Threshold phenomena and complexity: a statistical physics analysis of the random Satisfiability problem.

Rémi Monasson [1]

Laboratoire de Physique Théorique de l'ENS, 75005 Paris.

Abstract

Recently, it has been recognized that phase transitions play an important role in the probabilistic analysis of combinatorial optimization problems. This lecture aims at presenting the tools and concepts designed by physicists to deal with optimization or decision problems in an accessible language for computer scientists and mathematicians, with no prerequisites in physics. We focus here on the threshold phenomenon related to the random Satisfiability problem, and its relationship with resolution complexity.

[1] E-mail: monasson@lpt.ens.fr; URL: http://www.lpt.ens.fr/˜ monasson

School on complexity, CIRM, Marseille, 28 janvier - 1 Fevrier 2002.

1 Introduction

At the heart of statistical physics, discrete mathematics, and theoretical computer science, lie mathematically similar counting and optimization problems. This situation leads to a transgression of boundaries so that progress in one discipline can benefit the others. An old example of this is the work of Kasteleyn (a physicist) who introduced a method for counting perfect matchings over planar graphs (a discrete mathematics problem). Our belief is that a similar cross-fertilization of methods and models should arise in the study of combinatorial problems over *random structures*. Such problems have attracted the attention of a large community of researcher in the last decade, but a transgression of boundaries has only just begun.

The present lecture, devoted to the statistical mechanics approach to the random K-Satisfiability problem, is an attempt in this direction. The mathematician reader should keep in mind that the way physicists and mathematicians proceed is quite different. Theoretical physicists generally do not prove theorems, rather they attempt to understand problems by obtaining exact and approximate results based on reasonable hypotheses. In practice, these hypotheses are "validated" *a posteriori* through comparison with experiments or numerical simulations, and through consistency with the overall body of knowledge in physics. In this sense, theoretical physics must be distinguished from mathematical physics whose scope is to make rigorous statements. Of course, exact solutions play an important role in statistical physics in that they represent limiting cases where analytical or numerical techniques can be checked, but they are not the main focus of this discipline.

Though this lecture focuses on K-Satisfiability only, many various combinatorial problems of interest have been considered by mathematicians and physicists. For a recent account of ongoing researches from the statistical physics point of view, the reader is referred to [1].

2 Presentation of the random K-satisfiability problem.

The satisfaction of constrained Boolean formulae is a key issue in complexity theory. Many computational problems are known to be NP-complete [2,3] through a polynomial mapping onto the K-Satisfiability (SAT) problem, which in turn was the first problem shown to be NP-complete by Cook in 1971 [4].

Recently [5], there has been much interest in a random version of the K-SAT problem defined as follows. Consider N Boolean variables x_i, $i = 1, \ldots, N$. Call a clause C the logical OR of K randomly chosen variables, each of them being negated or left unchanged with equal probabilities. Then repeat this process by drawing independently M random clauses C_ℓ, $\ell = 1, \ldots, M$. The logical AND of all these clauses is a "formula", referred to as F. It is said to be satisfiable if there exists a logical assignment of the xs evaluating F to true, and unsatisfiable otherwise.

Numerical experiments have concentrated on the study of the probability $P_N(\alpha, K)$ that a randomly chosen F having $M = \alpha N$ clauses be satisfiable (Figure 1). For large sizes, a remarkable behaviour arises: P_N seems to reach unity for $\alpha < \alpha_c(K)$ and vanishes for $\alpha > \alpha_c(K)$ when $N \to \infty$ [6,5]. Such an abrupt threshold behaviour, separating a SAT phase from an UNSAT one, has indeed been rigorously confirmed for 2-SAT, which is in P, with $\alpha_c(2) = 1$ [7,8]. For larger $K \geq 3$, K-SAT is NP-complete and much less is known. The existence of a sharp transition has not been rigorously proved but estimates of the thresholds have been found : $\alpha_c(3) \simeq 4.3$ [9]. Moreover, some rigorous lower and upper bounds to $\alpha_c(3)$ (if it exists), $\alpha_{l.b.} = 3.26$ and $\alpha_{u.b.} = 4.51$ respectively have been established [10–12].

The interest in random K-SAT arises partly from the following fact: it has been observed numerically that hard random instances are generated when the problems are critically constrained, i.e., close to the SAT/UNSAT phase boundary [6,5] (Figure 2). The study of such hard instances represents a theoretical challenge towards an understanding of complexity and the analysis of exact algorithms. Moreover, hard random instances are also a test-bed for the optimization of heuristic search procedures, which are widely used in practice.

Statistical mechanics provides new intuition on the nature of the solutions of random K-SAT (or MAX-K-SAT) through the introduction of an order parameter which describes the geometrical structure of the space of solutions. In addition, it gives also a global picture of the dynamical operation of search procedures and the computational complexity of K-SAT solving.

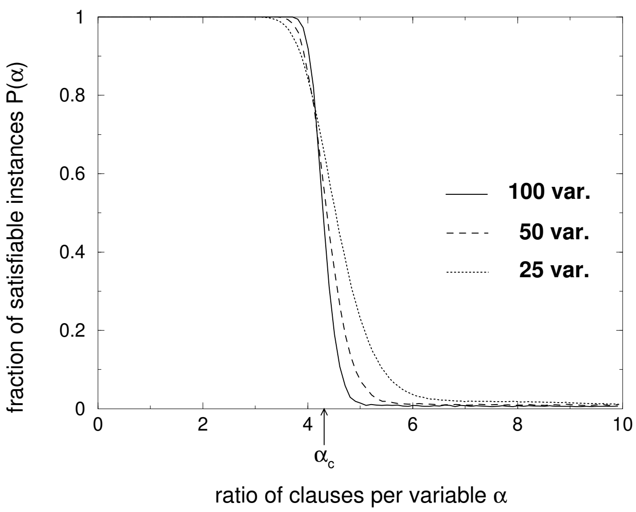

Fig. 1. Probability $P(\alpha)$ that a set of αN clauses be satisfiable for different sizes N (number of variables). This probability is obtained from numerical experiments, consisting in generating randomly 10,000 instances of the 3-SAT problem for each values of α and N, and measuring the fraction of them being satisfiable. In the large N limit, the threshold $\alpha_c \simeq 4.3$ separates the satisfiable ($\alpha < \alpha_c$, $P = 1$) and the non satisfiable ($\alpha > \alpha_c$, $P = 0$) phases. A similar transition also takes place for random 2-SAT instances, where all clauses involve two variables; the threshold is then equal to $\alpha_c = 1$.

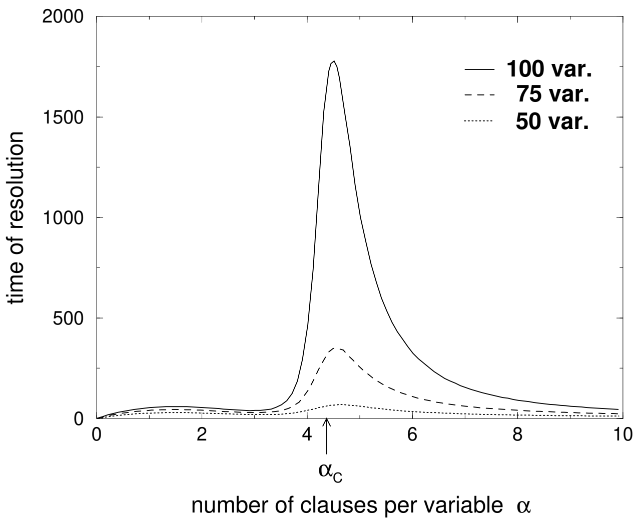

Fig. 2. Resolution time of 3-SAT instances as a function of the ratio of clauses per variable α and for three different sizes. The time is measured from the size of the search tree generated by the solving procedure (see Section 5). Data correspond to the median resolution time of 10,000 instances; the average time may be somewhat larger due to the presence of rare, exceptionally hard instances. The computational complexity is maximal at α_c. It is exponential in the vicinity of the threshold and in the unsatisfiable phase, but less and less as α increases.

Dictionary

K-SAT	*statistical physics*
Boolean variable x = True , False	*Ising spin* $s_x \in +1, -1$
Clauses	*Couplings and fields acting on spins*
Number of clauses violated by a logical configuration	*Energy E of the spins configuration*
some examples:	
2 SAT $\quad x$ or $\bar{y}$	$E = \frac{1}{4}(1-s_x)(1+s_y)$
(x or $\bar{y}$) and ($\bar{x}$ or z)	$E = \frac{1}{4}(1-s_x)(1+s_y) + \frac{1}{4}(1+s_x)(1-s_y)$
3 SAT $\quad x$ or $\bar{y}$ or z	$E = \frac{1}{8}(1-s_x)(1+s_y)(1-s_z)$
Minimal number of violated clauses	*Ground state energy*
The problem is satisfiable	*Ground state energy* $= 0$
The problem is not satisfiable	*Ground state energy* > 0

Fig. 3. Dictionary translating the 3-SAT problem (left) into statistical physics language (right).

3 Statistical physics framework.

3.1 Energy and partition function of K-SAT.

In order to unveil the relationship between random satisfiability problems and statistical physics, let us translate in Figure 3 the 3-SAT problem and its ingredients (Boolean variables, clauses, ...) in a language more familiar to physicists (Ising spins, interactions, ...). The main idea is to introduce an "energy" function, which is merely a cost function equal to the number of unsatisfied clauses for each variables-spins configuration, and study the ground state properties. The satisfiability of the 3-SAT problem is therefore equivalent to the vanishing of the minimal energy over spins configurations, also called ground state energy.

The logical values of an x_i can be represented by a binary variable S_i, called a spin, through the one-to-one mapping $S_i = -1$ (respectively $+1$) if x_i is false (resp. true). The random clauses can then be encoded into an $M \times N$ matrix $C_{\ell i}$ in the following way : $C_{\ell i} = -1$ (respectively $+1$) if the clause C_ℓ includes $\overline{x_i}$ (resp. x_i), $C_{\ell i} = 0$ otherwise. It can be checked easily that $\sum_{i=1}^N C_{\ell i} S_i$ equals the number of wrong literals in clause ℓ. Consider now the cost-function $E[\mathbf{C}, \mathbf{S}]$ defined as the number of clauses that are not satisfied by the logical assignment corresponding to configuration $\mathbf{S}$.

$$E[\mathbf{C}, \mathbf{S}] = \sum_{\ell=1}^{M} \delta \left(\sum_{i=1}^{N} C_{\ell i} S_i + 3 \right) \quad , \tag{1}$$

where $\delta(j) = 1$ if $j = 0$, zero otherwise, denotes the Kronecker function. The minimum (or ground state -GS) $E[\mathbf{C}]$ of $E[\mathbf{C}, \mathbf{S}]$, is the lowest number of violated clauses that can be achieved by the best possible logical assignment [13]. It is a random variable, that becomes highly concentrated around the average value $E_{GS} \equiv \overline{E[\mathbf{C}]}$ in the large size limit [14] (the overbar denotes the average over the clauses). The latter is accessible through the knowledge of the averaged logarithm of the generating function,

$$G[z, \mathbf{C}] = \sum_{\mathbf{S}} z^{E[\mathbf{C}, \mathbf{S}]} \tag{2}$$

since

$$E_{GS} = \lim_{z \to 0} \frac{\overline{\log G[z, \mathbf{C}]}}{\log z} \quad , \tag{3}$$

when the auxiliary parameter z is sent to zero. Being the minimal number of violated clauses, E_{GS} equals zero in the sat region and is strictly positive in the unsat phase. The knowledge of E_{GS} as a function of α therefore determines the threshold ratio $\alpha_c(K)$.

3.2 Average of the generating function G over disorder.

The calculation of the average value of the logarithm of the generating function (called partition function physics) in (3) is an awkward one. To circumvent this difficulty, we compute the n^{th} moment of G for integer-valued n and perform an analytic continuation to real n to exploit the identity $\overline{G[z,\mathbf{C}]^n} = 1 + n\,\overline{\log G[z,\mathbf{C}]} + O(n^2)$. The n^{th} moment of G is obtained by replicating n times the sum over the spin configurations $\mathbf{S}$ and averaging over the clause distribution [13]

$$G_n(z) \equiv \overline{G[z,\mathbf{C}]^n} = \sum_{\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n} \overline{z^{\,(E[\mathbf{C},\mathbf{S}^1]+E[\mathbf{C},\mathbf{S}^2]+\ldots+E[\mathbf{C},\mathbf{S}^n])}} \quad . \tag{4}$$

In order to compute the expectation values that appear in eq.(4)eq.(4), one notices that the r.h.s. of factorises over the sets of different clauses due to the absence of any correlation in their probability distribution. It follows

$$G_n(z) = \sum_{\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n} \Big(\zeta[\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n]\Big)^M \qquad , \tag{5}$$

with

$$\zeta[\{\mathbf{S}^a\}] = \overline{\exp\left[\log z \sum_{a=1}^{n} \delta\left(\sum_{i=1}^{N} C_i S_i^a + 3\right)\right]} \qquad , \tag{6}$$

with the bar denoting the uniform average over the set of $2^3\binom{N}{3}$ vectors of N components $C_i = 0, \pm 1$ and of squared norm equal to 3.

Resorting to the identity,

$$\delta\left(\sum_{i=1}^{N} C_i S_i^a + 3\right) = \prod_{i/C_i\neq 0} \delta\left(S_i^a + C_i\right) \qquad , \tag{7}$$

one may carry out the average over in disorder in eq.(6) to obtain

$$\zeta[\{\mathbf{S}^a\}] = \frac{1}{8}\sum_{C_1,\ldots,C_K=\pm1} \frac{1}{N^3} \sum_{i_1,i_2,i_3=1}^{N} \exp\left\{-\frac{1}{T}\sum_{a=1}^{n}\prod_{\ell=1}^{3}\delta\left[S_{i_\ell}^a + C_\ell\right]\right\} \tag{8}$$

up to negligible $O(1/N)$ contributions.

The averaged term in the r.h.s. of (4) depends on the $n \times N$ spin values only through the 2^n variables $\rho(\vec{\sigma})$ labelled by the vectors $\vec{\sigma}$ with n binary components; $\rho(\vec{\sigma})$ equals the number (divided by N) of labels i such that $S_i^a = \sigma^a$, $\forall a = 1,\ldots,n$. It follows that $\zeta[\{\mathbf{S}^a\}] = \zeta[\rho]$ where

$$\zeta[\rho] = \frac{1}{8}\sum_{C_1,C_2,C_3=\pm1}\ \sum_{\vec{\sigma}_1,\vec{\sigma}_2,\vec{\sigma}_3} \rho(-C_1\,\vec{\sigma}_1)\,\rho(-C_2\,\vec{\sigma}_2)\,\rho(-C_3\,\vec{\sigma}_3)\,\times \tag{9}$$

$$\times \exp \left\{ \log z \sum_{a=1}^{n} \prod_{\ell=1}^{3} \delta \left[\sigma_\ell^a - 1 \right] \right\} \quad .$$

We now substitute in (4) the summation over the 2^{nM} replicated spins configurations with the summation over the variables $\rho(\vec{\sigma})$, and write

$$G_n(z) = \sum_{\rho(\vec{\sigma})} \zeta[\rho]^M \, J(\rho) \quad , \tag{10}$$

where

$$J(\rho) = \frac{N!}{\prod_{\vec{\sigma}} (N\,\rho(\vec{\sigma}))!}, \tag{11}$$

is the multiplicity of (number of replicated spins configurations corresponding to) the "configuration" of variables ρ.

To leading order in N (by resorting to a saddle point integration), the final expression of the n^{th} moment of Z can be written as $G_n(z) \simeq z^{N\, F_{opt}(z)}$ where $F_{opt}(z)$ is the optimum (in fact the minimum for integer n) over all possible ρs of the functional [13]

$$F[z,\rho] = \frac{1}{\log z} \left\{ E[z,\rho] - \sum_{\vec{\sigma}} \rho(\vec{\sigma}) \log \rho(\vec{\sigma}) \right\} \quad , \tag{12}$$

with

$$E[z,\rho] = \alpha \log \left[\sum_{\vec{\sigma}_1,\vec{\sigma}_2,\vec{\sigma}_3} \rho(\vec{\sigma}_1)\, \rho(\vec{\sigma}_2)\, \rho(\vec{\sigma}_3) \; \exp \left(\log z \sum_{a=1}^{n} \prod_{\ell=1}^{3} \delta[\sigma_\ell^a - 1] \right) \right] \; . \tag{13}$$

Note that the optimum of F fulfills $\rho(\vec{\sigma}) = \rho(-\vec{\sigma})$ due to the unbias distribution of the disorder C (each literal and its negation appear with equal probabilities).

3.3 Order parameter and replica-symmetric saddle-point equations.

The optimization conditions over $F[\rho]$ provide 2^n coupled equations for the $\rho(\vec{\sigma})$s. Notice that F is a symmetric functional, invariant under any permutation of the replicas a, as is evident from equation (4). An extremum may thus be sought in the so-called replica symmetric (RS) subspace of dimension $n+1$ where $\rho(\vec{\sigma})$ is left unchanged under the action of the symmetric group.

In the limit of interest, $z \to 0$, and within the RS subspace, the ρs may be conveniently expressed as the moments of some probability density $P(m)$ over the range $-1 \leq m \leq 1$ [13],

$$\rho(\sigma_1, \sigma_2, \ldots, \sigma_n) = \int_{-1}^{1} dm\ P(m) \prod_{a=1}^{n} \left(\frac{1 + m\sigma^a}{2} \right) \quad . \tag{14}$$

$P(m)$ is not uniquely defined by (14) for integer values of n but acquires some precise meaning in the $n \to 0$ limit. It is the probability density of the expectation values of the spin variables over the set of ground states. Consider a formula Q and all the spin configurations $\mathbf{S}^{(\ell)}$, $\ell = 1, \ldots, L$ realizing the minimum $E[\mathbf{C}]$ of the cost-function $E[\mathbf{C}, \mathbf{S}]$, that is the solutions of the MAX-SAT problem defined by Q. Then define the average magnetizations of the spins

$$m_i = \frac{1}{L} \sum_{\ell=1}^{L} S_i^{(\ell)} \quad , \tag{15}$$

over the set of optimal configurations. Call $H(\mathbf{C}, m)$ the histogram of the m_is and $H(m)$ its quenched average, i.e., the average of $H(\mathbf{C}, m)$ over the random choices of formula Q. $H(m)$ is a probability density over the interval $-1 \leq m \leq 1$ giving information on the distribution of the variables induced by the constraint of satisfying all the clauses. In the absence of clauses, all assignments are solutions and all magnetizations vanish: $H(m) = \delta(m)$ (Dirac distribution) and variables are not constrained. Oppositely, variables that always take the same value in all solutions, if any, have magnetizations equal to $+1$ (or -1): such variables are totally constrained by the clauses.

As discussed in ref. [13], if the RS solution is the global optimum of (12) then $H(m)$ equals the above mentioned $P(m)$ in the limit of large sizes $N \to \infty$. Therefore, the order parameter arising in the replica calculation reflects the "microscopic" structure of the solutions of the K-SAT problem.

At this stage of the analysis it is possible to perform the analytic continuation $n \to 0$ since all the functionals have been expressed in term of the generic number of replicas n. Such a process leads to a self-consistent functional equation for the order parameter $P(m)$, which reads

$$P(m) = \frac{1}{1-m^2} \int_{-\infty}^{\infty} du\ \cos\left[\frac{u}{2} \ln\left(\frac{1+m}{1-m} \right) \right] \times$$

$$\exp\left[-\alpha K + \alpha K \int_{-1}^{1} \prod_{\ell=1}^{K-1} dm_\ell P(m_\ell) \cos\left(\frac{u}{2}\ln A_{K-1}\right)\right] \tag{16}$$

with

$$A_K \equiv A_K(\{m_\ell\}, z) = 1 + (z-1)\prod_{\ell=1}^{K}\left(\frac{1+m_\ell}{2}\right) \quad . \tag{17}$$

For the sake of generality, we consider here a generic number K of literals per clause, the case of 3-SAT being recovered for $K=3$. The corresponding optimum of the functional (12) reads

$$\begin{aligned} F_{opt}(z,K,\alpha) = \frac{1}{\log z}\Bigg\{ & \log 2 - (K-1)\,\alpha \int_{-1}^{1} \prod_{\ell=1}^{K} dm_\ell P(m_\ell) \log A_K \\ & + \frac{K\,\alpha}{2} \int_{-1}^{1} \prod_{\ell=1}^{K-1} dm_\ell P(m_\ell) \log A_{K-1} \\ & - \frac{1}{2}\int_{-1}^{1} dm P(m) \log(1-m^2) \Bigg\} \quad . \end{aligned} \tag{18}$$

It can be checked that equation (16) is recovered when optimizing the free-energy functional (18) over all (even) probability densities $P(m)$ on the interval [-1,1].

3.4 A toy model: random 1-SAT.

Before entering in the analysis of the saddle-point equations for K-SAT, it is worth considering the simple $K=1$ case which can be solved either by a direct combinatorial method or within the statistical mechanics approach. Though random 1-SAT does not present any critical behaviour (for finite α), its study allows an intuitive understanding of the meaning and correctness of the statistical mechanics approach.

For $K=1$, a sample of M clauses can be defined completely by giving directly the numbers t_i and f_i of clauses imposing that a certain Boolean variable S_i must be true or false respectively. The partition function corresponding to a given sample reads

$$G[z, \mathbf{C} = \{t,f\}] = \prod_{i=1}^{N}\left(z^{t_i} + z^{f_i}\right) \quad , \tag{19}$$

and the average over the disorder gives

$$
\begin{aligned}
\frac{1}{N}\overline{\ln G[z,\mathbf{C}]} &= \frac{1}{N}\sum_{\{t_i,f_i\}} \frac{M!}{\prod_{i=1}^{N}(t_i!f_i!)} \ln G[z,\{t,f\}] \\
&= \ln 2 + \frac{\alpha}{2}\log z + \sum_{\ell=-\infty}^{\infty} e^{-\alpha} I_\ell(\alpha)\ \ln\left(\frac{z^{\ell/2}+z^{-\ell/2}}{2}\right) \quad , \qquad (20)
\end{aligned}
$$

where I_l denotes the l^{th} modified Bessel function. The $z \to 0$ limit gives the ground state energy *density* (i.e. fraction of violated clauses)

$$
e_{GS}(\alpha) = \frac{\alpha}{2}[1 - e^{-\alpha} I_0(\alpha) - e^{-\alpha} I_1(\alpha)] \qquad (21)
$$

and the ground state entropy *density* (i.e. logarithm –divided by N– of the number of spins configurations having minimal energy)

$$
s_{GS}(\alpha) = e^{-\alpha} I_0(\alpha)\ \ln 2 \quad . \qquad (22)
$$

For any $\alpha > 0$, the ground–state energy density is positive and therefore the overall Boolean formula is false with probability one. Also, the entropy is finite, *i.e.*, the number of minima of the energy for any α is exponentially large. Such a result can be understood by noticing that there exist a fraction of unconstrained variables $e^{-\alpha} I_0(\alpha)$ which are subject to equal but opposite constraints $t_i = f_i$.

The above results are recovered in the statistical mechanics framework, thereby showing that the RS Ansatz is exact for all z and α when $K = 1$.

The solution of the saddle-point equation (16) can be found for any parameter z leading to the expression

$$
P(m) = \sum_{\ell=-\infty}^{\infty} e^{-\alpha} I_\ell(\alpha)\ \delta\left(m - \frac{1-z^\ell}{1+z^\ell}\right) \quad . \qquad (23)
$$

In the limit of interest $z \to 0$, this formula reads

$$
P(m) = e^{-\alpha} I_0(\alpha)\delta(m) + \frac{1}{2}(1 - e^{-\alpha} I_0(\alpha))\left(\delta(m-1) + \delta(m+1)\right) \quad . \qquad (24)
$$

As shown in Figure 4, the fraction of unconstrained variables is simply associated with the unfrozen spins and thus gives the weight of the Dirac distribution at $m = 0$. On the contrary, the non-zero value of the fraction of violated clauses, the ground-state energy density, is due to the presence of completely

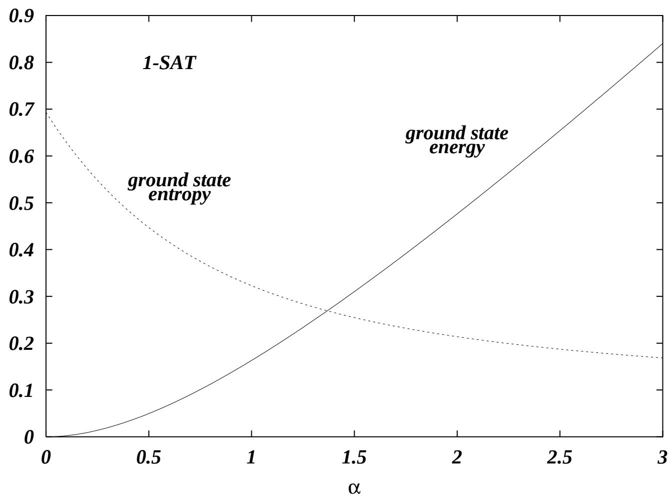

Fig. 4. Typical fraction of violated clauses (bold line) and entropy density (thin line) versus α in a random 1-SAT formula, in the limit $N \to \infty$.

frozen (over constrained) spins of magnetizations $m = \pm 1$. Such a feature remains valid for any K.

4 Sat and Unsat phases; Sat/Unsat transition.

4.1 Sat phase: structure of the space of solutions.

We start by considering the sat phase. An interesting quantity to look at is the typical number of solutions of the random K-SAT problem, *i.e.* the typical ground state entropy density, that is the logarithm of the number of solutions, $s_{GS}(\alpha)$ given by eq.(18) in the $z \to 0$ limit.

In the absence of any clauses, all assignments are solutions: $s_{GS}(\alpha = 0) = \ln 2$. We have computed the Taylor expansion of $s_{GS}(\alpha)$ in the vicinity of $\alpha = 0$, up to the seventh order in α. Results are shown in Figure 5. It is found that $s_{GS}(\alpha_c = 1) = .38$ and $s_{GS}(\alpha = 4.2) = .1$ for 2-SAT and 3-SAT respectively: just below threshold, solutions are exponentially numerous. This result is confirmed by rigorous work [15].

More involved calculations, including replica symmetry breaking (RSB) effects [17], have shown that the value of the entropy is insensitive to RSB in the sat phase. Therefore the RS calculation provides a precise estimate of the entropy (believed to be exact at low α ratios, see Talagrand's paper in this volume for

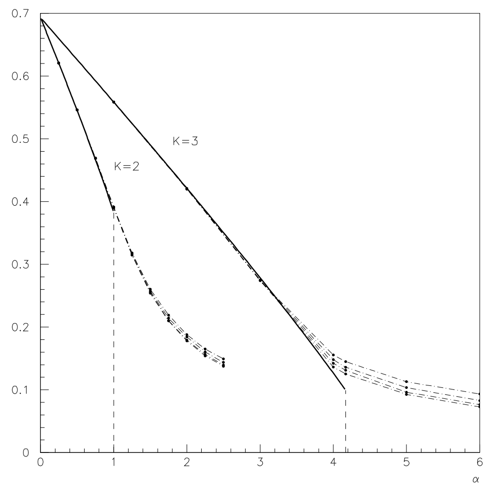

Fig. 5. RS estimate for the entropy density in random 2-SAT and 3-SAT below their thresholds. RSB corrections due to clustering are absent in 2-SAT and very small (within few a percent) in 3-SAT. The dots represent the results of exact enumerations over small systems (N ranging from 20 to 30, see ref. [16])

a discussion).

Recent analytical calculations for 3-SAT [17] (also confirmed by numerical investigations) indicate that the RS theory breaks down at a definite ratio α_{RSB} below α_c, where the solutions start to be organized into distinct clusters. The meaning of this statement is as follows. Think of the space of spins configurations as the N-dimensional hypercube. Optimal assignments are a subset of the set of 2^N vertices on the hypercube. Replica symmetry amounts to assume that any pair of vertices are a.s. separated by the same Hamming distance d, defined as the fraction of distinct spins in the corresponding configurations. In other words, solutions are gathered in a single cluster, of diameter $d\,N$. RSB variational calculations [17] show that this simplifying assumption is not generally true in the whole sat phase and that another scenario may take place close to threshold:

- Below α_{RSB} the space of solutions is replica symmetric. There exist one cluster of solutions characterized by a single probability distribution of local magnetizations. The Hamming distance d is a decreasing function of α, starting at $d(0) = 1/2$.
- At $\alpha_{RSB} \simeq 4.0$, the space of solutions breaks into an exponential number (in N) of different clusters. Each cluster contains an exponential number of solutions. The typical Hamming distance d_0 between solutions belonging to different clusters is close to 0.3 and remains almost constant (slightly decreasing) up to α_c, indicating that the centers of these clusters do not move on the hypercube when more and more clauses are added. Within each cluster, solutions tend to become more and more similar, with a rapidly decreasing intra-cluster Hamming distance d_1.

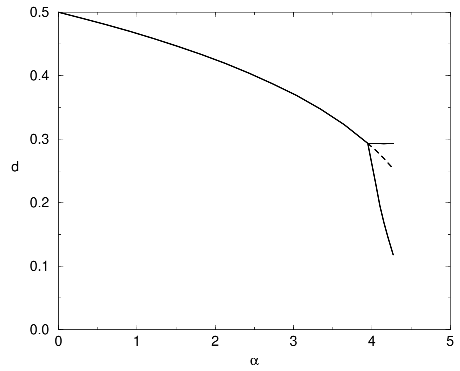

Fig. 6. Variational RSB estimate for the clustering of solutions below α_c for $3-SAT$. d is the typical Hamming distance between solutions. The splitting of the curves at $\alpha \simeq 4$ corresponds to clustering. There appear two characteristic distances, one within each cluster and one between solutions belonging to different clusters.

Figure 6 provides a qualitative representation of the clustering process. The fact that the Hamming distance can take two values at most is a direct consequence of the RSB Ansatz. In reality, the distance distribution could be more complicated. The key point is that statistical mechanics calculations strongly support the idea that the space of solutions has a highly organized structure, even in the sat phase.

4.2 Unsat phase: the backbone and the order of the phase transition.

In the unsat phase, it is expected that $O(N)$ variables become totally constrained, *i.e.* take on the same value in all the ground states. Such a hypothesis, which of course needs to be verified *a posteriori*, corresponds to a structural change in the probability distribution $P(m)$ which develops Dirac distributions with finite masses at $m = \pm 1$.

In the limit of interest $(z \to 0)$, to describe the accumulation of the magnetization on the borders of its domain $(m \in [-1; 1])$, we introduce the rescaled variable μ, implicitly defined by the relation $m = (1 - z^{2\mu})/(1 + z^{2\mu})$, see equation (23). Calling $R(\mu)$ the probability density of the μs, the saddle-point equations read

$$
\begin{aligned}
R(\mu) = \int_{-\infty}^{\infty} \frac{du}{2\pi} \cos(u\mu) \exp\Bigg[-\frac{\alpha K}{2^{K-1}} + \alpha K \times \\
\times \int_{0}^{\infty} \prod_{\ell=1}^{K-1} d\mu_\ell R(\mu_\ell) \cos(u \, \min(1, \mu_1, \ldots, \mu_{K-1})) \Bigg] \quad . \qquad (25)
\end{aligned}
$$

The corresponding ground state energy density reads, see (18),

$$
\begin{aligned}
e_{GS}(\alpha) = & \, \alpha(1-K) \int_0^\infty \prod_{\ell=1}^{K} d\mu_\ell R(\mu_\ell) \min(1,\mu_1,\ldots,\mu_K) \\
& + \frac{\alpha K}{2} \int_0^\infty \prod_{\ell=1}^{K-1} d\mu_\ell R(\mu_\ell) \min(1,\mu_1,\ldots,\mu_{K-1}) - \int_0^\infty d\mu R(\mu)\mu \quad . \qquad (26)
\end{aligned}
$$

It is easy to see that the saddle–point equation (25) is in fact a self–consistent identity for $R(\mu)$ in the range $\mu \in [0,1]$ only. Outside this interval, equation (25) is merely a definition of the functional order parameter R.

As discussed in detail in ref. [13], equations (25) admit an infinite sequence of more and more structured exact solutions of the form

$$
R(\mu) = \sum_{l=-\infty}^{\infty} r_\ell \, \delta\left(\mu - \frac{\ell}{q}\right) \quad , \qquad (27)
$$

having exactly q peaks in the interval $[0,1[$, whose centers are $\mu_\ell = \frac{\ell}{q}$, $\ell = 0,\ldots,q-1$. The corresponding energy density reads, from (27) and (26),

$$
\begin{aligned}
e_{GS} = & \frac{\alpha(1-K)}{q} \left[\left(\frac{1-r_0}{2}\right)^K + \sum_{j=1}^{q-1} \left(\frac{1-r_0}{2} - \sum_{l=1}^{j} r_l\right)^K \right] \\
& + \frac{\alpha K}{2q} \left[\left(\frac{1-r_0}{2}\right)^{K-1} + \sum_{j=1}^{q-1} \left(\frac{1-r_0}{2} - \sum_{l=1}^{j} r_l\right)^{K-1} \right] \\
& - \sum_{j=1}^{q} \frac{j}{q} \gamma_j \left(\frac{r_0}{2} + \frac{r_j}{2} + \sum_{l=1}^{j-1} r_l \right) \quad . \qquad (28)
\end{aligned}
$$

Though there might be continuous solutions to (25), it is hoped that the energy of ground state can be arbitrarily well approximated by the above large q solutions.

The location of the sat/unsat threshold can be obtained for any K by looking at the value of α beyond which the ground state energy becomes positive. For $2-SAT$ the exact result $\alpha_c(2) = 1$ is recovered whereas for $K > 2$ the RS energy becomes positive at a value of α (e.g., $\alpha_c(3) \simeq 4.6$ as shown in figure 7) which is sightly higher than the value estimated by numerical simulations.

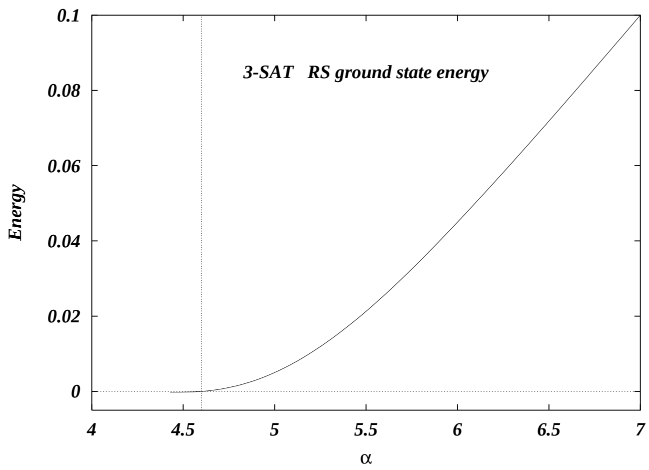

Fig. 7. RS estimate for the ground state energy density, i.e., the number of violated clauses divided by N in random 3-SAT. The prediction is given as a function of α, for $q \gg 1$ and in the limit $N \to \infty$. See ref. [13] for details.

4.3 A hint at replica symmetry breaking.

The RS theory provides an upper bound for the thresholds for any $K > 2$, whereas the exact values can be obtained only by adopting a more general functional form for the solution of the saddle-point equations which explicitly breaks the symmetry between replicas (see ref. [18] for a precise discussion). Such an issue is indeed a relevant, and largely open, problem in the statistical physics of random systems [19–26].

The general structure of the functional order parameter which describes solutions that break the permutational symmetry among replicas consists of a distribution of probability densities: each Boolean variable fluctuates from one cluster of solutions to another, leading to a site dependent probability density of local Boolean magnetizations. The distribution over all different variables then provides a probability distribution of probability distributions. The above scheme can in principle be iterated, leading to more and more refined levels of clustering of solutions. Such a scenario would correspond to the so-called continuous RSB scheme [27]. However the first step solution could suffice to capture the exact solution of random K-SAT, as happens in other similar random systems [27].

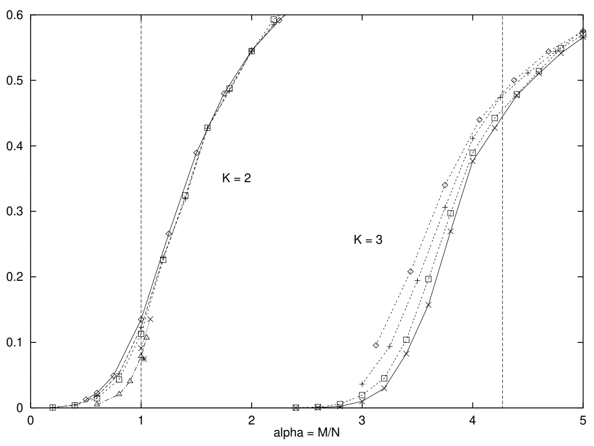

Fig. 8. Numerical estimates of the value of the backbone order parameter in 2-SAT and 3-SAT. The curves [28] are obtained by complete enumerations in small systems (up to $N = 500$ variables for 2-SAT and $N = 30$ for 3-SAT) averaged over many samples.

4.4 Abrupt vs. smooth phase transition.

Of particular interest are the fully constrained variables – the so called *backbone* component –, that is the x_is such that $m_i = \pm 1$. Within the RS Ansatz, the fraction of fully constrained variables $\gamma(\alpha, K)$ can be directly computed from the saddle-point equations. Clearly, $\gamma(\alpha, K)$ vanishes in the SAT region otherwise the addition of ϵN new clauses to F would lead to a contradiction with a finite probability for any $\epsilon > 0$. Two kinds of scenarii have been found when entering the unsat phase. For 2-SAT, $\gamma(\alpha, 2)$ smoothly increases above the threshold $\alpha_c(2) = 1$. For 3-SAT (and more generally $K \geq 3$), $\gamma(\alpha, 3)$ exhibits a discontinuous jump to a finite value γ_c slightly above the threshold. A finite fraction of variables become suddenly over constrained when crossing the threshold! Numerical results on the growth of the backbone order parameter are given in figure 8.

4.5 The random 2+p-SAT model.

The sat/unsat transition is accompanied by a smooth (respectively abrupt) change in the backbone component and therefore in the structure of the solutions of the 2-SAT (resp. 3-SAT) problem. A better way to understand how such a change takes place is to consider a mixed model, which continuously interpolates between 2-SAT and 3-SAT. The so-called $2 + p$-SAT model [28] includes a fraction p (resp. $1 - p$) of clauses of length two (resp. three). 2-SAT

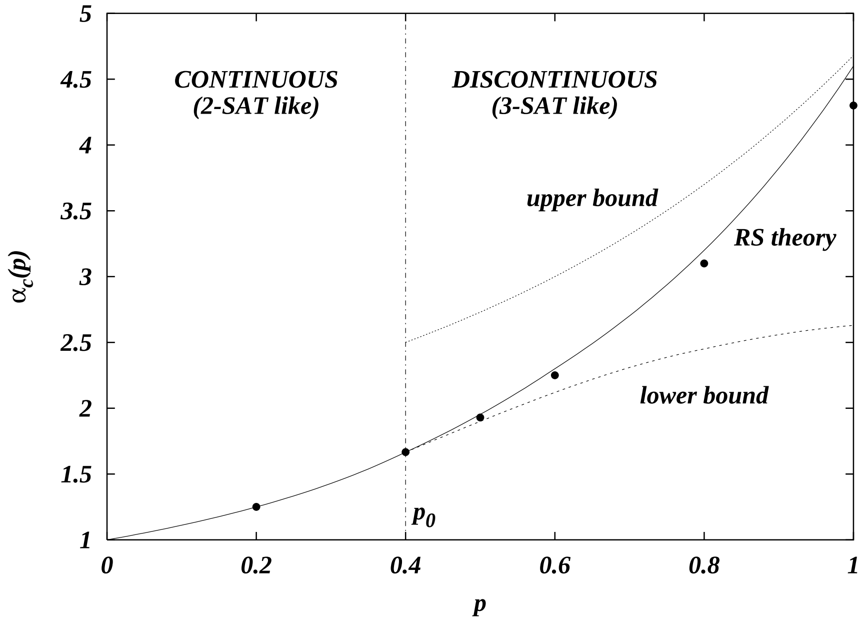

Fig. 9. $\alpha_c(p)$ versus p in random 2+p-SAT. Up to $p_0 \simeq .4$ $\alpha_c(p) = 1/(1-p)$, in agreement with rigorous results. For $p > p_0$ the transition becomes discontinuous in the backbone order parameter and the RS theory provides an upper bound for $\alpha_c(p)$ which is within a few percent of the results of numerical simulations (dots) [28,30].

is recovered for $p = 0$ and 3-SAT when $p = 1$. The RS theory predicts that, at the sat/unsat transition, the appearance of the backbone component becomes abrupt when $p > p_0 \simeq 0.4$ (see figure 9). On the contrary, when $p < p_0$, the transition is smooth as in the 2-SAT case. Such a scenario is consistent with both rigorous results based on the probabilistic analysis of simple algorithm [29] and with variational calculations [17] which include RSB effects.

An additional argument in favor of the above picture is given by the analysis of the finite-size effects on $P_N(\alpha, K)$ and the emergence of some universality for $p < p_0$. (The definition of P_N was given when we began discussing the properties of K-SAT.) A detailed account of these findings may be found in [28,30]. For $p < p_0$ the size of the critical window where the transition takes place is observed to remain constant and close to the value expected for 2-SAT. The critical behaviour is the same as for the percolation transition in random graphs (see also ref. [31]). For $p > p_0$ the size of the window shrinks following some non-universal exponents toward its statistical lower bound [32] but numerical data do not allow for any precise estimate. The balanced version of 2+p-SAT can be studied exactly and both the phase diagram and the critical exponents turn out to behave very similarly to the ones of $2 + p$-SAT [33].

As we shall conclude in the next section, the knowledge of the phase diagram of the 2+p-SAT model is very precious to understand the computational complexity of 3-SAT solving.

5 Computational complexity and dynamics.

5.1 Introduction.

Numerical experiments have shown that the typical solving time of search algorithms displays an easy-hard-easy pattern as a function of α with a peak of complexity close to the threshold. Since computational complexity is strongly affected by the presence of a phase transition, it is a natural question to ask whether the nature of this phase transition plays an important role too. The peak in the search cost seems indeed to scale polynomially with N (even using Davis-Putnam-like procedures) for the 2-SAT problem, where the transition is continuous, and exponentially with N in the 3-SAT case, for which the birth of the backbone is known to be discontinuous.

Precise numerical simulations [28,30] on the computational complexity of solving critical 2+p-SAT instances support the view that the crossover between polynomial and exponential scalings takes place at p_0, the very value of p separating continuous from discontinuous transitions. Though investigated 2+p-SAT instances are all critical and the problem itself is NP-complete for any $p > 0$, it is only when the phase transition is abrupt that hardness shows up (including the fastest known randomized search algorithms such as walksat [34]).

To understand why search algorithms require polynomial or exponential computational efforts, statistical studies of the solutions cannot be sufficient. A full dynamical study of how search procedures operate has to be carried out. Such studies had already been initiated by mathematicians in the easy region, where search tree are particularly simple and almost no backtracking occurs. Franco and Chao [35,36] have in particular analyzed the operation of DP algorithms with different kinds of heuristics and have shown that at small values of α, the typical complexity is linear in N. We now show how their results can be interpreted in the above framework of phase diagram (Figure 10), and how statistical physics allows to investigate also the whole range of values of α, including the hard phase[37,38].

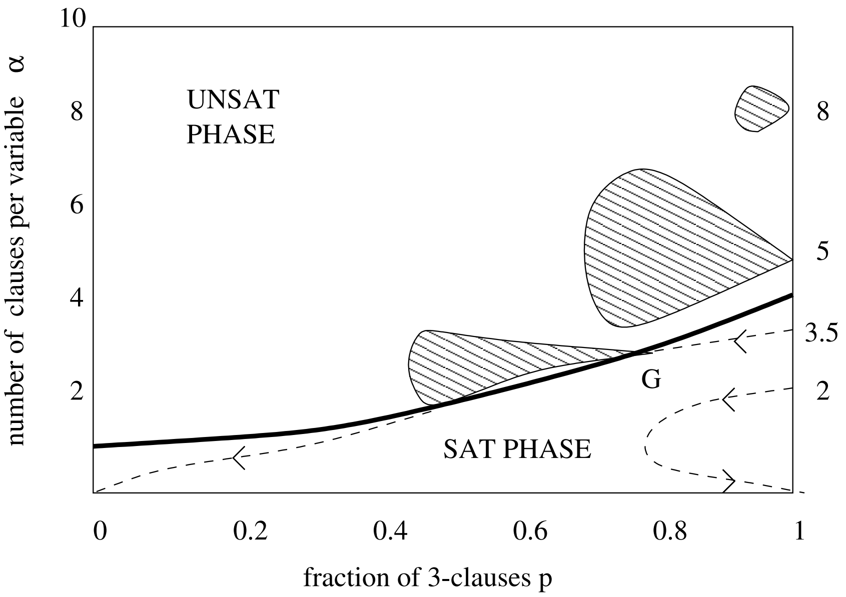

Fig. 10. Schematic phase diagram of 2+p-SAT and dynamical trajectories of DPLL. The threshold line $\alpha_c(p)$ (bold full line) separates sat (lower part of the plane) from unsat (upper part) phases. Extremities lie on the vertical 2-SAT (left) and 3-SAT (right) axis at coordinates ($p = 0, \alpha_c = 1$) and ($p = 1, \alpha_c \simeq 4.3$) respectively. Departure points for DPLL trajectories are located on the 3-SAT vertical axis and the corresponding values of α are explicitly given. Hatched regions symbolize tree trajectories in the unsat region, and dashed curve represent branch trajectories in the sat phase. Arrows indicate the direction of "motion" along branch trajectories parameterized by the fraction t of variables set by DPLL. For small ratios $\alpha < \alpha_L$, branch trajectories remain confined in the sat phase and end in S of coordinates $(1, 0)$, where a solution is found. At $\alpha_L \simeq 3.003$, the single branch trajectory hits tangentially the threshold line in T of coordinates $(2/5, 5/3)$ (T depends *a priori* on the heuristics used, but lies very close to the tricritical point). In the intermediate range $\alpha_L < \alpha < \alpha_c$, the branch trajectory intersects the threshold line at some point G (that depends on α). A dense tree then grows in the unsat phase, as happens when 3-SAT departure ratios are above threshold $\alpha > \alpha_c \simeq 4.3$. DPLL then reaches back the highest backtracking node in the search tree, that is, the first node when $\alpha > \alpha_C$, or node G for $\alpha_L < \alpha < \alpha_C$. In the latter case, a solution can be reached from a new descending branch while, in the former case, unsatisfiability is proven, see Figure 12.

5.2 Davis–Putnam-Loveland–Logemann and trajectories.

Why is the complexity of 3-SAT solving affected by the phase transition? To answer this question, let us first briefly expose the ubiquitous solving procedure of Davis–Putnam-Loveland–Logemann (DPLL), see Figure 11. DPLL is an exhaustive search procedure operating by trials and errors, the sequence of which can be graphically represented by a search tree (Figure 12). Briefly, **(1)** a node in the tree corresponds to a choice of a variable. An outgoing branch, attached to the value the latter (true or false), grows from the node. Thus a node gives birth to two branches at most. **(2)** Along a branch, the logical implications of the last choice are analyzed. **(3)** If a contradiction (violated clause) arises, the last choice is modified (backtracking of the tree) and the procedure goes on along a new branch (step 2); if all clauses are satisfied, a solution is found and the search process is over; otherwise, the algorithm resumes to step 1.

Computational complexity, the amount of operation necessary to solve the instance, is given by the size of the search tree, *i.e.* the number of nodes it contains. Performances can be improved by designing sophisticated heuristic rules for choosing variables (step 1).

The DPLL algorithm gives rise to a complex, non Markovian dynamical process, differing largely from the common modelization of the evolution of physical systems e.g. Monte Carlo dynamics. Our study of the operation of DPLL is based on the following, elementary observation (Figure 11). The recursive procedure of DPLL turns the initial instance of the 3-SAT problem into a mixed instance with clauses of length two and three (clauses of length unity are eliminated at step 2 of the procedure, see above). Therefore, if the initial instance is symbolized by a point of coordinates $p = 1, \alpha$ in the phase diagram of Figure 9-10, this representative point will evolve under the dynamical action of the algorithm and define a trajectory. Three ranges of initial ratios α must be distinguished on the vertical axis $p = 1$ of the 3-SAT problem.

5.3 Easy, hard, and mixed resolutions.

When $\alpha < \alpha_L = 3.003$, the search tree essentially reduces to a unique branch (Figure 12A), which ends up with a solution [35,36]. The trajectory corresponding to this branch depends on the heuristic rule implemented in DPLL (Figure 11) and is schematically represented on Figure 10. It first heads to the left and then reverses to the right until reaching a point on the 3-SAT axis at a small ratio without ever leaving the sat region. Further action of DPLL leads to a rapid elimination of the remaining clauses and the trajectory ends up at the right lower corner $p = 1, \alpha = 0$, where a solution is achieved. Thus, in the range of ratios $\alpha < \alpha_L \simeq 3.003$, 3-SAT is easy to solve: the computational complexity scales linearly with the size N (Figure 2).

For ratios above threshold ($\alpha > \alpha_C \simeq 4.3$), instances almost never have a solution but a considerable amount of backtracking is necessary before proving that clauses are incompatible. As shown in Figure 12B, a generic unsat tree includes many branches. The number of nodes grows exponentially with N, and is conveniently expressed as $2^{N\omega}$. The sequence of points (p, α) character-

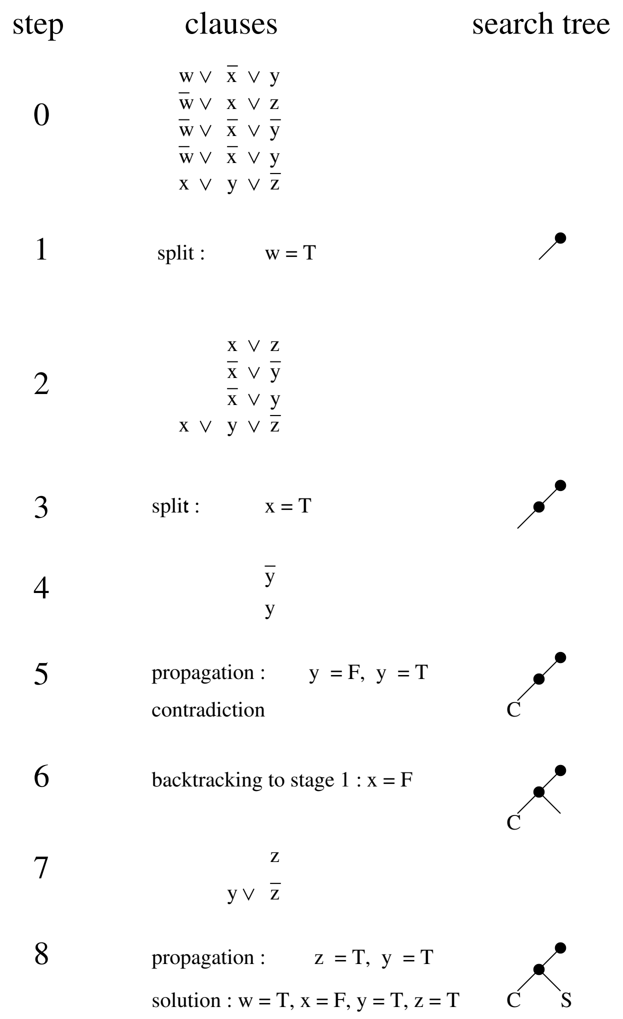

Fig. 11. Example of 3-SAT instance and Davis–Putnam- Loveland–Logemann resolution. **Step 0.** The instance consists of $M = 5$ clauses involving $N = 4$ variables x, y, w, z, which can be assigned to true (T) or false (F). $\bar{w}$ means (NOT w) and v denotes the logical OR. The search tree is empty. **1.** DPLL randomly selects a variable among the shortest clauses and assigns it to satisfy the clause it belongs to, e.g. $w =$T (splitting with the Generalized Unit Clause –GUC– heuristic) [6,9]. A node and an edge symbolizing respectively the variable chosen (w) and its value (T) are added to the tree. **2.** The logical implications of the last choice are extracted: clauses containing w are satisfied and eliminated, clauses including $\bar{w}$ are simplified and the remaining ones are left unchanged. If no unitary clause (*i.e.* with a single variable) is present, a new choice of variable has to be made. **3.** Splitting takes over. Another node and another edge are added to the tree. **4.** Same as step 2 but now unitary clauses are present. The variables they contain have to be fixed accordingly. **5.** The propagation of the unitary clauses results in a contradiction. The current branch dies out and gets marked with C. **6.** DPLL backtracks to the last split variable (x), inverts it (F) and creates a new edge. **7.** Same as step 4. **8.** The propagation of the unitary clauses eliminates all the clauses. A solution S is found and the instance is satisfiable. For an unsatisfiable instance, unsatisfiability is proven when backtracking (see step 6) is not possible anymore since all split variables have already been inverted. In this case, all the nodes in the final search tree have two descendent edges and all branches terminate by a contradiction C.

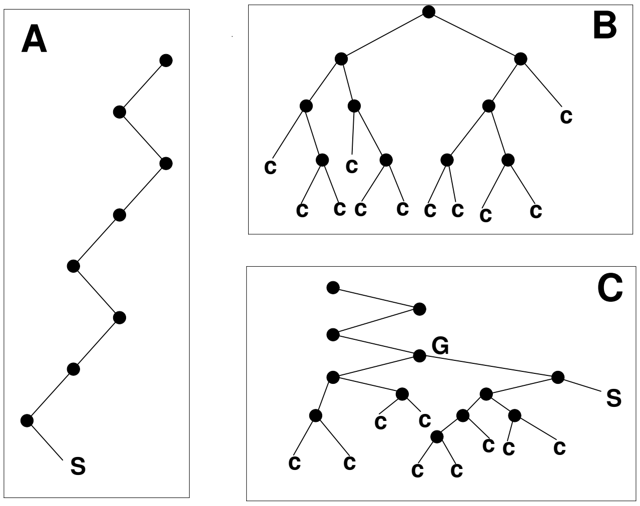

Fig. 12. Types of search trees generated by the DPLL solving procedure. Nodes (black dots) stand for the choices of variables made by the heuristic, and edges between nodes denote the elimination of unitary clauses. **A.** *simple branch:* the algorithm finds easily a solution without ever backtracking. **B.** *dense tree:* in the absence of solution, the algorithm builds a "bushy" tree, with many branches of various lengths, before stopping. **C.** *mixed case, branch + tree:* if many contradictions arise before reaching a solution, the resulting search tree can be decomposed in a single branch followed by a dense tree. The junction G is the highest backtracking node reached back by DPLL.

izing the evolution of the 2+p-SAT instance solved by DPLL does not define a line any longer, but rather a patch, or cloud of points with a finite extension in the phase diagram of Figure 10.

The value of ω can be calculated analytically as a function of the initial ratio, see next section. The main idea is to monitor the growth of the dominant, that is, most numerous branches in the tree until their extinction resulting from the onset of contradictions. Results are in excellent quantitative agreement with the numerical experiments of Figure 2[37]. Main qualitative features are: ω is positive in the whole unsat phase $\alpha > \alpha_c$, indicating that computational complexity is exponential in this region; ω is maximal at the threshold and gets smaller and smaller as α increases (as the instance is more and more over-constrained, it becomes easier and easier to detect a group of mutually incompatible clauses).

The intermediate region $\alpha_L < \alpha < \alpha_C$ juxtaposes the two previous behaviors, see tree Figure 12C. The branch trajectory, started from the point $(p = 1, \alpha)$ corresponding to the initial (and satisfiable) 3-SAT instance, crosses the critical line $\alpha_c(p)$ at some point G (Figure 10). The algorithm then enters the unsat phase and generates 2+p-SAT instances with no solution. A dense subtree, that DPLL has to go through entirely, forms beyond G (Figure 10). There

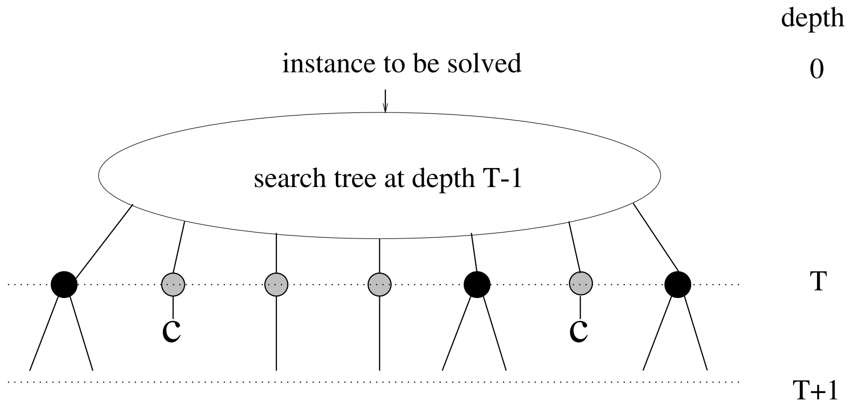

Fig. 13. Imaginary, parallel growth process of an unsat search tree used in the theoretical analysis of the computational complexity. Variables are fixed through unit-propagation, or the splitting heuristic as in the DPLL procedure, but branches evolve in parallel. T denotes the depth in the tree, that is the number of variables assigned by DPLL along each (living) branch. At depth T, one literal is chosen on each branch among 1-clauses (unit-propagation, grey circles not represented on Figure 1), or 2,3-clauses (split, black circles as in Figure 1). If a contradiction occurs as a result of unit-propagation, the branch gets marked with C and dies out. The growth of the tree proceeds until all branches carry C leaves. The resulting tree is identical to the one built through the usual, sequential operation of DPLL.

is no need for DPLL to backtrack above G (Figure 12C), since nodes immediately above G are located in the sat phase and carry 2+p-SAT instances with solutions. DPLL will eventually reach a solution (point S in Figure 12C). The corresponding branch shown Figure 10 is highly non typical and does not contribute to the complexity. The resolution complexity of the initial 3-SAT instance thus reduces, to the exponential order in N, to the complexity of the critical unsatisfiable 2+p-SAT instance located in G, and can be analytically predicted from our theory (next section). In this intermediary region, random 3-SAT instances are exponentially hard to solve.

5.4 Growth of the search tree and resolution time.

In this section, we briefly present the calculation of the size of resolution trees in the hard, exponential regime. Details of the computation may be found in [38].

DPLL solving procedure builds up the search tree in a sequential way, adding more and more nodes, branches up to the completion of the tree. As shown Figure 11, each node corresponds to an instance of the 2+p-SAT model, with defining parameters p, α. We have imagined a different building up, that results in the same complete tree but can be mathematically analyzed: the tree grows in parallel, layer after layer, with two branches simultaneously growing out of each node (Figure 13). Each branch obeys DPLL evolution rules (choice of variables, simplification of 1-clauses, and halt if a contradiction arises), such that the final tree is identical to the obtained through the original, sequential process. At each instant t of the building process, the tree consists of many branches, the extremities of which have distinct characteristic parameters. Let us consider the histogram $B(p, \alpha; t)$ of the values of these parameters. This histogram grows exponentially with the size N of the instance. It is thus convenient to consider its logarithm $\Omega(p, \alpha; t) = \frac{1}{N} \log_2 B(p, \alpha; t)$, which defines a surface in the three-dimensional space p, α, Ω.

The addition of a new layer to the search tree changes the characteristic parameters of the existing branches, and leads to the appearance, or halt of other branches. As a result, the search tree evolution can be recast in terms of a out-of-equilibrium growth process of a bidimensional surface Ω, where the "time" t is simply the depth of the tree. The equation of growth reads,

$$\frac{\partial \Omega}{\partial t} = H \left[\frac{\partial \Omega}{\partial p}, \frac{\partial \Omega}{\partial \alpha}, p, \alpha, t \right] \qquad , \tag{29}$$

where H depends upon the heuristic rule used by DPLL. The precise expression of H for GUC heuristic may be found in [38]. The number of branches $B(t)$ at depth t is equal, to the dominant order in N, to $2^{N\,\Omega^*(t)}$ where $\Omega^*(t)$ is the height of the top of the surface. The growth is interrupted at depth t_a, when contradictions appear at the extremities of all dominant branches. The final number of branches, *i.e.* of leaves C (Figure 12) is $B(t_a) = 2^{N\,\Omega^*(t_a)}$. It can be checked on Figure 12B that, for any complete tree, the number of leaves is simply equal to the number of nodes plus one. In the large N limit, the knowledge of $\Omega^*(t_a)$ gives a direct access to ω.

Acknowledgements — R.M. is partly funded by the ACI Jeunes Chercheurs "Optimisation combinatoire et verres de spins quantiques" from the Ministère de la Recherche.

References

[1] O. Martin, R. Monasson, R. Zecchina, Statistical mechanics methods and phase transitions in optimization problems, Theoretical Computer Science 265 (2001) 3–67.

[2] M. R. Garey, D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman, New York, 1979.

[3] C. H. Papadimitriou, Computational Complexity, Addison–Wesley, 1994.

[4] S. Cook, The complexity of theorem–proving procedures, in: Proc. 3rd Ann. ACM Symp. on Theory of Computing, Assoc. Comput. Mach., New York, 1971, p. 151.

[5] T. H. Eds., B. A. Huberman, C. Williams, Issue 1-2, special issue on phase transitions, Artificial Intelligence 81.

[6] D. Mitchell, B. Selman, H. Levesque, Hard and easy distributions of sat problems, in: Proc. of Am. Assoc. for Artif. Intell. AAAI-92, 1992, pp. 456–465.

[7] A. Goerdt, A threshold for unsatisfiability, J. Comput. System Sci. 53 (1996) 469.

[8] V. Chvàtal, B. Reed, Mick gets some (the odds are on his side), in: Proc. 33rd IEEE Symp. on Foundations of Computer Science, 1992, p. 620.

[9] S. Kirkpatrick, B. Selman, Critical behaviour in the satisfiability of random boolean expressions, Science 264 (1994) 1297.

[10] J. Franco, Results related to threshold phenomena research in satisfiability: lower bounds, Theoretical Computer Science 265 (2001) 147–158.

[11] D. Achlioptas, Lower bounds for random 3-sat via different equations, Theoretical Computer Science 265 (2001) 159–185.

[12] O. Dubois, Upper bounds on the satisfiability threshold, Theoretical Computer Science 265 (2001) 187–198.

[13] R. Monasson, R. Zecchina, Statistical mechanics of the random K-Sat problem, Phys. Rev. E 56 (1997) 1357–1361.

[14] A. Broder, A. Frieze, E. Upfal, On the satisfiability and maximum satisfiability of random 3-cnf formulas, Proc. 4th Annual ACM-SIAM Symp. on Discrete Algorithms (1993) 322.

[15] Y. Boufkhad, O. Dubois, Length of prime implicants and number of solutions of random cnf formulas, Theoret. Comput. Sci. 215 (1999) 1–30.

[16] R. Monasson, R. Zecchina, Entropy of the K-satisfiability problem, Phys. Rev. Lett. 76 (1996) 3881–3885.

[17] G. Biroli, R. Monasson, M. Weigt, A variational description of the ground state structure in random satisfiability problems, Euro. Phys. J. B 14 (2000) 551.

[18] R. Monasson, Optimization problems and replica symmetry breaking in finite connectivity spin glasses, J. Phys. A 31 (1998) 513.

[19] J. R. Banavar, D. Sherrington, N. Sourlas, Graph bipartitioning and statistical mechanics, J. Phys. A Lett. 20 (1987) L1–L8.

[20] M. Mézard, G. Parisi, Mean-field theory of randomly frustrated systems with finite connectivity, Europhys. Lett. 3 (10) (1987) 1067.

[21] L. Viana, A. J. Bray, Phase diagrams for dilute spin-glasses, J. Phys. C 18 (1985) 3037–3051.

[22] I. Kanter, H. Sompolinsky, Mean-field theory of spin-glasses with finite coordination number, Phys. Rev. Lett. 58 (1987) 164.

[23] C. D. Dominicis, Y. Goldschmidt, Replica symmetry breaking in finite connectivity systems: a large connectivity expansion at finite and zero temperature, J. Phys. A 22 (1989) L775.

[24] P. Mottishaw, C. D. Dominicis, On the stability of randomly frustrated systems with finite connectivity, J. Phys. A 20 (1987) L375.

[25] C. D. Dominicis, P. Mottishaw, Replica symmetry breaking in weak connectivity systems, J. Phys. A 20 (1987) L1267.

[26] Y. Goldschmidt, P. Lai, The finite connectivity spin glass: investigation of replica symmetry breaking of the ground state, J. Phys. A 23 (1990) L775.

[27] M. Mézard, G. Parisi, M. A. Virasoro (Eds.), Spin Glass Theory and Beyond, World Scientific, Singapore, 1987.

[28] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, L. Troyansky, Computational complexity from 'characteristic' phase transitions, Nature 400 (1999) 133–137.

[29] D. Achlioptas, L. Kirousis, E. Kranakis, D. Krizanc, Rigorous results for random (2+p)-sat, Theoretical Computer Science 265 (2001) 109–130.

[30] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, L. Troyansky, 2+p-sat: Relation of typical-case complexity to the nature of the phase transition, Random Structures and Algorithms 3 (1999) 414.

[31] B. Bollobàs, C. Borgs, J. Chayes, J. H. Kim, D. Wilson, The scaling window of the 2-sat transition, Random Structures and Algorithms 2000, in press.

[32] D. Wilson, The empirical values of the critical k-sat exponents are wrong 2000 preprint arXiv:math/0005136.

[33] M. Leone, F. Ricci-Tersenghi, R. Zecchina, Phase coexistence and finite size scaling in random combinatorial problems, ICTP preprint, 2000.

[34] B. Selman, H. Kautz, B. Cohen, Local search strategies for satisfiability testing, in: Proceedings of DIMACS, 1993, p. 661.

[35] M.-T. Chao, J. Franco, Probabilistic analysis of two heuristics for the 3-satisfiability problem, SIAM Journal on Computing 15 (1986) 1106–1118.

[36] M.-T. Chao, J. Franco, Probabilistic analysis of a generalization of the unit-clause literal selection heuristics for the k-satisfiability, Information Science 51 (1990) 289.

[37] S. Cocco, R. Monasson, Trajectories in phase diagrams, growth processes and computational complexity: how search algorithms solve the 3-satisfiability problem, Phys. Rev. Lett. 86 (2001) 1654.

[38] S. Cocco, R. Monasson, Analysis of the computational complexity of solving random satisfiability problems using branch and bound search algorithms, Eur. Phys. J. B 22 (2001) 505.