of constraints, see Figure 1. Our results are practically indistinguishable from the exact expansion of s_{rs} in powers of α performed in [9]. In Figure 2, we show the typical Hamming distance d_{rs} (7) between two solutions. d_{rs} monotonously decreases from $d = 0.5$ at $\alpha = 0$. This behavior signals a concentration of the solutions in configuration space.

In order to test the robustness of the results obtained with help of ansatz (10) we have repeated the above calculation with an exponential ansatz for $\tilde{P}(z)$. The values of the entropy were changed by less than 1%. We have also taken into account the presence of free spins through the ansatz

$$\tilde{P}(z) = (1 - A)\, \delta(z) + A\, G_{\Delta}(z) \tag{12}$$

for the rescaled distribution. The Dirac peak accounts for the variables which are not contained in the clauses as well as for spins having an effective field going to zero faster than linearly with the formal temperature T. Therefore $1 - \exp(-\alpha K)$, *i.e.* the fraction of variables present in the clauses, constitutes a rigorous upper bound for A which was explicitly violated by the ansätze considered so far. For 2-SAT at $\alpha = 1$, we find $A = 0.57 \pm 0.01$ (and 0.71 ± 0.02 if the Gaussian in (12) is again replaced by an exponential distribution), which has to be compared to the bound $1 - \exp(-2) = 0.865$. The strong dependence of A on the non-zero field part of the ansatz probably results from the inclusion of small but non-zero rescaled fields into the Dirac peak. For 3-SAT, differences are less drastic: whereas the upper bound is almost 1, the above ansatz gives $A = 0.94 \pm 0.02$ (and an A numerically indistinguishable from 1 for the exponential distribution) at $\alpha = 4$. In all cases, all these variations affect the entropy value by 1% at the most.

3.2 The replica symmetry breaking transition

We have already underlined in Section 2.3 that the replica symmetric ansatz is unable to reflect any non-trivial organization of the optimal assignments space. To investigate the ground state structure of K-SAT, we thus consider a one-step replica symmetry broken (RSB) ansatz. According to the discussion of Section 2.3, we choose

$$\mathcal{P}[\rho(z)] = \int_{-\infty}^{\infty} \mathrm{d}z\; G_{\Delta_0}(z) \times \delta\left[\rho(\tilde{z}) - \frac{G_{\Delta_1}(\tilde{z} - z)(2\cosh\tilde{z})^m}{\int \mathrm{d}z'\; G_{\Delta_1}(z' - z)(2\cosh z')^m}\right], \tag{13}$$

which coincides with the exact one-step expression for infinite-connectivity spin glass models. As in the RS hypothesis (9), $z = \beta h$ is a rescaled field which remains of order one in the zero-temperature limit. The detailed calculation of the variational RSB ground state entropy s_{rsb} in the satisfiable phase is exposed in Appendix A.

$$s_{\rm rsb} = \frac{\alpha}{m}\int\prod_{l=1}^{K}{\rm D}z_l\ \ln\frac{\int\prod_l {\rm D}\tilde{z}_l\left(\prod_l 2\cosh(\sqrt{\Delta_1}\tilde{z}_l+\sqrt{\Delta_0}z_l)-\prod_l\exp\{\sqrt{\Delta_1}\tilde{z}_l+\sqrt{\Delta_0}z_l\}\right)^m}{\int\prod_l {\rm D}\tilde{z}_l\left(\prod_l 2\cosh(\sqrt{\Delta_1}\tilde{z}_l+\sqrt{\Delta_0}z_l)\right)^m}$$
$$+\frac{1}{m}\int{\rm D}z\ \ln\int{\rm D}\tilde{z}\left(2\cosh(\sqrt{\Delta_1}\tilde{z}+\sqrt{\Delta_0}z)\right)^m$$
$$-\int{\rm D}z\frac{\int{\rm D}\tilde{z}\left(2\cosh(\sqrt{\Delta_1}\tilde{z}+\sqrt{\Delta_0}z)\right)^{m-1}\sinh(\sqrt{\Delta_1}\tilde{z}+\sqrt{\Delta_0}z)\ (\sqrt{\Delta_1}\tilde{z}+\sqrt{\Delta_0}z)}{\int{\rm D}\tilde{z}\left(2\cosh(\sqrt{\Delta_1}\tilde{z}+\sqrt{\Delta_0}z)\right)^m}\,. \tag{14}$$

The result reads

see equation (14) above.

This quantity has to be optimized with respect to the variational parameters Δ_0, Δ_1 and m. The numerical problem in calculating the solutions of the three equations

$$0=\frac{\partial s_{\rm rsb}}{\partial\Delta_0}=\frac{\partial s_{\rm rsb}}{\partial\Delta_1}=\frac{\partial s_{\rm rsb}}{\partial m} \tag{15}$$

for 3-SAT consists in the sixfold integration in the first term of (14). It is much easier to determine the critical $\alpha_{\rm s}$ where the first nontrivial solution of these equations can be found. In principle, due to the continuity of the entropy $s_{\rm rsb}$ at this transition, there are two possible scenarios [16],

- A continuous transition in $\Delta_0\to\Delta_{\rm rs}$, $\Delta_1\to 0$ takes place, where $\Delta_{\rm rs}$ here and in the following denotes the replica symmetric value. This could be connected to a nontrivial $m_{\rm s}$ with $0<m_{\rm s}<1$.
- A jump in Δ_1 towards a nontrivial value >0 takes place at the transition. To guarantee the continuity of $s_{\rm rsb}$, such a transition has to happen at $m=1$ and $\Delta_0=\Delta_{\rm rs}$.

In the following two subsections we will consider both possibilities. Whereas a transition of the first type can be found at a certain $\alpha_{\rm s}$, the second possibility can be ruled out. The value of $\alpha_{\rm s}$ found this way constitutes an upper bound for the exact threshold. We can indeed not exclude that, by taking into account a larger variety of density functionals (13), a non-trivial RSB solution could appear already at smaller values of α.

3.2.1 The continuous transition

In order to determine the critical value $\alpha_{\rm s}$ for the replica symmetry breaking transition inside the SAT phase, we have to explicitly use equation (15),

$$\frac{\partial s_{\rm rsb}}{\partial\Delta_1}=0, \tag{16}$$

and expand it to first order in Δ_1. As a result of the expansion, the interior integrals over the $\tilde{z}_l$ in (14) can be carried out analytically, leaving only three integral to be evaluated numerically. At the zeroth order, the replica symmetric saddle point equation for Δ_0 is recovered. At the first order, the coefficient of the linear term in Δ_1 vanishes at

$$\alpha_{\rm s}=3.955\pm 0.005, \tag{17}$$

thus allowing a non-zero solution for Δ_1 to develop. This value is in surprisingly good agreement with a critical slowing down found numerically by Svenson and Nordahl [17]. They considered a simple zero-temperature Glauber dynamics for random satisfiability and coloring problems. In the case of 3-SAT this dynamics showed an exponential relaxation down to (almost) zero energy density for $\alpha<4$, whereas the relaxation became algebraic for $\alpha>4$ – converging towards non-zero energy. As we shall see in Section 3.3, the increase of Δ_1 at $\alpha_{\rm s}$ coincides with the emergence of a non trivial structure of the optimal assignments of a typical 3-SAT instance. Due to the continuous nature of the transition at $\alpha_{\rm s}$ it is probable that higher-lying, metastable states blow up simultaneously with the breaking up of the ground state structure. If it were so, the relationship between Svenson's and Nordahl's result and the static transition at $\alpha_{\rm s}$ could perhaps be explained along the lines developed in the context of the off-equilibrium dynamics of spin-glasses[18].

In order to calculate also $m_{\rm s}$ at the transition, we have to take into account either the second order terms of equations (16) slightly above $\alpha_{\rm s}$, or to explicitly solve the full saddle point equations (15) in the limit $\alpha\to\alpha_{\rm s}^+$. We have followed the second route [19] and found $m_{\rm s}\approx 0.8$. The corrections to the entropy of solutions is however very weak, $s_{\rm rs}=0.917$ while $s_{\rm rsb}=0.911$ at $\alpha=4.2$.

For 2-SAT, no such transition can be found before the SAT-UNSAT-transition. A numerical investigation of the $2+p$-SAT model makes us conjecture that the existence of a replica symmetry breaking transition within the SAT phase is related to the appearance of a discontinuous SAT-UNSAT transition, see Section 4.

3.2.2 Nonexistence of a discontinuous transition

As already discussed above, one could also imagine a discontinuous transition with a jump in Δ_1 even at a lower value of α. Due to the continuity of the ground-state entropy, we would expect this transition to be continuous in m, *i.e.* to happen at $m=1$.

The most interesting saddle point equation to be considered here is hence the m-equation. Exactly at the transition, a non-zero solution for Δ_1 should show up

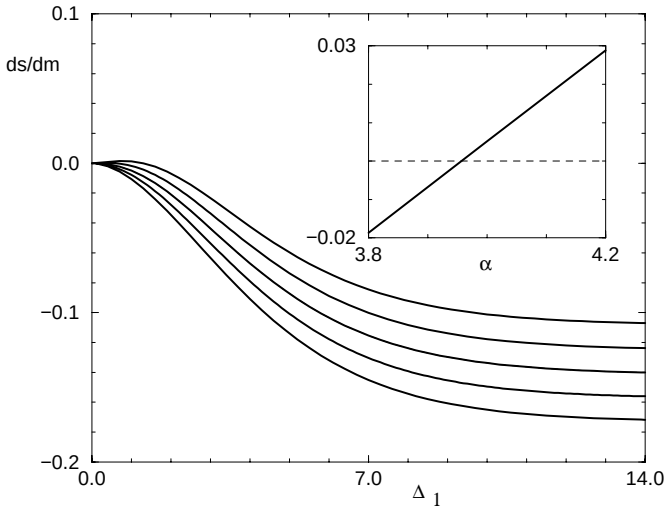

Fig. 3. $\partial_m s_{\rm rsb}$ $(m = 1, \Delta_0 = \Delta_{\rm rs}, \Delta_1)$ as function of Δ_1. The figure shows the curves for $\alpha = 3.7, 3.8, 3.9, 4.0, 4.1$ (from bottom to top), illustrating the non-existence of a discontinuous transition. Inset: the second derivative $\partial^2_{\Delta_1}(\partial_m s_{\rm rsb})$ vanishes at $\alpha_{\rm s} = 3.955 \pm 0.005$.

in the equation

$$0 = \frac{\partial s_{\rm rsb}}{\partial m} \quad (m = 1, \Delta_0 = \Delta_{\rm rs}, \Delta_1) \ . \tag{18}$$

The other two variational equations (16) are automatically fulfilled at this point. In Figure 3, we display $\partial_m s_{\rm rsb}$ $(m = 1, \Delta_0 = \Delta_{\rm rs}, \Delta_1)$ for several values of α as a function of Δ_1. For $\alpha < \alpha_{\rm s}$, *i.e.* below the continuous transition found in the last subsection, the function is monotonously decreasing with Δ_1, towards some finite asymptotic value. This clearly rules out any discontinuous transition in Δ_1 – the only zero of this function lies at $\Delta_1 = 0$. At the continuous transition the behavior in the vicinity of $\Delta_1 = 0$ changes. The sign of the second derivative of $\partial_m s_{\rm rsb}$ with respect to Δ_1 changes whereas the first derivative is always zero. This confirms again the local instability of the replica symmetric solution leading to the continuous transition found in the previous paragraph.

3.3 Multiplicity of clusters

In this section, we give a geometrical interpretation of the RSB transition. Above $\alpha_{\rm s} \simeq 3.96$, the ground state configurations are divided into an exponential number of well separated clusters. We are interested in the distribution of the entropy densities s of these clusters, *i.e.* we want to count the clusters containing $\sim {\rm e}^{Ns}$ satisfying assignments. This number will be denoted by ${\rm e}^{N\omega(s)}$. The quantity $\omega(s)$ is hereafter referred to as the *multiplicity* of s. By definition of ω,

$$Z_m = \int {\rm d}s \ {\rm e}^{N\omega(s)} \left({\rm e}^{Ns}\right)^m \tag{19}$$

$$= \lim_{\beta\to\infty} \sum_{\Gamma} \left(\sum_{\{S_i\}\in\Gamma} {\rm e}^{-\beta\mathcal{H}[\{C\},\{S\}]} \right)^m \tag{20}$$

where Γ denotes the clusters and $\mathcal{H}$ the energy cost function of Section 2.1. m is now a control parameter that can be varied to obtain $\omega(s)$. A straightforward calculation of (19) show that $\tau(m) = \log Z_m/N$ is simply the Legendre transform of $\omega(s)$ in the large N limit,

$$\tau(m) = \lim_{N\to\infty} \frac{1}{N} \log Z_m = {\rm extr}_s[\omega(s) + ms]. \tag{21}$$

Thus to access the multiplicity ω, we resort to the calculation of τ, following closely the lines of [20] (see also [21] and [22] for related calculations on the $p-$spin glass model and neural networks). We use again the replica trick, $\overline{\ln Z_m} = \lim_{n\to 0} \partial_n \overline{(Z_m)^n}$, and represent also the mth power in (20) by a positive integer-valued m. This leads to $n.m$ replicas of the original system obeying the one-step RSB algebra[20]. Within our Gaussian variational scheme, we easily find

$$\tau(m) = m \ {\rm extr}_{\Delta_0,\Delta_1} \left[s_{\rm rsb}(m, \Delta_0, \Delta_1) \right] \tag{22}$$

where $s_{\rm rsb}$ is given by (14). Consequently, the dominant clusters considered so far and obtained by optimization of $s_{\rm rsb}$ over m have zero multiplicity $\omega = -m^2 \partial s_{\rm rsb}/\partial m = 0$, *i.e.* their number is less than exponential in the system size N. Simultaneously, there exist exponentially numerous clusters with lower entropies such that the total number of satisfying assignments they contain remains much smaller than ${\rm e}^{N s_{\rm rsb}}$. These subdominant states appear as soon as α gets larger than $\alpha_{\rm s}$. Below this transition there is no positive multiplicity at all; almost all solutions are collected in one large cluster.

In Figure 4 we show the results for $\alpha = 4.2$ (we have checked the qualitative similarity of the curves for other values of α) obtained through a numerical solution of the variational equations in Δ_0 and Δ_1. At a certain $s = s_{\rm rsb} \simeq 0.911$ the multiplicity becomes positive, and the curve starts with slope $-m_{\rm rsb}$, where $m_{\rm rsb} \simeq 0.72$ is the value of m that optimizes $s_{\rm rsb}$. At $m = 0$, *i.e.* where the slope of ω over s vanishes, we find again the replica symmetric entropy $s_{\rm rs} \simeq 0.917$ calculated at the beginning of this section. However, the curve is only reliable up to the cusp: there, the second derivative ${\rm d}^2 s_{\rm rsb}/{\rm d}m^2$ changes sign and the corresponding variational solution becomes unstable [23]. Note also that the RS entropy $s_{\rm rs} = \tau(1)$ is found back at $m = 1$ in the unphysical negative ω region. As happens also in the case of the spherical $p-$spin glass, there could already be an instability of the one-step solution due to instable replicon modes [24], but the deviations from the given curve are expected to be very weak.

Although the order of magnitude of the multiplicity calculated above is small, some drastic changes take place

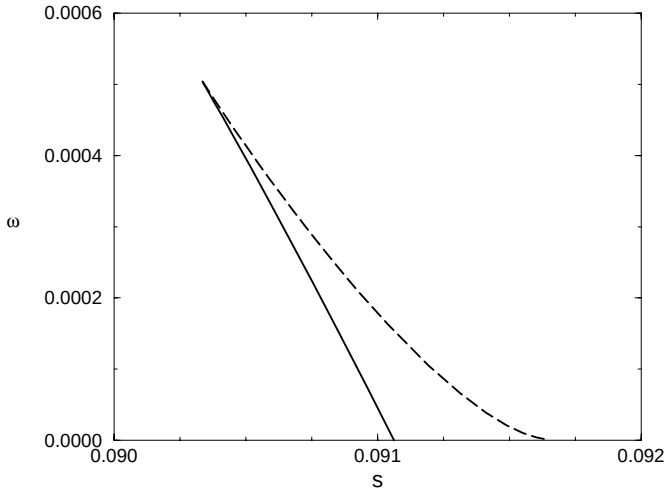

Fig. 4. Multiplicity $\omega(s)$ of states with entropy s at $\alpha = 4.2$. The curve intersects the zero multiplicity axis at $s_{\rm rsb} \simeq 0.911$ and $s_{\rm rs} \simeq 0.917$. The full line shows the reliable part of the curve. Along the dashed line, the second derivative of $s_{\rm rsb}$ is positive and the one-step replica symmetry broken ansatz is not longer valid. The curves for other values of $\alpha > \alpha_{\rm s}$ are qualitatively similar.

at $\alpha_{\rm s}$. This can best seen on the typical Hamming distances between solutions. The quantity

$$d_1 = \frac{1}{2} - \frac{1}{2}\int \mathrm{D}z \times \frac{\int \mathrm{D}\tilde{z} \cosh^m(\sqrt{\Delta_0}z + \sqrt{\Delta_1}\tilde{z}) \tanh^2(\sqrt{\Delta_0}z + \sqrt{\Delta_1}\tilde{z})}{\int \mathrm{D}\tilde{z} \cosh^m(\sqrt{\Delta_0}z + \sqrt{\Delta_1}\tilde{z})} \tag{23}$$

describes the average distance between two solutions inside the same cluster, whereas

$$d_0 = \frac{1}{2} - \frac{1}{2}\int \mathrm{D}z \times \left[\frac{\int \mathrm{D}\tilde{z} \cosh^m(\sqrt{\Delta_0}z + \sqrt{\Delta_1}\tilde{z}) \tanh(\sqrt{\Delta_0}z + \sqrt{\Delta_1}\tilde{z})}{\int \mathrm{D}\tilde{z} \cosh^m(\sqrt{\Delta_0}z + \sqrt{\Delta_1}\tilde{z})}\right]^2 \tag{24}$$

stands for the distance between two clusters. The results for the thermodynamically dominant states are shown in Figure 2. We observe that d_0 is almost constant in α, *i.e.* the relative positions of the clusters in configuration space remain roughly unchanged if new clauses are added to a given sample. In contrast to this behavior, the disappearance of solutions as α grows is accompanied with a rapid decrease of the cluster diameter d_1.

3.4 Breakdown of the scaling of the effective fields

All the ansätze we have used in the description of the SAT phase were based on the assumption of effective fields linearly vanishing in the zero-temperature limit. As we have argued, this scaling is no longer valid above the SAT-UNSAT transition. So we can extract a first variational estimate of the critical value $\alpha_{\rm c}$ from the divergence of Δ in the replica symmetric and of Δ_1 in the replica symmetry broken case. Compared with the numerical value $\alpha_{\rm c} = 4.25$–4.3 [6,7,4] the resulting values 4.76 (RS), resp. 4.66 (RSB) are rather crude approximation. The replica symmetric value can already be improved by taking (12) instead of (10). We find $\alpha_{\rm c} = 4.622$ with $A = 0.935$, which is rather similar to the iterative replica symmetric result in [9]. As we shall show in the next Section, physically more elaborate approximations are needed to obtained better results for $\alpha_{\rm c}$.

4 The SAT-UNSAT transition

4.1 Replica symmetric calculation

4.1.1 Variational RS free-energy

According to Sections 2.2 and 2.5, we propose the following variational (replica symmetric) field distribution in the zero temperature limit

$$P(h) = (1-B)\delta(h) + \frac{B}{\sqrt{\Delta}}\,\Phi\left(\frac{h}{\sqrt{\Delta}}\right). \tag{25}$$

$\Phi(x)$ is an even and decreasing probability distribution with argument $x = O(1)$. B denotes the fraction of frozen spins and Δ ($= O(1)$ when $T \to 0$) the typical squared magnitude of the effective fields acting on frozen spins.

Once $\Phi(x)$ has been chosen, we plug the trial variational function (25) into (5). The resulting variational problem involves two parameters B and Δ only and is therefore considerably simpler than the initial one. In the zero temperature limit we obtain from (5),

$$f_{\rm rs}(B,\Delta,\alpha,p) = -2\sqrt{\Delta} \times \left(\frac{B}{\pi}\int_0^{+\infty} \frac{\mathrm{d}\nu}{\nu}\,\Phi'_{\rm ft}(\nu)\ln[1-B+B\Phi_{\rm ft}(\nu)] - \alpha\int_0^{1/(2\sqrt{\Delta})} \mathrm{d}h\left\{(1-p)B^2\left[\Phi_{\rm cc}(h)\right]^2 + pB^3\left[\Phi_{\rm cc}(h)\right]^3\right\}\right) \tag{26}$$

where

$$\Phi_{\rm ft}(\nu) = \int_{-\infty}^{+\infty} \mathrm{d}x\ \mathrm{e}^{-\mathrm{i}x\nu}\Phi(x) \tag{27}$$

$$\Phi_{\rm cc}(h) = \int_h^{+\infty} \mathrm{d}x\Phi(x) \tag{28}$$

are respectively the Fourier transform and the complementary cumulative function of Φ. The above free-energy (26) corresponds to the 2+p-SAT problem which smoothly interpolates between 2-SAT ($p = 0$) and 3-SAT ($p = 1$), *cf.* Section 1. For the sake of completeness, we give in Appendix B the derivation of (26) for the special case of a Gaussian distribution $\Phi(x) = G_1(x)$.

As we shall show below, within this simplified version of the variational problem it is possible to obtain results which are only slightly different from the ones obtained from the best replica symmetric solution [9]. The simplicity of this approach leads to a more transparent description of the SAT-UNSAT transition.

4.1.2 A smooth transition: the 2-SAT problem

We start with $p = 0$, *i.e.* the 2-SAT case. From previous numerical and analytical studies, it is known that the fraction of frozen spins is continuous at the transition and is zero for $\alpha < \alpha_c$ [4,9]. Actually a numerical analysis of $f_{rs}(B, \Delta, \alpha, 0)$ excludes the possibility of a first order transition in B and Δ. To locate the critical value of α, we expand $f_{rs}(B, \Delta, \alpha, 0)$ around $B = 0$ and $\Delta = 0$. To the leading order and neglecting irrelevant terms in Δ, we find

$$f_{rs}(B, \Delta, \alpha, 0) \simeq -2\sqrt{\Delta}\left(B^2 f_{rs}^{(2)}(\alpha) + B^3 f_{rs}^{(3)}\right) \quad (29)$$

where

$$f_{rs}^{(2)}(\alpha) = \frac{1}{\pi}\int_0^{+\infty} \frac{d\nu}{\nu} \Phi'_{ft}(\nu)[\Phi_{ft}(\nu) - 1] - \alpha \int_0^{\infty} dh\, [\Phi_{cc}(h)]^2 \quad (30)$$

$$f_{rs}^{(3)} = -\frac{1}{2\pi}\int_0^{+\infty} \frac{d\nu}{\nu} \Phi'_{ft}(\nu)[\Phi_{ft}(\nu) - 1]^2. \quad (31)$$

$f_{rs}^{(3)}$ is clearly positive. Therefore the maximum of f_{rs} is located at $B = 0$ if $f_{rs}^{(2)}(\alpha) \geq 0$ and at $B > 0$ if $f_{rs}^{(2)}(\alpha) < 0$.

It is easy to demonstrate that the threshold α_c, determined through the condition $f_{rs}^{(2)}(\alpha_c) = 0$ is *always* equal to unity *independently* of the choice of the probability distribution Φ. To do so, we rewrite the first term on the r.h.s. of (30), that is $f_{rs}^{(2)}(0)$ using the definition (27) of Φ_{ft},

$$f_{rs}^{(2)}(0) = -\int_{-\infty}^{\infty} dx dy\, x\, \Phi(x)\Phi(y)w(x, y) \quad (32)$$

with

$$w(x, y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{d\nu}{i\nu} e^{i\nu x}(e^{i\nu y} - 1) = \frac{1}{2}\text{sign}(x + y) - \frac{1}{2}\text{sign}(x). \quad (33)$$

Inserting (33) in (32), a simple calculation leads to

$$f_{rs}^{(2)}(0) = \int_0^{\infty} dh\, [\Phi_{cc}(h)]^2, \quad (34)$$

and therefore to the reported result $\alpha_c = 1$. Surprisingly, the variational RS calculation is able to recover the exact threshold of 2-SAT [5] in a very robust manner. Note however, that the continuous growth of the backbone B above α_c depends on the choice of Φ.

4.1.3 A discontinuous transition: the 3-SAT problem

We now focus on the 3-SAT case ($p = 1$). Previous numerical and analytical studies have shown that spins freeze discontinuously at the transition [4,9]. Thus, we cannot locate the threshold through an expansion of $f_{rs}(B, \Delta, \alpha, 1)$ as in the 2-SAT case. The full variational calculation can nevertheless be simplified due to the following observation. In the SAT phase ($B = 0$), the free-energy is identically zero. Within a first-order transition scenario, the threshold will be the value of α at which the free energy of the UNSAT phase ($B \neq 0$, $\Delta \neq 0$) changes sign to become thermodynamically stable. The calculation of α_c becomes simpler once the free-energy (26) is rewritten as

$$f_{rs}(B, \Delta, \alpha, 1) = 2\sqrt{\Delta}\, B^3 \left(-s_{rs}(B) + \alpha\, e_{rs}(\Delta)\right) \quad (35)$$

with

$$s_{rs}(B) = \frac{1}{\pi B^2}\int_0^{+\infty} \frac{d\nu}{\nu} \Phi'_{ft}(\nu) \ln[1 - B + B\, \Phi_{ft}(\nu)] \quad (36)$$

$$e_{rs}(\Delta) = \int_0^{1/(2\sqrt{\Delta})} dh\, [\Phi_{cc}(h)]^3. \quad (37)$$

Calling B_c (respectively Δ_c) the argument where s_{rs} (resp. e_{rs}) reaches its minimal (resp. maximal) value, we obtain from (36,37) the following expression of the threshold

$$\alpha_c = \frac{\min_B s_{rs}(B)}{\max_\Delta e_{rs}(\Delta)} = \frac{s_{rs}(B_c)}{e_{rs}(\Delta_c)}. \quad (38)$$

The maximum of e_{rs} is obviously located at $\Delta_c = 0$ whereas the precise value of B_c depends of the field distribution Φ. We list below the results obtained for three different choices.

- *Gaussian distribution:* $\Phi(x) = G_1(x)$, $B_c \simeq 0.935$, $\alpha_c \simeq 4.622$.
- *Exponential distribution:* $\Phi(x) = \frac{1}{2}e^{-|x|}$, $B_c \simeq 0.976$, $\alpha_c \simeq 4.617$.
- *Lorentzian distribution:* $\Phi(x) = \frac{1}{\pi\,(1+x^2)}$, $B_c \simeq 0.986$, $\alpha_c \simeq 4.983$.

Above the threshold, B and Δ both increase with α from their critical values B_c and $\Delta_c (= 0)$. The variational approach is thus able to reproduce the qualitative picture of the mixed nature of the phase transition (second order in Δ and first order in the backbone size B) that emerged from the iterative RS solution [9] and numerics [4]. This prediction is quite robust with respect to the choice of $\Phi(x)$. Even a Lorentzian distribution gives reasonable results for α_c and B_c though its large-field tail is not physically sensible, see Section 5.2.

From a quantitative point of view, the above results for the Gaussian and the exponential case differ from the iterative RS solution $B_c \simeq 0.94$, $\alpha_c \simeq 4.60$ [9] by a few percent only. However, the latter was derived through a much less convenient iterative scheme [9].

4.1.4 The tricritical point p_0

As we have seen above, the main difference between 2-SAT and 3-SAT lies in the behaviour of the fraction of frozen spins at the transition. In other words, the backbone size at threshold vanishes in the former case ($B_{\rm c}(p=0)=0$), while it exhibits a discontinuous jump in the latter case ($B_{\rm c}(p=1)>0$). It is natural to expect the existence of a tricritical point p_0 separating continuous SAT-UNSAT transitions ($p<p_0$) from discontinuous ones ($p>p_0$) [4,10].

When $p<p_0$, the transition can be studied through an expansion of the free-energy (26) in powers of the backbone size, see (29),

$$f_{\rm rs}(B,\Delta,\alpha,p) \simeq -2\sqrt{\Delta}\left(B^2\ f_{\rm rs}^{(2)}(\alpha,p)+B^3\ f_{\rm rs}^{(3)}(\alpha,p)\right) \tag{39}$$

where, using (32,33,34),

$$f_{\rm rs}^{(2)}(\alpha,p) = \big(1-(1-p)\ \alpha\big)\ \int_0^\infty {\rm d}h\,[\Phi_{\rm cc}(h)]^2 \tag{40}$$

$$f_{\rm rs}^{(3)}(\alpha,p) = -\frac{1}{2\pi}\int_0^{+\infty}\frac{{\rm d}\nu}{\nu}\,\Phi'_{\rm ft}(\nu)[\Phi_{\rm ft}(\nu)-1]^2 - \alpha\ p\int_0^\infty {\rm d}h\,[\Phi_{\rm cc}(h)]^3\,. \tag{41}$$

As long as $f_{\rm rs}^{(3)}$ remains positive, the threshold is situated at $\alpha_{\rm c}(2+p)=1/(1-p)$ (40). As in the 2-SAT case, this result does not depend on the distribution Φ in (25) and coincides with the rigorous result found in [25] for $p<\frac{2}{5}$.

At a given p and slightly above the threshold, the backbone size scales as

$$B \sim \frac{\alpha-\alpha_{\rm c}(2+p)}{f_{\rm rs}^{(3)}\big(\alpha_{\rm c}(2+p),p\big)}, \tag{42}$$

up to a constant multiplicative factor. The tricritical point p_0 can thus be found through the condition $f_{\rm rs}^{(3)}\big(\alpha_{\rm c}(2+p),p\big)=0$. This statement remains unaffected by the inclusion of higher order terms in Δ in the expansion (39). The corresponding values of p_0 for the three choices of Φ of the previous paragraph are: $p_0\simeq 0.437$ for the Gaussian distribution, $p_0=3/7\simeq 0.429$ for the exponential distribution and $p_0\simeq 0.418$ for the Lorentzian distribution. As expected, these values are slightly higher than the prediction of the iterative RS solution, $0.4\le p_0<0.416$ [10].

We shall now show under some assumptions exposed in Appendix C that the tricritical point is precisely located at $p_0=2/5$. To do so, we proceed in two steps. Firstly, we recall that the equality $\alpha_{\rm c}(p)=1/(1-p)$ for $p\le 0.4$ has been rigorously demonstrated in [25]. Secondly, using the RS variational approach, we have seen above that $\alpha_{\rm c}(2+p)=1/(1-p)$ up to a tricritical p_0 which depends of Φ through the condition $f_{\rm rs}^{(3)}(1/(1-p_0),p_0)=0$. Consider now two different ansätze $\Phi^{(1)}$ and $\Phi^{(2)}$ such that the corresponding tricritical points satisfy $p_0^{(1)}<p_0^{(2)}$. Then, for p in the range $p_0^{(1)}<p<p_0^{(2)}$, we have $\alpha_{\rm c}^{(2)}(2+p)=1/(1-p)$ by definition of $p_0^{(2)}$ and $\alpha_{\rm c}^{(1)}(2+p)<1/(1-p)$ (the 2-clauses part of the formula is almost surely satisfiable if and only if $\alpha(1-p)\le\alpha_{\rm c}(2)=1$ giving thus this upper bound to the threshold, see [4]). Let us choose α with $\alpha_{\rm c}^{(1)}(p)<\alpha<\alpha_{\rm c}^{(2)}(p)$. For ansatz 2, the free energy $f_{\rm rs}^{(2)}(\alpha)$ vanishes while within ansatz 1, $f_{\rm rs}^{(1)}(\alpha)>0$. Since the free-energy has to be maximized (see Sect 2.1), the first ansatz has to be preferred to the second one [13]. Consequently, p_0 has to be minimized over the choice of possible distributions Φ and an upper bound to the true value of p_0 is provided by the minimal value of p_0 within the RS variational calculation. We show in Appendix 5.3 that the latter already equals 2/5.

4.1.5 Comments on the variational RS calculation

The main ingredient into our trial variational function is the separation between the effective fields of order one seen by frozen spins and the fields of order $T=1/\beta\to 0$ acting on unfrozen spins. The crucial importance of this separation can be easily seen *a posteriori* with a simple Gaussian ansatz for $P(h)$, which amounts to set B to unity in (26). In this case, one finds $\alpha_{\rm c}\simeq 4.76$ for 3-SAT, a rather high value. For 2-SAT, the situation is even worse: the predicted value for the threshold, $\alpha_{\rm c}\simeq 1.7$ is totally wrong while the correct value $\alpha_{\rm c}=1$ was successfully obtained by optimizing over B. This result is not surprising: slightly above $\alpha_{\rm c}$, there are only few constrained spins whereas the Gaussian ansatz with $B=1$ and $\Delta>0$ amounts to consider that all spins are frozen.

Besides its technical simplicity, the variational calculation provides a better understanding of the transition. While the iterative RS scheme used in [9] was rather involved and the resulting shape of the field distribution remained unclear, the two-parameter variational theory presented here stresses unambiguously the mixed nature of the SAT-UNSAT transition [4]: of first order with respect to the backbone size B and continuous with respect to the intensity Δ of the effective fields related to excited configurations. However, from a quantitative point of view, the predicted value of $B_{\rm c}$ is much larger than the numerical result [4]. This discrepancy stems from an intrinsic weakness of replica symmetry, which is unable to distinguish between different kinds of frozen spins (belonging or not to the backbone). As a result, the parameter B obtained from the variational RS calculation takes into account all frozen variables and thus overestimates the backbone size. We shall see in next section how replica symmetry breaking has to be introduced to solve this problem.

4.2 Replica symmetry breaking calculation

The variational calculation exposed in Section 4.1 as well as the iterative replica symmetric solution of [9,10] provide qualitative insights into the physical features of the SAT-UNSAT transition. From a quantitative point of view, the RS ansatz however fails to predict accurately the threshold

$$f_{\rm rsb}(B_1,B_0,r,\mu) = \frac{1}{\mu}\int_0^{+\infty} {\rm d}x{\rm d}y\; L(x,y)\left[1-B_1-B_0+B_1{\rm e}^{-x}+B_0{\rm e}^{-y}\right]\ln\left[1-B_1-B_0+B_1{\rm e}^{-x}+B_0{\rm e}^{-y}\right]$$
$$-\frac{B_1}{\mu}\ln\left(\int_{-\infty}^{+\infty}{\rm D}x\;{\rm e}^{\,\mu|x|}\right)-\frac{2r}{\sqrt{2\pi}}B_0-\frac{\alpha}{\mu}B_1^K\ln\left(1-\frac{2\mu}{2^K H^K(0)}\int_0^{+\infty}{\rm d}y\;{\rm e}^{-2\mu y}\;H^K(y)\right)$$
$$+2\alpha\sum_{q=0}^{K-1}\binom{q}{K}B_1^q B_0^{K-q}\int_0^{+\infty}{\rm d}x\left(\int_{x/r}^{+\infty}{\rm D}y\right)^{K-q}\frac{{\rm e}^{-2\mu x}H^q(x)}{2^q H(0)-2\mu\int_0^x{\rm d}y\;{\rm e}^{-2\mu y}H^q(y)}, \qquad (45)$$

of 3-SAT: the estimate $\alpha_{\rm c}\simeq 4.6$ lies above numerical findings $\alpha_{\rm c}\simeq 4.25$–$4.30$ [4,6,7]. This by itself indicates that a replica symmetry broken (RSB) theory of K-SAT has to be sought for [9–11]. Another strong hint is of course the appearance of RSB in the ground state structure already in the SAT phase, as discussed in Sections 3.2 and 3.3.

4.2.1 Structure of the RSB field distributions

A major qualitative weakness of the RS ansatz underlined in Section 4.1.5 lies in its inability to distinguish spins frozen always in the same direction (backbone) from spins frozen up and down depending on the particular ground state cluster. For this reason, the RS analysis can only predict that the fraction of frozen spins is ~ 0.93 but does not tell us the size of the subset that truly belongs to the backbone of solutions.

Results on the backbone can be obtained within the replica symmetry breaking analysis. In the latter, the backbone may be defined as the fraction of frozen spins S_i which do not change direction from cluster to cluster. The distribution of effective fields $\rho_i(h)$ of such a spin has its whole support on the positive (or the negative) semi-axis only, see Section 2.3. Conversely, frozen spins that do not belong to the backbone can fluctuate from state to state: their corresponding probability distribution $\rho_i(h)$ may extend over the entire real axis.

On the basis of the previous considerations, we propose the following (one step) RSB variational ansatz,

$$\mathcal{P}[\rho(h)] = (1-B_1-B_0)\;\delta\Big[\rho(h)-\delta(h)\Big] + B_0\int_{-\infty}^{+\infty}{\rm d}\tilde{h}\;\phi_0(\tilde{h})\;\delta\Big[\rho(h)-\psi_0(h,\tilde{h})\Big] + B_1\int_{-\infty}^{+\infty}{\rm d}\tilde{h}\;\phi_1(\tilde{h})\;\delta\Big[\rho(h)-\psi_1(h,\tilde{h})\Big]. \qquad (43)$$

In the above expression, $\delta[.]$ denotes a functional Dirac distribution [11]. The first term on the r.h.s. of (43) is the contribution due to unfrozen spins, whereas the two other terms include two kinds of frozen spins. B_0 gives the fraction of variables in the backbone. $\psi_0(h,\tilde{h})$ denotes the distribution of the effective fields h at one site while fluctuations from site to site are taken into account through $\tilde{h}$ and the distribution $\phi_0(\tilde{h})$ [11,13,26]. Thus, at fixed $\tilde{h}$, the distribution $\psi_0(h,\tilde{h})$ of h has a support on the semi axis having the same sign as $\tilde{h}$. The last term of (43) is associated to frozen spins not belonging to the backbone. The effective field distribution $\psi_1(h,\tilde{h})$ has therefore no *a priori* restriction on the sign of h.

To obtain mathematically tractable expressions, we have made the following choices for the above field distributions:

$$\phi_0(\tilde{h}) = G_{\Delta_0}(h)$$
$$\psi_0(h,\tilde{h}) = \delta(h-\tilde{h})$$
$$\phi_1(\tilde{h}) = \delta(\tilde{h})$$
$$\psi_1(h,\tilde{h}) = G_{\Delta_1}(h-\tilde{h}). \qquad (44)$$

The arbitrary choice for ϕ_1 simply means that close to the transition the typical value of $\tilde{h}$ is much smaller than h's one. Indeed, as in the RS case, effective fields acting on frozen spins are expected to vanish at the transition (coming from the high α phase). Ansatz (44) is the simplest one compatible with the sign restriction on the support of ψ_0 and the unbiased distribution of literals in clauses imposing $P[\rho(h)]=P[\rho(-h)]$.

4.2.2 Analysis of the transition for 3-SAT

The trial variational function (43) with (44) can be plugged into (8) and (3). As the temperature T is sent to zero, the resulting variational problem involves four parameters: B_1, B_0, $r=\sqrt{\Delta_0/\Delta_1}$ and $\mu=\beta m\sqrt{\Delta_1}$. The variances Δ_0 and Δ_1 of the fields vanish at the transition (see Appendix C) and enter the free-energy through the finite ratio $r=\sqrt{\Delta_0/\Delta_1}$. Moreover, at very low temperatures, the breakpoint parameter m naturally behaves as $O(T)$. Since the number of states having an excess free energy F with respect to the lowest lying state scales as ${\rm e}^{\beta m F}$ [13], βm keeps finite when $T=0$ not to spoil RSB effects. Furthermore, to match the SAT phase ($m=O(1)$, *i.e.* $\beta m=\infty$), βm has to diverge at the threshold $\alpha_{\rm c}$ when coming from the RSB-UNSAT phase. This divergence makes $\mu=\beta m\sqrt{\Delta_1}$ finite at the transition.

The variational RSB free-energy is computed in Appendix C and written below,

see equation (45) above

where $L(x,y)$, that depends on r and μ is the double inverse Laplace transform of

$$\mathcal{K}(\mu\sqrt{a},\mu r\sqrt{b}) = \int_0^{+\infty}{\rm d}x{\rm d}y\;{\rm e}^{-xb-ya}\;L(x,y), \qquad (46)$$

with $\mathcal{K}(a,b)$ defined as

$$\mathcal{K}(a,b) = \int_{-\infty}^{+\infty} \mathrm{D}y \ln \left[\int_{-\infty}^{+\infty} \mathrm{D}x \mathrm{e}^{|ax+by|} \right]. \qquad (47)$$

The function $H(x)$ equals

$$H(x) = \int_{x}^{+\infty} \mathrm{D}y \, \mathrm{e}^{\mu y}. \qquad (48)$$

To compute f_{rsb}, we have expanded the logarithm in the first term of the r.h.s. of (45) in powers of B_1 and B_0, using then the definition (46) of L to perform the integrations over x and y. The main difficulty with this procedure is that results obviously depend on the number j of terms considered in the series expansion, see Appendix D. In the simple case $\mu = 0$, we have checked that the optimal (and j-dependent) free-energy $f_{\mathrm{rsb}}(j)$ reaches its exact value f_{rsb} with $1/j^2$ corrections as j grows. In the general case $\mu \neq 0$, numerical results support this scaling: $f_{\mathrm{rsb}}(j) = f_{\mathrm{rsb}} + O(1/j^2)$. For instance we show in the inset of Figure 5 the threshold $\alpha_{\mathrm{c}}(j)$ *versus* $1/j^2$. Using this procedure, we have found that the SAT-UNSAT transition takes place at $\alpha_{\mathrm{c}} \simeq 4.480 \pm 0.003$. This value is still higher than the numerical one, $\alpha_{\mathrm{c}} \simeq 4.25$–$4.30$ [7,4], but definitely improves the RS result $\alpha_{\mathrm{c}} \simeq 4.60$ and lies below the best known rigorous upper bounds [27]. Moreover, at the transition we find $B_0 \simeq 0.13 \pm 0.01$, $B_1 \simeq 0.79 \pm 0.01$, $\mu \simeq 0.88 \pm 0.02$, and $r \simeq 1.4 \pm 0.1$.

What is the meaning of μ? Consider at a given $\alpha > \alpha_{\mathrm{c}}$, the clusters corresponding to the ground state energy E_{GS}, *i.e.* with the minimal number $= O(N)$ of violated clauses. Higher-lying clusters Γ exist which make slightly more mistakes: $E_\Gamma = E_{\mathrm{GS}} + e_\Gamma$ with $e_\Gamma = O(1)$. Imagine now that we add $c = O(1)$ new clauses to the instance we have considered so far. Clearly, the previous ground state assignments are not necessarily optimal any longer and can be supplanted by configurations belonging to some clusters Γ such that $e_\Gamma < c$. Therefore, these clusters and the distribution of the related e_Γ are of interest to understand how an instance of K-SAT can adapt in response to some change in the constraints. From a more quantitative point of view, let us define the linear susceptibility χ_Γ as the (free-)energy change of the cluster Γ when the typical magnitude of the effective fields grows from zero up to $\sqrt{\Delta_1}$, divided by $\sqrt{\Delta_1}$. Within the RSB variational approach, the number of quasi-optimal clusters having a susceptibility equal to χ scales as

$$\mathcal{N}(\chi) \sim \mathrm{e}^{\mu\chi}. \qquad (49)$$

The above equation unveils the meaning of the parameter μ associated with the breaking of replica symmetry.

4.2.3 Comments on the replica symmetry breaking solutions

As stressed in Section 2.4, it is crucial to distinguish fields of the order of one from vanishing fields when $T \to 0$. The importance of this separation for the RSB solution can be checked within the variational subspace $B_1 + B_0 = 1$ in (45), that is discarding unfrozen spins. In this case, we find $\alpha_{\mathrm{c}} \simeq 4.66$ and $B_1 = 1$, $B_0 = 0$. This result is quantitatively and qualitatively erroneous, because the value of α_{c} is even higher than the value predicted within the replica symmetric analysis and the fraction of spins belonging to the backbone is zero. The value of α_{c} can be improved fixing $B_0 = 0$ and optimizing over B_1. In this case we find $\alpha_{\mathrm{c}} \simeq 4.51$, $B_1 \simeq 0.925$, $\mu \simeq 0.8$ and there is no backbone.

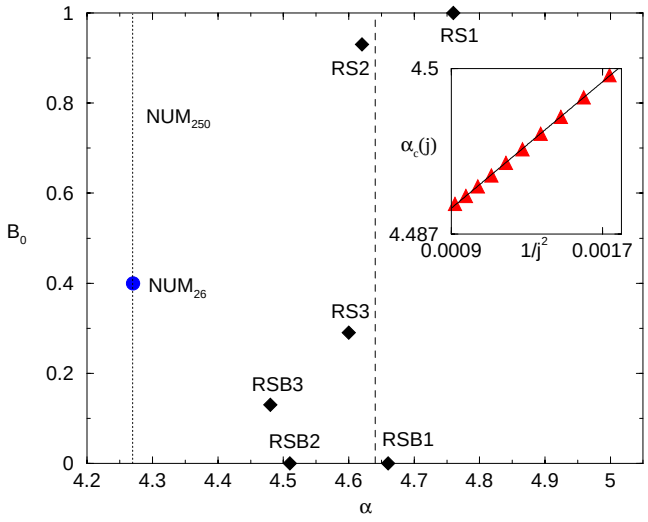

Fig. 5. Values of α_{c} and B_0 at the SAT-UNSAT transition for the different ansätze presented in Section 4. RS1, RS2 and RS3 correspond respectively to the replica symmetric ansätze with one Gaussian, with one Gaussian and a Dirac peak, and with two Gaussians and a Dirac peak. RSB1, RSB2 and RSB3 are their generalizations to the replica symmetry broken case. The dashed line gives the best known rigorous upper bound on the value of α_{c} [27]. The dotted line and the circle respectively show the values of α_{c} and B_0 found in numerical simulations [4]. Note that the value of α_{c} is more reliable than the estimate of B_0 due to the sample sizes used to determine the former ($N = 250$) and the latter ($N = 26$). Inset: scaling of $\alpha_{\mathrm{c}}(j)$ as a function of $1/j^2$ (where j is the number of terms considered in the series expansion of the effective entropy contribution).

The backbone is taken into account when relaxing the constraint $B_0 = 0$. The corresponding variational calculation has been exposed in the previous paragraph. Let us briefly comment on the results. First of all, we note that the fraction of frozen spins $B_1 + B_0 \sim 0.92$ changes by a few percents with respect to the RS case. This value is quite robust and should be quantitatively correct [28]. Conversely, the fraction of spins belonging to the backbone $B_0 \simeq 0.13$ is underestimated with respect to numerical findings [4], which predict a value of 0.4 for small instances. This probably stems from the choices of the field distributions (44) which break replica symmetry for spins not belonging to the backbone only. Therefore in the variational treatment the latter are thermodynamically favoured and the computed fraction of spins belonging to the backbone is smaller than the true one. Breaking replica symmetry also for these spins would presumably

permit to obtain better values for α_c and B_0. This would however be a hard task due to the technical difficulties arising in the numerical computation.

To strengthen this intuition, we consider the $\mu \to 0$ limit of the RSB free-energy (45) which amounts to treating the two kinds of frozen spins on the same footing. The latter becomes simplified and corresponds to the RS free-energy obtained from the following RS field distribution,

$$P(h) = (1 - B_1 - B_0)\,\delta(h) + B_1 \frac{\mathrm{e}^{-h^2/2\Delta_1}}{\sqrt{2\pi\Delta_1}} + B_0 \frac{\mathrm{e}^{-h^2/2\Delta_0}}{\sqrt{2\pi\Delta_0}}\,. \tag{50}$$

Optimizing $f_{\mathrm{rsb}}(B_1, B_0, r, \mu = 0)$, we found a transition at $\alpha_c \simeq 4.60$, which quantitatively coincides with the best know RS solution. The corresponding values of the variational parameters are $B_1 \simeq 0.65$, $B_0 \simeq 0.29$ and $r \simeq 0.49$. Once more, the total fraction of frozen spins is close to 0.94. However, as conjectured, the relative fraction of backbone spins increases drastically by a factor of two with respect to the result of Section 4.2.

The values of α_c and B_0 predicted using the above mentioned ansätze are shown Figure 5 and compared to the results of numerical simulations [7,4] and the rigorous bounds found in [27].

5 Discussion and conclusion

5.1 K-SAT picture arising from the variational calculation

The variational calculations of the last two sections lead us to propose the following picture of the 3-SAT problem. At very low α each variable x_i is under-constrained, *i.e.* both SAT instances which result from fixing x_i either to true or to false are satisfiable with probability one. By adding new clauses, the number α of constraints per variable is increased and the solution space shrinks. The latter is made of a single cluster without any particular internal structure. Its diameter d decreases monotonously with the number of clauses, thus signaling a concentration of the satisfying assignments in configuration space, see Figure 2.

When α reaches $\alpha_s \simeq 3.96$, the set of all solutions continuously breaks up into an exponential number (in N) of geometrically separated clusters, see Figure 6 for a schematic representation. The instance remains nevertheless satisfiable, and the variables are still under-constrained. If we further increase the number αN of clauses, the typical distance d_0 between clusters remains nearly unchanged. The decrease of the entropy of solutions is thus essentially due to the decrease of the average diameter d_1 of the clusters (Fig. 2).

Increasing α the system becomes unsatisfiable with probability one at a certain value α_c, *i.e.* it undergoes a SAT-UNSAT transition. In the optimal assignments (which minimize the number of violated clauses), a large fraction of variables (approximatively 90%) becomes over-constrained. The mixed nature of the SAT-UNSAT transition can be seen explicitly: whereas the fraction of frozen spins jumps up discontinuously, the effective fields measuring the strength of the constraints on each variable grow continuously. Moreover the existence of different clusters of optimal configurations allows the distinction between two groups of over-constrained variables. The first group (backbone) contains variables keeping the same truth value in all optimal configurations. In the second group, the variables have a cluster-dependent value. In other words, optimal configurations corresponding to different thruth values of the second group variables necessarily belong to distinct clusters and lie at $O(N)$ distances from each other.

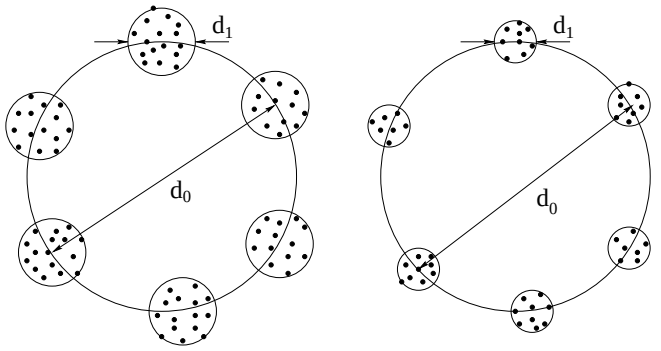

Fig. 6. Schematic representation of the solution space structure of 3-SAT for two values of α with $\alpha_s < \alpha < \alpha_c$ resulting from the one-step replica symmetry broken variational ansatz. The solutions (represented by dots) are organized in clusters. Whereas the distance d_0 between the clusters remain almost unchanged as α is increased by adding new clauses, the cluster size d_1 decreases quickly.

It is important to note that even in the UNSAT regime these frozen spins coexist with under-constrained variables. These unfrozen variables lead to a positive entropy at the transition [28], a behavior which is intrinsically different from the case of infinite connectivity models [29].

Besides, we should mention that the actual cluster distribution could be even more complicated, *e.g.* through the existence of clusters of clusters etc. The existence of only two typical distances, and thus the distinction between two kinds of frozen spins is intrinsic to our one-step replica symmetry broken ansatz. We nevertheless expect that the main qualitative features of 3-SAT are already captured in our one-step broken variational description.

Finally, we note that in the $2 + p$-SAT case for $p < p_0 = 2/5$ the picture arising from the variational calculation is much simpler. In fact, α_s and α_c coincide and the transition from the under-constrained SAT regime to the over-constrained UNSAT phase is smooth. The geometrically non-trivial intermediate phase does not exist at all.

5.2 Critical behavior and exponents

The exact RS saddle-point equation [9] shows that the probability $P(h)$ that the effective field equals $h \gg 1$ on a given site is bounded from above by the probability that this site is connected to h neighbours and then decreases at least exponentially with h. Combining this observation

with the variational calculation presented in this paper, an investigation of the optimization equations over B and H reveals that slightly above the threshold, the free-energy exhibits a singularity of the type

$$F_{\rm rs}(\delta\alpha) \sim N\delta\alpha\,(-\log\delta\alpha)^{-\eta} \qquad (\delta\alpha = \alpha - \alpha_{\rm c} \geq 0). \tag{51}$$

The actual scaling of $\Phi(h)$ at large fields h gives only rise to logarithmic singularities, *e.g.* $\eta = \frac{1}{2}$ for a Gaussian distribution, $\eta = 1$ for an exponential one. Note that equation (51) also holds at the RSB level within the ansatz (43).

These predictions can be related to the recent finite-size scaling (FSS) numerical studies of the K-SAT model [7,4]. Let us call $E_{\rm GS}(\alpha, N)$ the average ground state energy for a finite number N of Boolean variables. Close to the threshold, we expect the curves of $E_{\rm GS}$ as functions of α obtained for different sizes N to collapse onto each other when properly rescaled. In other word, FSS should hold and there should exist some exponents ν and γ such that

$$E_{\rm GS}(\alpha, N) \simeq N^{\gamma}\ \mathcal{E}\left(N^{1/\nu}\ \delta\alpha\right), \tag{52}$$

when $\delta\alpha = \alpha - \alpha_{\rm c} \ll 1$ and $N \gg 1$. ν characterizes the smaller and smaller width of the transition region from the SAT to the UNSAT phase as N grows. It has been numerically calculated using a Davis-Putnam procedure [7]: $\nu \simeq 3$ for 2+p-SAT as long as $p < p_0$ and ν decreases for larger p down to $\simeq 1.5$ for 3-SAT. γ can be simply interpreted: N^{γ} is the minimal number of violated clauses at threshold and thus $\gamma < 1$. The rescaled ground state energy $\mathcal{E}(y)$ is a monotonously increasing function of its argument: $\mathcal{E}(y) \to 0$ when $y \to -\infty$ (right boundary of the SAT phase); $\mathcal{E}(0)$ is finite; $\mathcal{E}(y) \sim y$ when $y \to \infty$ (left boundary of the UNSAT phase). The latter scaling ensures that $E_{\rm GS}$ grows above the threshold (that is at fixed $\delta\alpha$ while N becomes larger and larger) as

$$E_{\rm GS}(\alpha, N) \simeq N^{\gamma}\ N^{1/\nu}\ \delta\alpha, \tag{53}$$

to coincide with (51) up to logarithmic singularities. Imposing that $E_{\rm GS}(\alpha, N) = O(N)$ in the UNSAT regime, identity (53) gives the hyperscaling relation

$$\gamma = 1 - \frac{1}{\nu}\cdot \tag{54}$$

Whereas ν may be computed for large formulae, involving thousands of variables, no such powerful method exists so far to estimate γ. Therefore, identity (54) may be precious to derive indirectly γ from the knowledge of ν.

5.3 Perspectives

As discussed in Section 2.5, the threshold $\alpha_{\rm c}$ separates $O(T)$ fields (SAT regime) from $O(1)$ ones (UNSAT phase). This change of scaling of effective fields is nicely apparent within the variational calculation presented in Sections 3.1, 3.2, 4.1, and 4.2. Looking at compatible ansätze for the SAT and the UNSAT phase, the same values for $\alpha_{\rm c}$ can be obtained either starting from the SAT phase from a diverging renormalized variance of the effective fields (which were assumed to vanish linearly with T), or coming from the UNSAT phase from a vanishing variance of the effective fields (which remained finite in the limit $T \to 0$). A deeper understanding of the scaling of the fields in the vicinity of the critical point $\alpha = \alpha_{\rm c}$, $T = 0$ would be of interest for at least two reasons. First, it would allow the calculation of the entropy in the UNSAT phase, which has been out of reach yet. Secondly, the structure of the invariant measure $P(h)$ could be studied carefully to gain some information on its singularities, its fractal structure, etc. [30]. At finite but low temperatures, one might expect that the support of $P(h)$ includes different regions corresponding to different scalings with T, *i.e.* to distinct physical phenomena coexisting in the model. This potential richness of the order parameter cannot be present in infinite-connectivity spin glasses and could give rise to new properties at the mean-field level.

Further work is clearly required to confirm the geometrical picture of the space of solutions sketched in Section 5.1. From a numerical point of view, an analysis of the distances between solutions would be of interest to check the existence of a non trivial (non necessarily bimodal) distribution for d. It would be also worth trying to improve our analytical approach by using richer trial field distributions in the RSB calculations. However, the most promising route is probably to attempt to use the information presently available on the optimal (and quasi-optimal, see Sect. 4.2) assignments of K-SAT to understand the drastic change of behaviour of algorithms close to the threshold.

We are grateful to O. Dubois, S. Kirkpatrick and R. Zecchina for useful discussions, and J. Berg for carefully reading the manuscript. M.W. acknowledges financial support by the German Academic Exchange Service (DAAD).

Note added in proof

After submission of this paper we have been aware of a new rigorous upper bound for 3-SAT, $\alpha_{\rm c} \leq 4.506$ proven by O. Dubois. This upper bound lies slightly above our RSB result.

Appendix A: RSB free-energy in the SAT phase

In this appendix we show how the one-step replica symmetry broken ground state entropy is calculated. We start from (3), perform the limit $\beta \to \infty$:

$$s = \lim_{n\to 0} -\frac{1}{n}\sum_{\boldsymbol{\sigma}} c(\boldsymbol{\sigma})\ln c(\boldsymbol{\sigma}) + \frac{\alpha}{n}\ln\left[\sum_{\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_K} c(\boldsymbol{\sigma}_1)\ldots c(\boldsymbol{\sigma}_K)\prod_{a=1}^{n}\left(1 - \prod_{a=1}^{n}\delta_{\sigma_l^a,1}\right)\right] \tag{A.1}$$

and plug in ansatz (13), that is

$$c(\boldsymbol{\sigma}) = \int_{-\infty}^{\infty} \mathrm{d}z\ G_{\Delta_0}(z) \times \prod_{a=1}^{n/m} \frac{\int \mathrm{d}\tilde{z}\ G_{\Delta_1}(\tilde{z}-z) \exp\left\{\tilde{z}\sum_{b=(a-1)m+1}^{am}\sigma^b\right\}}{\int \mathrm{d}\tilde{z}\ G_{\Delta_1}(\tilde{z}-z)(2\cosh\tilde{z})^m}. \quad \text{(A.2)}$$

We will calculate both terms on the r.h.s. of (A.1) separately. We start with the effective entropic term and follow closely the analytical continuation scheme proposed in [11],

$$\begin{aligned} -\lim_{n\to 0}\frac{1}{n}\sum_{\boldsymbol{\sigma}} c(\boldsymbol{\sigma})\ln c(\boldsymbol{\sigma}) = \\ -\int \mathcal{D}\hat{\nu}\mathcal{D}\nu\ \exp\left\{-\mathrm{i}\int_{-\infty}^{\infty}\mathrm{d}y\ \hat{\nu}(y)\ \nu(y)\right\} c[\mathrm{i}\nu]\ln c[\mathrm{i}\nu] \\ \times \frac{1}{m}\ln\left\{\int_{-\infty}^{\infty}\frac{\mathrm{d}x\ \mathrm{d}y}{2\pi}\mathrm{e}^{-\mathrm{i}xy}(2\cosh x)^m \exp\hat{\nu}(\mathrm{i}y)\right\} \end{aligned} \quad \text{(A.3)}$$

with

$$c[\nu] = \int \mathcal{D}\rho\mathcal{P}[\rho]\exp\left\{\int_{-\infty}^{\infty}\mathrm{d}y\ \nu(y) \times \ln\left[\int_{-\infty}^{\infty}\mathrm{d}h\ \rho(h)\frac{\mathrm{e}^{\beta h y}}{(2\cosh\beta h)^m}\right]\right\}. \quad \text{(A.4)}$$

Note that this form does not depend on K, *i.e.* on the length of the clauses, and the following calculations are consequently valid for any K. With ansatz (13), the last expression depends on $\nu(y)$ only through its first three moments,

$$c[\nu] = c(\nu_0,\nu_1,\nu_2) = \int_{-\infty}^{\infty}\mathrm{d}z\ G_{\Delta_0}(z)\exp\left\{-\nu_0 \times \ln\int_{-\infty}^{\infty}\mathrm{d}\tilde{z}\ G_{\Delta_1}(\tilde{z}-z)(2\cosh\tilde{z})^m + z\nu_1 + \Delta_1\nu_2\right\} \quad \text{(A.5)}$$

with $\nu_l = \int_{-\infty}^{\infty}\mathrm{d}y\ \nu(y)\ y^l/l!$ $(l=0,1,2)$. The effective entropic part can now be calculated according to (A.3) if we introduce the series expansion $\hat{\nu}(y) = \sum_{l=0}^{\infty}\nu_l y^l/l!$. Using $\int \mathrm{d}y\ \nu(y)\hat{\nu}(y) = \sum_{l=0}^{\infty}\nu_l\hat{\nu}_l$ the integrals over the ν_l and $\hat{\nu}_l$ with $l\geq 3$ can be executed trivially; ν_l does not exist outside the first exponential leading to a Dirac-function in $\hat{\nu}_l$, which vanishes consequently. Thus we obtain

$$\begin{aligned} -\lim_{n\to 0}\frac{1}{n}\sum_{\boldsymbol{\sigma}} c(\boldsymbol{\sigma})\ln c(\boldsymbol{\sigma}) = \\ -\int\prod_{l=0}^{2}\left(\frac{\mathrm{d}\nu_l\ \mathrm{d}\hat{\nu}_l}{2\pi}\mathrm{e}^{-\mathrm{i}\nu_l\hat{\nu}_l}\right) c(\mathrm{i}\nu_0,\mathrm{i}\nu_1,\mathrm{i}\nu_2)\ln c(\mathrm{i}\nu_0,\mathrm{i}\nu_1,\mathrm{i}\nu_2)\frac{1}{m} \\ \times\ln\left[\int\frac{\mathrm{d}x\ \mathrm{d}y}{2\pi}\mathrm{e}^{-\mathrm{i}xy}(2\cosh x)^m\exp\{\hat{\nu}_0+\mathrm{i}\hat{\nu}_1 y-\frac{1}{2}\hat{\nu}_2 y^2\}\right] \\ = S_0 + S_1, \end{aligned} \quad \text{(A.6)}$$

where

$$\begin{aligned} S_0 &= -\frac{1}{m}\int\frac{\mathrm{d}\nu_0\ \mathrm{d}\hat{\nu}_0}{2\pi}\mathrm{e}^{-\mathrm{i}\nu_0\hat{\nu}_0}\ \hat{\nu}_0\ c(\mathrm{i}\nu_0,0,0)\ln c(\mathrm{i}\nu_0,0,0) \\ &= +\frac{1}{m}\int\mathrm{d}z\ G_{\Delta_0}(z)\ \ln\int\mathrm{d}\tilde{z}\ G_{\Delta_1}(\tilde{z}-z)\,(2\cosh\tilde{z})^m \end{aligned} \quad \text{(A.7)}$$

and

$$\begin{aligned} S_1 &= -\frac{1}{m}\int\frac{\mathrm{d}\nu_1\ \mathrm{d}\hat{\nu}_1\ \mathrm{d}\nu_2\ \mathrm{d}\hat{\nu}_2}{(2\pi)^2}\mathrm{e}^{-\mathrm{i}(\nu_1\hat{\nu}_1+\nu_2\hat{\nu}_2)} \\ &\quad\times c(0,\mathrm{i}\nu_1,\mathrm{i}\nu_2)\ln c(0,\mathrm{i}\nu_1,\mathrm{i}\nu_2) \\ &\quad\times\ln\left[\int\frac{\mathrm{d}x\ \mathrm{d}y}{2\pi}\mathrm{e}^{-\mathrm{i}xy}(2\cosh x)^m\exp\{\mathrm{i}\hat{\nu}_1 y-\frac{1}{2}\hat{\nu}_2 y^2\}\right] \\ &= -\frac{1}{m}\left(\Delta_0\frac{\partial}{\partial\Delta_0}+\Delta_1\frac{\partial}{\partial\Delta_1}\right)\int\mathrm{d}z\ G_{\Delta_0}(z) \\ &\quad\times\ln\int\mathrm{d}\tilde{z}\ G_{\Delta_1}(\tilde{z}-z)\,(2\cosh\tilde{z})^m. \end{aligned} \quad \text{(A.8)}$$

By calculating the derivatives and rescaling the integration variables we finally find the corresponding contribution in (14).

Let us now calculate the effective energy, *i.e.* the explicitly α-dependent contribution in (14) starting from the last term in (A.1) by plugging in (8). The sum over the replicated spin variables can easily be carried out. This directly gives

$$E = -\lim_{n\to 0}\frac{\alpha}{n}\ln\left[\int\prod_{l=1}^{K}\left(\mathrm{d}z_l\ G_{\Delta_0}(z_l)\right) \times\left(\frac{\int\prod_{l=1}^{K}\left(\mathrm{d}\tilde{z}_l\ G_{\Delta_0}(\tilde{z}_l-z_l)\right)\left(\prod_l 2\cosh\tilde{z}_l-\prod_l\mathrm{e}^{\tilde{z}_l}\right)^m}{\int\prod_{l=1}^{K}(\mathrm{d}\tilde{z}_l\ G_{\Delta_0}(\tilde{z}_l-z_l)(2\cosh\tilde{z}_l)^m)}\right)^{\frac{n}{m}}\right]. \quad \text{(A.9)}$$

In the limit $n\to 0$ and after a rescaling and translation of the integration variables to normally distributed Gaussian variables, we find the corresponding expression in (14).

Appendix B: RS Gaussian ansatz for the SAT-UNSAT transition

In this appendix, we compute the replica symmetric variational free-energy for a Gaussian distribution $\Phi(x) = G_1(x)$. Using (26), we straightforwardly obtain for the K-SAT problem

$$\begin{aligned} f_{\mathrm{rs}}^{\mathrm{Gauss}}(B,\Delta,\alpha,K) = 2B\sqrt{\Delta} \\ \times\left\{\int_{-\infty}^{\infty}\frac{\mathrm{d}x}{2\pi}\mathrm{e}^{-x^2/2}\ln\left(1-B+B\ \mathrm{e}^{-x^2/2}\right)\right. \\ \left.+\alpha\ B^{K-1}\int_0^{1/(2\sqrt{\Delta})}\mathrm{d}x\left[\int_x^{\infty}\mathrm{D}y\right]^K\right\}. \end{aligned} \quad \text{(A.10)}$$

The corresponding free-energy for the $2+p$-SAT model can be easily obtained by a linear combination of expression (A.10) for $K=2$ (with weight $1-p$) and $K=3$ (with weight p).

However for completeness we give a derivation of free-energy (A.10) from (3) without any reference to [11]. The order parameter $c(\boldsymbol{\sigma})$ reduces to

$$c(\boldsymbol{\sigma}) = \frac{1-B}{2^n} + B\int_{-\infty}^{+\infty} \mathrm{D}x \prod_{a=1}^{n} \frac{\mathrm{e}^{\beta\sqrt{\Delta}x\sigma^a}}{2\cosh\beta\sqrt{\Delta}x}\cdot \quad \text{(A.11)}$$

The first term on the r.h.s. of (3) may then be written in the limit $n\to 0$ and $\beta\to\infty$ as

$$\frac{1}{\beta n}\sum_{\boldsymbol{\sigma}} c(\boldsymbol{\sigma})\ln c(\boldsymbol{\sigma}) = \frac{1}{\beta n}\frac{\partial}{\partial l}\sum_{\boldsymbol{\sigma}} c(\boldsymbol{\sigma})^l\big|_{l=1} = -\frac{\sqrt{\Delta}}{\sqrt{\pi}}\frac{\partial}{\partial l}\left[\sum_{j=0}^{l}\binom{l}{j}(1-B)^{l-j}B^j(j-\sqrt{j})\right]\Bigg|_{l=1}. \quad \text{(A.12)}$$

The difficulty in the computation of the entropic contribution is the analytic continuation in l. For the r.h.s. of (A.12), this can be easily achieved through the relation

$$j-\sqrt{j} = j\left(1-\int_{-\infty}^{+\infty}\frac{\mathrm{d}x}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2 j}\right). \quad \text{(A.13)}$$

Using (A.13), the sum over l in (A.12) can be carried out and expression (A.10) recovered. The last term of (5) is found when inserting the RS expression of $c(\boldsymbol{\sigma})$ in (3) and dividing by n. Using the variational expression (4) of $P(h)$, this term reduces to

$$\alpha B^K\sqrt{\Delta}\int_0^{+\infty}\mathrm{D}x_1\ldots\mathrm{D}x_K\,\min\left[1/\sqrt{\Delta},2x_1,\ldots,2x_K\right]$$

when $\beta\to\infty$. It is easy to check that the above expression coincides with the last term of (A.10).

Appendix C: Variational upper bound to p_0

For a given distribution Φ, the tricritical point p_0 is obtained through the condition $f_{\mathrm{rs}}^{(3)}(1/(1-p_0),p_0)=0$. Using (41), we obtain

$$p_0[\Phi] = \frac{A[\Phi]}{A[\Phi]+B[\Phi]}, \quad \text{(A.14)}$$

where

$$A[\Phi] = -\frac{1}{2\pi}\int_0^{+\infty}\frac{\mathrm{d}\nu}{\nu}\,\Phi'_{\mathrm{ft}}(\nu)[\Phi_{\mathrm{ft}}(\nu)-1]^2 \quad \text{(A.15)}$$

$$B[\Phi] = \int_0^{\infty}\mathrm{d}h\,[\Phi_{\mathrm{cc}}(h)]^3. \quad \text{(A.16)}$$

The extremization condition of p_0 with respect to the even distribution Φ may thus be written as

$$\frac{\delta p_0}{\delta\Phi(x)}[\Phi] = \lambda, \qquad \forall x\geq 0, \quad \text{(A.17)}$$

where λ is a Lagrange multiplier ensuring the normalization of Φ. Equation (A.17) involves the functional derivatives of A and B,

$$\frac{\delta A}{\delta\Phi(x)}[\Phi] = 2\,x\int_{-\infty}^{\infty}\mathrm{d}y\,\Phi(y)\big[\Phi_{\mathrm{cc}}(x+y)-\Phi_{\mathrm{cc}}(x)\big] + 2\int_{-\infty}^{\infty}\mathrm{d}y\,y\,\Phi(y)\big[\Phi_{\mathrm{cc}}(x-y)-\theta(y-x)\big] \quad \text{(A.18)}$$

$$\frac{\delta B}{\delta\Phi(x)}[\Phi] = 3\int_0^x \mathrm{d}y\,\big[\Phi_{\mathrm{cc}}(y)\big]^2, \quad \text{(A.19)}$$

where $\theta(\cdot)$ denotes the Heaviside function. By subtracting the values of the functional derivatives of p_0 (A.17) in $x=0$ and $x=\infty$, the Lagrange multiplier λ disappears. We obtain $B[\Phi]/A[\Phi]=3/2$ and therefore from (A.14),

$$\min_{\Phi}\; p_0[\Phi] = \frac{2}{5}\cdot \quad \text{(A.20)}$$

The determination of the optimal distribution Φ, although of interest, see Section 5.2 would be more difficult. Note that we have implicitely assumed in the functional differentiations (A.17, A.18, A.19) that Φ included no Dirac distributions. The value of $p_0=2/5$ directly comes from this hypothesis as shown in [10].

Appendix D: RSB free-energy for the SAT-UNSAT transition

Within the RSB ansatz (43), the order parameter $c(\boldsymbol{\sigma})$ reduces to:

$$c(\boldsymbol{\sigma}) = \frac{1-B_1-B_0}{2^n} + B_1\prod_{b=1}^{\frac{n}{m}}\frac{\displaystyle\int_{-\infty}^{+\infty}\mathrm{D}h\,\mathrm{e}^{\beta h\sqrt{\Delta_1}s_b}}{\displaystyle\int_{-\infty}^{+\infty}\mathrm{D}h(2\cosh\beta h\sqrt{\Delta_1})^m} + B_0\int_{-\infty}^{+\infty}\mathrm{D}h\prod_{a=1}^{n}\frac{\mathrm{e}^{\beta h\sigma^a\sqrt{\Delta_0}}}{2\cosh\beta h\sqrt{\Delta_0}}\cdot \quad \text{(A.21)}$$

In the following, we compute the effective entropy contribution by taking the derivative of $\sum_{\boldsymbol{\sigma}} c(\boldsymbol{\sigma})^l$ with respect

to l, see (A.12). For $\beta \to \infty$, we find

$$S = \frac{\sqrt{\Delta_1}}{\mu}\frac{\partial}{\partial l}\left\{\sum_{p=0}^{l}\binom{l}{p}(1-B_1-B_0)^{l-p} \times \sum_{q=0}^{p}\binom{p}{q}B_1^q B_0^{p-q}\int_{-\infty}^{+\infty}\mathrm{D}h_{q+1}\ldots\mathrm{D}h_p \times\left[\ln\left(\int_{-\infty}^{+\infty}\mathrm{D}h_1\ldots\mathrm{D}h_q \mathrm{e}^{\,\mu|(h_1+\cdots h_q)+(h_{q+1}+\cdots h_p)r|}\right) - q\ln\left(\int_{-\infty}^{+\infty}\mathrm{D}h\ \mathrm{e}^{2\mu|h|}\right) - \frac{\mu}{\sqrt{\Delta_1}}(p-q)\int_{-\infty}^{+\infty}\mathrm{D}h|h|\sqrt{\Delta_0}\right]\right\}\Bigg|_{l=1}, \tag{A.22}$$

where $\mu = m\beta\sqrt{\Delta_1}$ and $r = \sqrt{\Delta_0/\Delta_1}$. As in the replica-symmetric computation of Appendix B, the main difficulty is the analytic continuation in l. For the last two terms in (A.21) the sum over l can be performed explicitly, therefore the analytic continuation can be trivially performed. It is easy to check that the contribution to S of these two terms leads to the second and the third terms of (45) multiplied by $\sqrt{\Delta_1}$. For the first term of (A.22) the analytic continuation in l is more tricky. First of all using the convolution properties of Gaussian functions this term is reduced to

$$S_I = \frac{\sqrt{\Delta_1}}{\mu}\frac{\partial}{\partial l}\left\{\sum_{p=0}^{l}\binom{l}{p}(1-B_1-B_0)^{l-p} \times\sum_{q=0}^{p}\binom{p}{q}B_1^q B_0^{p-q}\mathcal{K}(\mu\sqrt{q},\mu r\sqrt{p-q})\right\}\Bigg|_{l=1}, \tag{A.23}$$

where the function $\mathcal{K}(a,b)$ has been defined in (47). Then the analytic continuation in l can be achieved using the function $L(x,y)$ defined in (46, 47) and the first term of (45) (multiplied by $\sqrt{\Delta_1}$) is recovered. Although written in a compact way, the resulting S_I is not very useful for numerical purposes. We have rather use the equivalent expression

$$S_I = -\frac{\sqrt{\Delta_1}}{\mu}\sum_{l=1}^{+\infty}\frac{1}{l}\sum_{p=0}^{l}(-1)^p\binom{l}{p}(B_1+B_0)^{l-p} \times\sum_{q=0}^{p}\binom{p}{q}B_1^q B_0^{p-q}\Big[B_1\mathcal{K}(\mu\sqrt{q+1},\mu r\sqrt{p-q}) + B_0\mathcal{K}(\mu\sqrt{q},\mu r\sqrt{p-q+1}) + (1-B_1-B_0)\mathcal{K}(\mu\sqrt{q},\mu r\sqrt{p-q})\Big]. \tag{A.24}$$

We now turn to the effective energy contribution which reads [11],

$$E = -\frac{\alpha}{\beta}\int_V \mathcal{D}\rho_1\ldots\mathcal{D}\rho_K\ \mathcal{P}[\rho_1]\ldots\mathcal{P}[\rho_K] \times\frac{1}{m}\ln\left[\int_{-\infty}^{+\infty}\mathrm{d}h_1\ldots\mathrm{d}h_K\ \rho_1(h_1)\ldots\rho_K(h_K) \times\left(1+(\mathrm{e}^{-\beta}-1)\frac{\mathrm{e}^{\beta\sum_{j=1}^{K}h_j}}{\prod_{j=1}^{K}2\cosh\beta h_j}\right)^m\right]. \tag{A.25}$$

Plugging the trial variational functional (43) into (A.25), we find for $\beta\to\infty$,

$$E = -\frac{\alpha\sqrt{\Delta_1}}{\mu}\sum_{q=0}^{K}\binom{K}{q}B_1^q B_0^{K-q} \times\int_0^{+\infty}\mathrm{D}h_{q+1}\ldots\mathrm{D}h_K\left\{-q\ln\int_{-\infty}^{+\infty}\mathrm{D}h\mathrm{e}^{\mu|h|} + \ln\left[\int_{-\infty}^{+\infty}\mathrm{D}h_1\mathrm{e}^{\mu|h_1|}\ldots\mathrm{D}h_q\mathrm{e}^{\mu|h_q|}\left(1+\prod_{i=1}^{q}\theta(h_i) \times\left[\mathrm{e}^{-\mu\ \min(1/\sqrt{\Delta_1},2h_1,\ldots,2h_q,2rh_{q+1},\ldots,2rh_K)}-1\right]\right)\right]\right\}, \tag{A.26}$$

where $\theta(h)$ is the Heaviside function. First of all we focus on the $q=K$ term, which is the only one that does not vanish for $B_1=1$. In this case a simple integration by parts leads to

$$-\frac{\alpha B_1^K\sqrt{\Delta_1}}{\mu}\ln\left(1-\frac{2\mu}{2^K H^K(0)}\int_0^{1/(2\sqrt{\Delta_1})}\mathrm{e}^{-2\mu h}H^K(h)\right), \tag{A.27}$$

where the function H has been defined in (48). The other terms in the sum over q lead to two different contributions. For $h_{q+1},\ldots,h_K > 1/2\sqrt{\Delta_0}$ by an integration by parts the integrals in the qth term reduce to:

$$\left(\int_{\frac{1}{2\sqrt{\Delta_0}}}^{+\infty}\mathrm{D}h\right)^{K-q}\ln\left(1-\frac{2\mu}{2^q H^q(0)}\int_0^{\frac{1}{2\sqrt{\Delta_1}}}\mathrm{e}^{-2\mu h}H^q(h)\right). \tag{A.28}$$

On the other hand if there is at least one field among $h_{q+1},\ldots,h_K$ which is smaller than $1/2\sqrt{\Delta_0}$. After a little algebra we find that the integrals in the qth term reduce to

$$(K-q)\int_0^{\frac{1}{2\sqrt{\Delta_0}}}\mathrm{D}h\left(\int_h^{+\infty}\mathrm{D}h\right)^{K-q-1} \times\ln\left[1-\frac{2\mu}{2^q H^q(0)}\int_0^{rh}\mathrm{d}h'\mathrm{e}^{-2\mu h'}H^q(h')\right]. \tag{A.29}$$

Collecting (A.28) and (A.29) and integrating by parts, we find that the sum over q (for $q \neq K$) reduces to

$$2\alpha\sqrt{\Delta_1}\sum_{q=0}^{K-1}\binom{K}{q}B_1^q B_0^{K-q}\int_0^{\frac{1}{2\sqrt{\Delta_0}}}\mathrm{d}h\left(\int_{h/r}^{+\infty}\mathrm{D}h'\right)^{K-q} \times\frac{\mathrm{e}^{-2\mu h}H^q(h)}{2^q H(0)-2\mu\int_0^h \mathrm{d}h'\mathrm{e}^{-2\mu h'}H^q(h')}\cdot \quad \text{(A.30)}$$

Finally gathering (A.30) and (A.27) one obtain the final form of the effective energy part.

It is easy to verify that both Δ_1 and Δ_0 vanish at the transition while the ratio $r=\sqrt{\Delta_0/\Delta_1}$ has a non trivial value. Dividing the variational free energy by $\sqrt{\Delta_1}$ the entropic contribution depends on Δ_1 and Δ_0 through r alone. Therefore as in the replica symmetric analysis the optimization on Δ_1 has to be performed for the energetic contribution only. This procedure leads to $\Delta_1=\Delta_0=0$ at the transition. As a consequence within the variational approach the analysis of the SAT-UNSAT transition reduces to the study of the variational free energy (45).

References

1. P. Cheeseman, B. Kanefsky, W.M. Taylor, in *Proc. 13th Int. Joint Conf. Artif. Intell. (IJCAI-91)*, edited by J. Mylopoulos, K. Reiter, (Morgan Kaufmann, San Mateo, California, 1991), p. 331; D. Mitchell, B. Selman, H. Levesque, in *Proc. 10th Natl. Conf. on Artif. Intell. (AAAI-92)*, (AAAI Press/MIT Press, Cambridge, Massachusetts, 1992), p. 440; *Frontiers in problem solving: phase transitions and complexity*, edited by T. Hogg, B.A. Huberman, C. Williams, *Artif. Intell.* **81** (I+II) (1996).
2. M.R. Garey, D.S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, 1979).
3. C. Papadimitriou, *Computational Complexity*, (Addison-Wesley, Readings, 1994).
4. R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, L. Troyansky, Nature **400**, 133 (1999).
5. A. Goerdt, in *Proc. 7th Int. Symp. on Mathematical Foundations of Computer Science* (1992), p. 264; V. Chantal, B. Reed, in *Proc. 33rd IEEE Symp. on Foundations of Computer Science* (IEEE Comp. Soc. Press, New York, 1992), p. 620.
6. J.M. Crawford, L.D. Auton, in *Proc. 11th Natl. Conf. on Artif. Intell. (AAAI-93)* (AAAI Press, Menlo Park, California, 1993), p. 21.
7. B. Selman, S. Kirkpatrick, Science **264**, 1297 (1994).
8. C. De Dominicis, P. Mottishaw, J. Phys. A **20**, L375 (1987).
9. R. Monasson, R. Zecchina, Phys. Rev. Lett. **76**, 3881 (1996); Phys. Rev. E **56**, 1357 (1997).
10. R. Monasson, R. Zecchina, J. Phys. A **31**, 9209 (1998).
11. R. Monasson, J. Phys. A **31**, 513 (1998).
12. A.Z, Broder, A.M. Frieze, E. Upfal, in *Proc. 4th Annual ACM–SIAM Symp. on Discrete Algorithms*, 322 (1993).
13. M. Mézard, G. Parisi, M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
14. This is a general problem, which arises in the study of disordered spin systems with finite connectivity. So far, an exact replica symmetry breaking solution has been found for the diluted p-spin model when $p \to +\infty$ only [8].
15. E.I. Shaknovich, A.M. Gutin, Europhys. Lett. **8**, 327 (1991); M. Mézard, G. Parisi J. Phys. A **23**, L-1229 (1990); J. Phys. I France **1**, 809 (1991).
16. E. Gardner, Nucl. Phys. B **257**, 747 (1985); A. Crisanti, H.J. Sommers, Z. Phys. B **87**, 341 (1992).
17. P. Svenson, M. G. Nordahl, Phys. Rev. E **59**, 3983 (1999).
18. L.F. Cugliandolo, J. Kurchan, Phys. Rev. Lett. **71**, 173 (1993); J.P. Bouchaud, L.F. Cugliandolo, J. Kurchan, M. Mézard, in *Spin Glasses and Random Fields*, edited by A.P. Young (World Scientific, Singapore, 1998).
19. The sixfold numerical integration is somewhat subtle. Even if the integrand shows an asymtotically exponential behavior we have used the Gauss-Hermite quadrature. This works well only if $\sqrt{\Delta_1}\tilde{h}$ is small compared to $\tilde{h}^2$ outside the region covered by the abscissas used in the quadrature. However, this is fulfilled in our case because the interesting values of Δ_1 are of order one. We have cross-checked our results by using a Monte-Carlo integration and a series expansion of the integrand allowing a reduction of the number of integrations to 2 at the cost of a double series. In the case of $\partial_m s_{\mathrm{1RSB}}(m=1,\Delta_0=\Delta_{\mathrm{RS}},\Delta_1)$ the results where precise up to $\Delta_1 \approx 20$, then a systematic deviation showed up.
20. R. Monasson, Phys. Rev. Lett. **75**, 2847 (1995).
21. S. Franz, G. Parisi, J. Phys. I France **5**, 1401 (1995).
22. R. Monasson, D. O'Kane, Europhys. Lett. **27**, 85 (1994); S. Cocco, Tesi di Laurea, Roma (1995); S. Cocco, R. Monasson, R. Zecchina, Phys. Rev. E **54**, 717 (1996); M. Weigt, A. Engel, Phys. Rev. E **55**, 4552 (1997).
23. Th.M. Nieuwenhuizen, J. Phys. I France **4**, 1819 (1996); M.E. Ferrero, G. Parisi, P. Ranieri, J. Phys. A **29**, L569 (1996).
24. T. Temesvari, C. de Dominicis, I. Kondor, J. Phys. A **27**, 7569 (1994).
25. D. Achilioptas, L.M. Kirousis, E. Kranakis, D. Krinzac, in *Proc. of RALCOM 97* (1997), p. 1.
26. K.Y.M. Wong, D. Sherrington, J. Phys. A **21**, L459 (1988).
27. L. Kirousis, E. Kranakis, D. Krizanc, in *Proc. of the 4th European Symposium on Algorithms*, (1992), p. 27–38; O. Dubois, Y. Boufkhad, J. Algorithms **24**, 395–420 (1997).
28. The ground state entropy is finite due to the existence of under-contrained spins, *i.e.* within our approach it is bounded at α_{c} by $(1-B_0-B_1)\ln 2 \approx 0.055$ from above. This value coincides within our numerical precision with the SAT entropy $s_{\mathrm{rsb}}(\alpha=4.48)\approx 0.054$, but is smaller than the entropy $s_{\mathrm{rsb}} \approx 0.073$–$0.081$ at the numerical estimates for $\alpha_{\mathrm{c}}=4.25$–4.3. The latter entropy values would require at least 10–12% of under-constrained spins.
29. In fact, a zero-entropy criterion is frequently used in discrete infinite connectivity models in order to locate $T=0$ phase transitions, *e.g.* in the case of the learning behavior of binary neural networks [31].
30. U. Behn, J.L. van Hemmen, R. Kühn, A. Lange, V.A. Zagrebnov, Physica D **68**, 401 (1993).
31. W. Krauth, M. Mézard, J. Phys. France **50**, 3059 (1989); G. Györgyi, Phys. Rev. Lett. **64**, 2957 (1990).