2+ P-SAT: Relation of Typical-Case Complexity to the Nature of the Phase Transition

Rémi Monasson[1], Riccardo Zecchina[2], Scott Kirkpatrick[3], Bart Selman[4], Lidror Troyansky[5]

[1] *Laboratoire de Physique Théorique de l'ENS, Paris, France*

[2] *International Centre for Theoretical Physics, Trieste, Italy*

[3] *IBM, Thomas J. Watson Research Center, Yorktown Heights, NY*

[4] *Computer Science Dept., Cornell University, Ithaca, NY*

[5] *Computer Science Dept., Hebrew University, Jerusalem, Israel*

Heuristic methods for solution of problems in the NP-Complete class of decision problems often reach exact solutions, but fail badly at "phase boundaries," across which the decision to be reached changes from almost always having one value to almost always having a different value. We report an analytic solution and experimental investigations of the phase transition that occurs in the limit of very large problems in K-SAT. The nature of its "random first-order" phase transition, seen at values of K large enough to make the computational cost of solving typical instances increase exponentially with problem size, suggests a mechanism for the cost increase. There has been evidence for features like the "backbone" of frozen inputs which characterizes the UNSAT phase in K-SAT in the study of models of disordered materials, but this feature and this transition are uniquely accessible to analysis in K-SAT. The random first-order transition combines properties of the 1st order (discontinuous onset of order) and 2nd order (with power law scaling, e.g. of the width of the critical region in a finite system) transitions known in the physics of pure solids. Such transitions should occur in other combinatoric problems in the large N limit. Finally, improved search heuristics may be developed when a "backbone" is known to exist.

Constraint satisfaction, the automated search for a configuration of a complex system which satisfies a set of rules or inequalities, is often difficult, and occurs widely in practice. The simplest example of a constraint satisfaction problem, K-SAT [1,4], is commonly used as a testbed for heuristic algorithms intended for wider use and was the first problem proved to be in the complexity class NP-Complete [5,6], in which the worst case instances are believed to always require computing effort exponential in N, the number of input degrees of freedom. In random K-SAT, the system parameters are a string of N bits, and the rules to be satisfied are a set of M clauses. If the string consists of bits $\{x_i = 0,1\}_{i=1,\dots,N}$, we construct an instance of K-SAT by first randomly choosing K distinct possible indices i and then, for each of them, a literal z_i (i.e. the corresponding x_i or its negation $\bar{x}_i$ with equal probability). A clause C is defined as the logical OR of the K literals chosen. Next, we repeat this process to obtain M independently chosen clauses $\{C_\ell\}_{\ell=1,\dots,M}$ and ask for all of them to be true at the same time, i.e. we take the logical AND of the M clauses. This gives a formula F, which may be written as

$$F = \bigwedge_{\ell=1}^{M} C_\ell = \bigwedge_{\ell=1}^{M} \left(\bigvee_{i=1}^{K} z_i^{(\ell)} \right) \quad , \tag{1}$$

where $\bigwedge$ and $\bigvee$ stand for the logical AND and OR operations respectively. An assignment of the $\{x_i\}$'s satisfying all clauses is a solution of the K–SAT problem. If no such assignment exists, F is unsatisfiable. The formulae F constructed at random keeping the ratio $\alpha = M/N$ constant as $M, N \to \infty$ provide a natural ensemble of test problems, with α characterizing whether the F are typically under- or over-constrained.

The value of K is important. 2–SAT can be solved by a linear time algorithm [7]. There is a critical value of α, $\alpha_c(2) = 1$, below which the likelihood of an F being UNSAT vanishes in the limit $N \to \infty$, and above which it goes to 1. For $K \geq 3$, K–SAT is NP-Complete and rigorous results are few. Computer experiments [8,9] on K-SAT for $K = 2, 3$ and higher have located the phase transition and provided critical exponents for the sharpening of the critical region which occurs with increasing sample size N.

The technique used, finite-size scaling, will be discussed in more detail below. It has recently been proven [10] that certain monotonic properties (such as SAT) in combinatorial ensembles do have sharp thresholds. "Sharp" means in our case that for any $\alpha < \alpha_c(K, N)$, the probability that a formula in K-SAT can be satisfied goes to 1 as $N \to \infty$, while for any

$\alpha > \alpha_c(K, N)$ this probability tends to 0. While Friedgut's result leaves open the question of whether $\alpha_c(K, N)$ has a limit as $N \to \infty$, experiment suggests that this is the case for K-SAT.

One reason for the recent interest in the threshold is the growing recognition that "easy-hard-easy" patterns of computational cost are characteristic of heuristic algorithms developed to solve problems whose cost in the worst case increases exponentially with problem size N, and that the hardest instances occur near [11] phase boundaries like α_c [12–14].

There is a strong analogy between these problems and the properties of disordered materials, alloys or even glasses, studied by constructing models whose energy function expresses the constraints [15]. Strongly disordered models with conflicting interactions similar to the randomly negated literals in K–SAT are known as "spin glasses" [16]. Fu and Anderson [2] first conjectured that spin glasses are the models underlying NP-Complete decision problems and vice-versa. They cite the example of weighted graph partitioning, which is equivalent to an Ising or Potts spin glass.

A technique of calculating expectation values of observables in random many-parameter systems, called the replica method [17] predicts that ordering of a new type is possible in the presence of microscopic randomness. In the replica method, the calculation of the average over the disorder leads to an effective energy or cost function which describes many identical copies of one instance of the model system with the dynamical variables (usually Ising spins taking on values +1 or -1) in different replicas coupled by some non-linear function. The onset of ordering can be identified by the fact that a single stable state occurs in multiple replicas. This corresponds to a physical ground state that is highly irregular in structure due to the randomness of the problem. A more subtle kind of ordering (called Replica Symmetry-Breaking) occurs when distinct stable states of this sort are found in different subsets of the replicas, signalling that there may be infinitely many distinct ground states with energies infinitesimally close to the true ground state. The extent of or absence of this new kind of order can be quantified by an "order parameter," which in general emerges from the replica formulation. We describe both types of result in sections I and II. Details of the calculations for K-SAT are given in section II. While the replica procedures are not rigorous, certain steps can be proven, and some results have been verified by other means.

We discuss these issues in section II.

Spin glass models with realistic connectivity are difficult to explore experimentally, either on real substances or by computer simulation of models in thermal equilibrium at some finite temperature. Experimental study of the "easy-hard-easy" phase transitions in combinatorics is more tractable. Although these are spin glass models, the properties of interest are ground state properties, and a large body of model-specific heuristics exists, which gives powerful means of exploring these ground states. We have previously applied replica methods and determined characteristics of the 3-SAT transition [18–20]. In section III, we report additional results which provide new insights, hopefully of use to both fields. Note, however, that the methods of statistical physics predict the most probable, or typical, behavior of a system with many degrees of freedom, so we shall be describing the typical complexity of K–SAT, not its worst case.

I. MIXTURES OF $K = 2$ AND 3: OVERVIEW OF RESULTS

In order to understand what occurs between $K = 2$ and $K = 3$, we have studied [21] formulae containing mixtures of 2- and 3-clauses: consider a random formula with M clauses, of which $(1 - p)M$ contain two literals and pM contain 3 literals, with $0 \leq p \leq 1$. This "$2 + p$–SAT" model smoothly interpolates between 2–SAT ($p = 0$) and 3–SAT ($p = 1$). The problem is NP–complete, since any instance of the model for $p > 0$ contains a sub-formula of 3-clauses. But our interest here is in the complexity of "typical" problem instances.

We seek $\alpha_c(2 + p)$, the threshold ratio M/N of the above model at fixed p. We know $\alpha_c(2) = 1$ and $\alpha_c(3) \simeq 4.27$. F cannot be almost always satisfied if the number of 2–clauses (respectively 3–clauses) exceeds N (resp. $\alpha_c(3)N$). As a consequence, the critical ratio must be bounded by $\alpha_c(2 + p) \leq \min\left(\frac{1}{1-p}, \frac{\alpha_c(3)}{p}\right)$.

The $2 + p$–SAT model can be mapped onto a diluted spin glass model with N spins S_i: $S_i = 1$ if the Boolean variable x_i is true, $S_i = -1$ if x_i is false. Then, to any configuration is associated an energy E, or cost-function, equal to the number of clauses violated. Random couplings between the spins are induced by the clauses. *The most important result of the replica approach [19] is the emergence, in the large N, M limit and at fixed p and α, of order parameters describing the statistics of optimal assignments, which minimize the number of* *violated clauses.* In this section, we give an overview of the results from statistical mechanics. The next section gives a more detailed description of the analysis.

Consider an instance of the $2 + p$-SAT problem. We use the $\mathcal{N}_{GS}$ ground state configurations to define

$$m_i = \frac{1}{\mathcal{N}_{GS}} \sum_{g=1}^{\mathcal{N}_{GS}} S_i^g \tag{2}$$

the average value of spin S_i over all optimal configurations. Clearly, m_i ranges from -1 to $+1$ and $m_i = -1$ (respectively $+1$) means that the corresponding Boolean variable x_i is always false (resp. true) in all ground states. The distribution $P(m)$ of all m_i gives the microscopic structure of the ground states. The accumulation of magnetizations m around ± 1 represents a "backbone" of almost completely constrained variables, whose logical values cannot vary from solution to solution, while the center of the distribution $P(m \simeq 0)$ describes weakly constrained variables. The threshold α_c will coincide with the appearance of an extensive backbone density of fully constrained variables x_i, with a finite probability weight at $m = \pm 1$. A simple argument shows that the backbone must vanish when $\alpha < \alpha_c$. Consider adding one clause to a SAT formula found below α_c. If there is a finite fraction of backbone spins, there will be a finite probability that the added clause creates an UNSAT formula, which cannot occur.

For $\alpha < \alpha_c$, the solution exhibits a simple symmetry property, usually referred to as Replica Symmetry (RS), which leads to an order parameter which is precisely the magnetization distribution $P(m)$ defined above. An essential qualitative difference between 2-SAT and 3-SAT is the way the order parameter $P(m)$ changes at the threshold. This discrepancy can be seen in the fraction $f(K, \alpha)$ of Boolean variables which become fully constrained, at and above the threshold. As said above, $f(K, \alpha)$ is identically null below the threshold. For 2-SAT, $f(2, \alpha)$ becomes strictly positive above $\alpha_c = 1$ and is continuous at the transition : $f(2, 1^-) = f(2, 1^+) = 0$. On the contrary, $f(3, \alpha)$ displays a discontinuous change at the threshold : $f(3, \alpha_c^-) = 0$ and $f(3, \alpha_c^+) = f_c(3) > 0$.

While for the continuous transitions, the exact value of the threshold can be derived within the RS scheme, for the discontinuous case the RS prediction gives only upper bounds. The exact value of the threshold can be predicted only by a proper choice of the order parameter at the transition point, i.e. by a more general symmetry breaking scheme, a problem which is still open. However, the predictions of the RS equations, such as the number of solutions, remain valid up to α_c, and the RS prediction for the nature of the threshold should be qualitatively correct.

For the mixed $2+p$–SAT model, the key issue is therefore to understand how a discontinuous 3–SAT-like transition may appear when increasing p from zero up to one and how it affects the computational cost of finding solutions near threshold. Applying the method of ref. [19], we find for $p < p_0$ ($p_0 = 0.41$), there is a **continuous** SAT/UNSAT transition at $\alpha_c(2+p) = \frac{1}{1-p}$. This has recently been verified by rigorous analysis up to $p = 0.4$ [22]. The RS theory appears to be correct for $\alpha < \alpha_c(2+p)$, and thus gives both the critical ratio and the typical number of solutions, as in the $K = 2$ case. The SAT/UNSAT transition should coincide with a replica symmetry breaking transition, as discussed in [19]. So, for $p < p_0$, the model shares the characteristics of random 2–SAT.

For $p > p_0$, the transition becomes **discontinuous** and the RS transition gives an upper bound for the true $\alpha_c(2+p)$. The RS theory correctly predicts a discontinuous appearance of a finite fraction of fully constrained variables which jumps from 0 to f_c when crossing the threshold $\alpha_c(2+p)$. However, both values of $f_c(2+p)$ and α_c are slightly overestimated, e.g. for $p = 1$, $\alpha_c^{RS}(3) \simeq 4.60$ and $f_c^{RS}(3) \simeq 0.6$ whereas experiments give $\alpha_c(3) \simeq 4.27$ and $f_c(3) \sim 0.5$. A replica symmetry breaking theory will be necessary to predict these quantities. For $p > p_0$, the random 2+p–SAT problem shares the characteristics of random 3–SAT.

This transition differs from phase transitions in most ordered materials in that the ground state is highly degenerate at α_c. The entropy, i.e. the logarithm (base 2) of the typical number of optimal solutions, can be computed exactly within the RS scheme for any p and $\alpha < \alpha_c(2+p)$. The entropy at the transition point decreases as a function of p, from 0.56 for $p = 0$ to 0.13 for $p = 1$, as plotted in Fig. 1.

II. STATISTICAL MECHANICS ANALYSIS OF THE 2+P–SAT MODEL

In this section we describe the analytical calculation of the typical ground state properties of the 2+P–SAT model using the replica method. A brief discussion concerning rigorous results and prospects for making the replica results rigorous is also included.

A. The energy-cost function

The logical variables x_i can be represented by N binary variables S_i, called spins, through the one-to-one mapping $S_i = -1$ (respectively $+1$) if x_i is false (resp. true). We then encode the random clauses into a $M \times N$ 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. Note that $\sum_{i=1}^{N} C_{\ell i} S_i$ gives the the net number of literals satisfying 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 $\mathbf{S}$.

$$E[\mathbf{C}, \mathbf{S}] = \sum_{\ell=1}^{(1-p)M} \delta\left[\sum_{i=1}^{N} C_{\ell i} S_i ; -2\right] + \sum_{\ell=(1-p)M+1}^{M} \delta\left[\sum_{i=1}^{N} C_{\ell i} S_i ; -3\right] \quad , \tag{3}$$

where $\delta[.;.]$ denotes the Kronecker function, which is 1 if its arguments are equal, zero otherwise. The minimum (or ground state) $E[\mathbf{C}]$ of $E[\mathbf{C}, \mathbf{S}]$, the lowest number of violated clauses that can be achieved by the best possible logical assignment [19], is a random variable which becomes totally concentrated around its mean value $\ll E[\mathbf{C}] \gg$ in the large size limit [23]. The latter is accessible through the knowledge of the averaged logarithm of the generating function

$$Z[\mathbf{C}] = \sum_{\mathbf{S}} \exp\left(-E[\mathbf{C}, \mathbf{S}]/T\right) \tag{4}$$

since $\ll E[\mathbf{C}] \gg = -T \ll \log Z[\mathbf{C}] \gg + O(T^2)$ when the auxiliary parameter T is eventually sent to zero. Since $\ll E[\mathbf{C}] \gg = 0$ in the SAT region and is positive in the UNSAT phase, calculating $\ll E[\mathbf{C}] \gg$ locates $\alpha_c(K)$.

B. The average over the disorder

The calculation of the average value of the logarithm of Z from Eq. (4) is an awkward one. To circumvent this difficulty, we compute the n^{th} moment of Z for integer-valued n and perform an analytical continuation to real n to exploit the identity $\ll Z[\mathbf{C}]^n \gg = 1 + n \ll \log Z[\mathbf{C}] \gg + O(n^2)$. The n^{th} moment of Z is obtained by replicating n times the sum over the spin configuration $\mathbf{S}$ and averaging over the clause distribution [19]

$$\ll Z[\mathbf{C}]^n \gg = \sum_{\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n} \ll \exp\left(-\sum_{a=1}^{n} E[\mathbf{C},\mathbf{S}^a]/T\right) \gg \quad . \tag{5}$$

The average over the clauses can be performed because their probability distributions are uncorrelated. We obtain

$$\ll Z[\mathbf{C}]^n \gg = \sum_{\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n} (\zeta_2[S^a])^{(1-p)M} \ (\zeta_3[S^a])^{pM} \qquad , \tag{6}$$

where each factor is defined by $(K=2,3)$

$$\zeta_K[S^a] = \ll \exp\left(-\frac{1}{T}\sum_{a=1}^{n}\delta\left[\sum_{i=1}^{N} C_i S_i^a; -K\right]\right) \gg \qquad . \tag{7}$$

We stress that $\ll . \gg$ now denotes the unbiased average over the set of $2^K \binom{N}{K}$ vectors of N components $C_i = 0, \pm 1$ and of squared norm equal to K.

Resorting to the identity,

$$\delta\left[\sum_{i=1}^{N} C_i S_i^a; -K\right] = \prod_{i/C_i \neq 0} \delta\left[S_i^a; -C_i\right] \qquad , \tag{8}$$

one may carry out the average over disorder in eq.(7) to obtain

$$\zeta_K[S^a] = \frac{1}{2^K} \sum_{C_1,\ldots,C_K=\pm 1} \frac{1}{N^K} \sum_{i_1,\ldots,i_K=1}^{N} \exp\left\{-\frac{1}{T}\sum_{a=1}^{n}\prod_{\ell=1}^{K}\delta\left[S_{i_\ell}^a; -C_\ell\right]\right\} \quad , \tag{9}$$

to the largest order in N.

It is crucial to remark that $\zeta_K[S^a]$ in (9) depends on the $n \times N$ spins only through some 2^n quantities $c(\sigma)$ labelled by the vectors σ with n binary components; $c(\sigma)$ equals the number (divided by N) of labels i such that $S_i^a = \sigma^a$, $\forall a = 1,\ldots,n$ [24]. Indeed, one can rewrite $\zeta_K[S^a] = \zeta_K[c]$ with

$$\zeta_K[c] = \frac{1}{2^K} \sum_{C_1,\ldots,C_K=\pm 1} \sum_{\vec{\sigma}_1,\ldots,\vec{\sigma}_K} c(-C_1\,\vec{\sigma}_1)\ldots c(-C_K\,\vec{\sigma}_K) \exp\left\{-\frac{1}{T}\sum_{a=1}^{n}\prod_{\ell=1}^{K}\delta\left[\sigma_\ell^a; 1\right]\right\} \quad . \tag{10}$$

Notice that $c(\vec{\sigma}) = c(-\vec{\sigma})$ due to the even distribution of the disorder C.

Introducing the *effective energy* function,

$$\begin{aligned} E_{eff}[c] = {} & \alpha\,(1-p)\ \ln\left[\sum_{\vec{\sigma},\vec{\tau}} c(\vec{\sigma})c(\vec{\tau})\ \exp\left(-\frac{1}{T}\sum_{a=1}^{n}\delta[\sigma_a;1]\delta[\tau_a;1]\right)\right] \\ & +\alpha\, p\ \ln\left[\sum_{\vec{\sigma},\vec{\tau},\vec{\omega}} c(\vec{\sigma})c(\vec{\tau})c(\vec{\omega})\ \exp\left(-\frac{1}{T}\sum_{a=1}^{n}\delta[\sigma_a;1]\delta[\tau_a;1]\delta[\omega_a;1]\right)\right] \qquad , \end{aligned} \tag{11}$$

we may rewrite the n^{th} moment of the generating function Z (5) as

$$\ll Z^n \gg = \int \prod_{\vec{\sigma}} dc(\vec{\sigma}) e^{-E_{eff}[c]} \sum_{\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^n} \prod_{\vec{\sigma}} \delta\left(c(\vec{\sigma}) - \frac{1}{N}\sum_{i=1}^{N}\prod_{a=1}^{n}\delta[S_i^a;\sigma^a]\right) \quad . \tag{12}$$

The sum over the spins in the last term of the above equation can be computed, and gives rise to a combinatorial factor

$$\frac{N!}{\prod_{\vec{\sigma}}(Nc(\vec{\sigma}))!} = \exp\left(-N\sum_{\vec{\sigma}} c(\vec{\sigma}) \ln c(\vec{\sigma})\right) \quad , \tag{13}$$

to the leading exponential order in N. As a consequence, the n^{th} moment of Z using the Laplace method is $\ll Z^n \gg \simeq \exp(N\ F_{max})$ where F_{max} is the maximum over all possible cs of the functional [19]

$$F[c] = -\sum_{\vec{\sigma}} c(\vec{\sigma}) \log c(\vec{\sigma}) - T E_{eff}[c] \quad , \tag{14}$$

with the constraint

$$\sum_{\vec{\sigma}} c(\vec{\sigma}) = 1. \tag{15}$$

C. The replica symmetric theory

The optimisation conditions over $F[c]$ provide 2^n coupled equations for the cs. Notice that F is a symmetric functional, invariant under any permutation of the replicas a. A maximum may thus be sought in the so-called replica symmetric (RS) subspace of dimension $n+1$ where $c(\sigma)$ is left unchanged under the action of the symmetric group. Within the RS subspace, the occupation fractions may be conveniently expressed as the moments of a probability distribution $P(m)$ over the range $-1 \leq m \leq 1$ [19].

$$c(\sigma_1, \sigma_2, \ldots, \sigma_n) = \int_{-1}^{1} dm\ P(m) \prod_{a=1}^{n}\left(\frac{1+m\sigma^a}{2}\right) \quad . \tag{16}$$

$P(m)$ is the distribution of Boolean magnetizations previously introduced in the paper.

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 lead to a self-consistent functional equation for the order parameter $P(m)$.

In the limit of interest $T \to 0$, in order to properly describe the accumulation of the Boolean magnetization to the border of its domain ($m \in [-1,1]$), it is convenient to introduce the rescaled variables z, implicitly defined by the relation $m = \tanh(z/T)$. Calling $R(z)$ the probability distribution of the zs, we obtain

$$R(z) = \int_{-\infty}^{\infty} \frac{du}{2\pi} \cos(uz) \exp\Big\{ -\alpha(1-p) + 2\alpha(1-p) \int_0^\infty dz_1 R(z_1) \cos(u \min(1,z_1)) \\ -\frac{3}{4}\alpha p + 3\alpha p \int_0^\infty dz_1 dz_2 R(z_1) R(z_2) \cos(u \min(1,z_1,z_2)) \Big\} \qquad . \tag{17}$$

As discussed in detail in ref. [19], the above type of equations admit an infinite sequence of rapidly converging exact solutions of the form

$$R(z) = \sum_{\ell=-\infty}^{\infty} r_\ell \, \delta\left(z - \frac{\ell}{q}\right) \qquad . \tag{18}$$

In the above expression, $1/q$ is the resolution of the rescaled variable z which eventually goes to zero. Equation (17), leads to the following set of coupled equations for the coefficients r_ℓ's

$$r_\ell = \int_0^{2\pi} \frac{d\theta}{2\pi} \cos(\ell\theta) \exp\left(\sum_{j=1}^{q} \gamma_j (\cos(j\theta) - 1) \right) \tag{19}$$

for all $\ell = 0, \ldots, q-1$ where

$$\gamma_j/\alpha = 2\,(1-p)\, r_j + 3\, p\, r_j \left(1 - r_0 - 2\sum_{\ell=1}^{j-1} r_\ell - r_j\right) \quad , \; \forall j = 1, \ldots, q-1$$

$$\gamma_q/\alpha = (1-p)\left(1 - r_0 - 2\sum_{\ell=1}^{j-1} r_\ell\right) + \frac{3}{4}\, p \, \left(1 - r_0 - 2\sum_{\ell=1}^{j-1} r_\ell\right)^2 \qquad . \tag{20}$$

By looking for the value of α at which the internal energy (11) becomes positive, we are able to recover the results discussed previously in the text. For $p < p_0$, the transition is continuous and the solution of the equations up to α_c is simply $r_0 = 1, r_\ell = 0$ ($l = 1, ..., q$). At α_c, the coefficients r_ℓ become continuously positive. For $p > p_0$, the coefficients r_ℓ jump discontinuously to a finite value beyond α_c. It follows that in order to find the point where the discontinuous transition first takes place, one should look, within the RS scheme, for the point p_0 at which the derivative of the order parameters r_ℓ at α_c diverge.

We may expand the saddle point equations (19,20) to the second order in the parameters r_ℓ and $s \equiv 1 - r_0$. We find

$$r_\ell = \alpha(1-p) r_\ell + \frac{3}{2}\,\alpha\, p\, r_\ell\,\left(s - 2\sum_{k=1}^{\ell-1} r_k - r_\ell\right) - \alpha^2\,(1-p)^2\, r_\ell\, s + \frac{1}{2}(1-p)^2\alpha^2 \sum_{j=1}^{\ell-1} r_j r_{\ell-j} + \tag{21}$$

$$+ (1-p)^2\alpha^2 \sum_{j=1}^{q-\ell-1} r_j r_{\ell+j} + \frac{1}{2}(1-p)^2\alpha^2 r_{q-\ell}(s - 2\sum_{k=1}^{\ell-1} r_k) \quad , \ (\ell = 1, ..., q-1)$$

and, for $\ell = 0$,

$$s = \alpha(1-p)s + 3\alpha p \left[\sum_{j=1}^{q-1} r_j(s - 2\sum_{\ell=1}^{j-1} r_\ell - r_j) + \frac{1}{4}(s - 2\sum_{\ell=1}^{q-1} r_\ell)^2\right] - \tag{22}$$

$$- \alpha^2(1-p)^2 \left[\sum_{j=1}^{q-1} r_j^2 + (\frac{s}{2} - \sum_{\ell=1}^{q-1} r_\ell)^2\right] - \frac{1}{2}\alpha^2(1-p)^2 s^2 \ .$$

The analysis of the linear terms in the above equations shows that the threshold is given by

$$\alpha_c(2+p) = \frac{1}{1-p} \qquad , \qquad (0 \le p < p_0). \tag{23}$$

Next, we expand around the latter by posing $\alpha = \frac{1}{1-p} + x$, $r_\ell = B_\ell x$ and $s = Ax$. At the critical point p_0, the above quantities $\{B_\ell, A\}$ should diverge in order to have a first order jump when $x \to 0^+$. We then assume that $B_\ell = \lambda_\ell A$, with $\lambda_\ell = O(1)$ and $A \to \infty$, discarding irrelevant $O(x^2)$ corrections to the order parameters. We find q equations for p_0 and λ_ℓ, $\ell = 1, \ldots, q-1$.

$$0 = \frac{3}{4}\,\frac{1-2p_0}{1-p_0} - \sum_{j=1}^{q-1}\lambda_j + \sum_{j=1}^{q-1}\lambda_j^2 + \left(\sum_{j=1}^{q-1}\lambda_j\right)^2 \tag{24}$$

and, for $\ell = 1, \ldots, q-1$,

$$0 = \frac{3}{2}\frac{p_0}{1-p_0}\lambda_\ell \left(1 - 2\sum_{j=1}^{\ell-1}\lambda_j - \lambda_\ell\right) + \frac{1}{2}\sum_{j=1}^{\ell-1}\lambda_j\lambda_{\ell-j} +$$

$$\sum_{j=1}^{q-\ell-1}\lambda_j\lambda_{\ell+j} + \lambda_{q-\ell}\left(\frac{1}{2} - \sum_{j=1}^{q-1}\lambda_j\right) - \lambda_\ell \tag{25}$$

Though we have not been able to find an exact solution to (24,25), the above equations can be easily solved iteratively, leading to a value of $p_0 \simeq 0.41$. A more detailed discussion of the equations for p_0 is given in [20], in which the connection with the calculation of ref. [22] is made explicit by showing that $p_0 = 2/5$ is indeed a lower bound for p_0.

The exactness of the above results depends on the validity of the RS assumption introduced in the functional saddle-point equations. For $p < p_0$, such an assumption turns out to be correct, leading to a threshold value which as been proven to be exact also by other methods [22] For $p > p_0$, the change in the order parameter $P(m)$ (or $R(z)$) at the threshold becomes discontinuous and the solutions of the RS equations account only for an upper bound of the true threshold. The exact solution of the self–consistency equations lies outside the RS subspace to which we have restricted our analysis. The exact determination of the SAT/UNSAT threshold for discontinuous transitions requires the introduction of a more general (and much more complicated) symmetry breaking scheme in the equations, the so called Replica Symmetry Breaking, which embodies the RS subspace as a particular case.

It is worth noting that in the SAT region $\alpha < \alpha_c(p)$, the RS theory is believed to be exact and allows for the estimation of quantities of interest such as the typical number of solutions or the probability distribution of the variables over all solutions. Some rigorous probabilistic results are given in [25].

D. Comments on the replica approach

In the previous paragraphs, replicas are introduced as a trick to compute the average value $\ll \log Z[\mathbf{C}] \gg$ from the integer moments $\ll Z[\mathbf{C}]^n \gg$ of the generating function (4). As long as the number of variables N is finite, $a(n) \equiv \ll Z[\mathbf{C}]^n \gg$ is an analytic function of n and grows less than exponentially at large n, $a(n) < (2^N)^n$. Invoking a theorem of Carlson, $a(n)$ is uniquely known from its values at the nonnegative integers. Therefore the analytic continuation to real $n \to 0$ is unambiguous. However, due to the saddle-point calculation of Section II.B the limit $N \to \infty$ is made first and the analytic continuation requires the introduction of some additional hypotheses.

The most natural continuation scheme, called Replica Symmetry (RS) has recently been shown to be exact at high temperature T [26,27] for the K-SAT model. Though not explicitly proven in [26,27], it is reasonable to think that RS should also hold at $T = 0$ for small ratios of clause per variable α. Indeed, in a simple constraint satisfaction model, the RS hypothesis has been shown [28] to be exact in the range $0 < \alpha < \alpha_c$ giving thus access to the exact value of the threshold α_c [29]. Note that self-consistency criteria of the calculation of the local stability of the saddle-point over $c(\vec{\sigma})$ found in Section II.B are satisfied by the RS hypothesis in this range.

What happens when RS fails, *e.g.* above (respectively in the vicinity of) α_c for 2-SAT (resp. 3-SAT) ? Among all models for which RS fails, the so called Random Energy Model (REM) [30] has been the only one rigorously solved so far. Its exact solution can be reproduced [31] using another Replica Symmetry Broken (RSB) scheme designed by Parisi [17]. Within such a scheme, the analytic continuation is performed by an iterative hierarchical procedure characterized by a closed algebraic structure at each stage of the hierarchy [17]. The known random mean-field models (i.e. models with a complete graph of interaction) appear to be divided in two main classes. A first one for which the complete solution requires an infinite iteration of the Parisi scheme (e.g. the SK model [17]) and a second one for which the first step already provides the correct result. In the latter case, the successive steps of the RSB scheme lead to saddle point equations having as solutions the first step result [31]. It is worth noting here that in the random 3-SAT problem, like several random mean-field models known to exhibit a discontinuous transition and be solvable by the one step RSB ansatz, the energy can be expressed as a sum of products of up to three spins S_i. Therefore, the use of the one-step RSB hypothesis has promise for analyzing the SAT/UNSAT transition of 3-SAT. We expect that there will be differences resulting from the fact that the 3-SAT model involves a sparse graph of interactions. This permits heterogeneous orderings not possible in the mean-field models, for which all degrees of freedom are frozen in the ordered phase.

III. EXPERIMENTAL TESTS

We have run experiments to test the theory between $K = 2$ and $K = 3$, finding thresholds and assigning an exponent ν for the narrowing of the critical region by finite-size scaling. The data obtained for this study is collected and presented in Figs 2a and 2b, which show the fraction of formulae that are unsatisfiable as a function of α for sample sizes from $N = 50$ to the largest practical size for each value of p. To obtain the data in Figs 2a and 2b, we take a sample of formulae with the desired value of p, and for each formula, starting well inside the SAT phase, add clauses until the formula first becomes unsatisfiable. The cumulative distribution of the values of α at which this occurs provides the curves in Figs 2. From

10,000 to 40,000 formulae were studied for each value of N and p shown. For cases with $p > p_0$, our scripts used as their core the TABLEAU implementation of the Davis-Putnam search algorithm [32,33]. For cases with $p < p_0$ a variant called MODOC was used [34]. This adds binary resolution to eliminate 2-clauses by the relation $(p \vee q) \wedge ((\neg q) \vee r) = p \vee r$.

Finite-size scaling of the critical region is done by plotting all quantities against the rescaled variable

$$y = N^{\frac{1}{\nu}}(\alpha - \alpha_c(K,N))/\alpha_c(K,N), \tag{26}$$

which "opens up" the critical region in samples with large N so that data from all sizes collapse onto a single universal curve [8]. If $\alpha_c(K,N)$ were constant, all the curves for a given value of K would pivot about a single point, $\alpha_c(K,\infty)$. This occurs in the finite-size scaling analysis of the random graph ensemble and is a good approximation at large K for K-SAT [8]. But there are significant additional size dependences present for small values of K, as evidenced in Fig. 3, which shows $K = 2$, $(p = 0.0)$ on a fine scale. The successive crossings of pairs of curves for increasing values of N provides a rough measure of $\alpha_c(K,N)$ (e.g. estimate $\alpha_c(K,50)$ as the point where the fraction UNSAT for $N = 50$ crosses the fraction UNSAT for $N = 100$). If we make the ad hoc assumption that the added size dependence is due to the variation with N of $\alpha_c(K,N)$, then we do not have to specify $\alpha_c(K,N)$ for each N. The data reduction required is to choose values of $\alpha_c(K,\infty), \nu$ for which all the transformed curves are parallel, that is, shifted by $(\alpha_c(K,N) - \alpha_c(K,\infty))/\alpha_c(K,\infty)$. An example of such a reduction is shown in Fig 4, for the case $p = 0.0$. Using this methodology, we obtained rescaled curves for all of the data which varied smoothly with p, as shown in Fig. 5.

We find (Fig. 6) good agreement between the observed and predicted values of α_c, with an error which increases slowly from $p = 0.41$ to $p = 1.0$. We also show the bounds obtained by rigorous methods in Fig. 6. Lower bounds are obtained by showing that some analyzable algorithms, such as unit clause propagation [7] find SAT solutions with a finite probability [35]. Upper bounds use the fact that the probability of finding a satisfying assignment is bounded by the expected number of solutions. Refined versions of this argument [36,37] partially eliminate the high degeneracy of some solutions [38]. Both methods have been applied to $(2+p)$–SAT in [22], yielding the dashed and dotted lines plotted in Fig. 6.

In figure 7, we show values of ν obtained from finite-size scaling analysis described above. Below p_0, the exponent ν is roughly constant and equal to 2.8, the value found for 2SAT. This indicates that the critical behavior along the second order transition line in Fig. 6 is dominated by the 2-clauses in the formulae. When we include additional corrections to scaling in $y(\alpha, N)$ and in the probability of UNSAT, following the classic prescription [3], we find that ν may be as large as 3, the value that occurs in the percolation transition for undirected random graphs [39]. The UNSAT phase for $K = 2$ is one in which "constraint loops" become so ubiquitous that almost certainly there is some literal that implies its converse. It is likely that the 2SAT transition results from percolation of these loops, and is in the same universality class as random graph percolation, differing only in the corrections to its scaling behavior.

Above p_0, ν drops rapidly to 1.5. For $K \gg 3$, the values of ν tend to 1.0, a result which can be understood in the "annealed" limit discussed in [19,38].

It is surprising that finite size scaling holds in the presence of discontinuous behavior of the order parameter characterizing the backbone. But this discontinuity is accompanied by smooth behavior of other thermodynamic quantities, e.g. entropy, as first discussed by Gross and Mézard in [31]. First order transitions in pure solids involve two (or a finite number of) phases and do not exhibit critical fluctuations or scaling laws with non-integer exponents. The random first-order transition taking place for $p > p_0$, into an infinite number of distinct ground states, displays features of both first and second orders. This mixed behaviour has also recently been observed in random-field models [40].

Previous work showed that the cost of running the best heuristics, depth-first search algorithms with backtracking [32] increased exponentially with N for any value of α for $K = 3$, with a prefactor that could be mapped into a universal function by plotting it as a function of y [14]. The cost was maximized at $\alpha_{.5}(K, N)$, so we have obtained cost data at this value of α for $p = 0., .2, .4$ and $.6$ over the range of N that could be searched. The plot in Fig. 8a shows that the median cost increases linearly with N for $p < p_0$. It increases dramatically over a smaller range of N for $p > p_0$. Fig. 8b confirms that this increase is exponential already for $p = 0.6$.

Discontinuous nucleation of UNSAT regions due to the breakdown of replica symmetry and the “backbone” of frozen spins conveniently explain the apparent inevitably high cost of heuristic search near the phase boundary. Heuristics which proceed by ”asserting” the possible value of a spin make early mistakes by mis-asserting a backbone spin, and take a long time to backtrack to correct their mistakes. Even if the backbone can be identified before the depth-first search begins, the problem that remains is one of organizing the search over the remaining spins which lie on the boundaries of ”nuclei” or partial solutions to find the lowest energy arrangements of the whole solution, also involving much wasted effort to explore an exponential subspace.

The experiments, shown in Figs. 9a ($K = 2$) and 9b ($K = 3$), confirm that the appearance of the backbone is discontinuous for $K = 3$, and support the prediction of a continuous appearance of the backbone for $K = 2$. Above the threshold, the fraction of frozen spins found in small samples by exhaustive enumeration to locate all ground states is relatively insensitive to N. At and below the threshold, the fraction of frozen spins decreases rapidly with increasing N. While the samples which could be studied are too small to permit extrapolation, the results are consistent with $< f(K, \alpha) >$ vanishing below α_c.

The existence of a “backbone” has previously been reported in the traveling salesman problem [41], with only a few bonds differing in many near-optimal tours. This observation has recently been turned to advantage by heuristics [42] which identify the backbone links and concentrate attention on the small subproblems which remain. This may prove to be a generally valid approach. Efficient means of finding the backbone will be specific to each problem type, but should nonetheless provide a step ahead in algorithm efficiency. Moreover, many worst-case NP-complete problems occurring in practice contain a mix of tractable and intractable constraints. Our results suggest that search algorithms that exploit as much of the tractable structure as possible may in fact scale polynomially in the typical case. In much of the work on search methods in, for example, Artificial Intelligence and Operations Research, one already informally follows the methodology of exploiting tractable problem structure in worst-case intractable problems. However, our hybrid model provides the first formal explanation why such a methodology can work so well in practice. Below a certain threshold fraction of intractable constraints, the overall behavior is dominated by the tractable substructure of the problem, leading to an overall efficient, polynomial time solution method.

---

[1] Email: monasson@lpt.ens.fr

[2] Email: zecchina@ictp.trieste.it

[3] Email: kirk@watson.ibm.com

[4] Email: selman@cs.cornell.edu

[5] Email: lidrort@cs.huji.ac.il

[1] for an introductory discussion see HAYES, B., *American Scientist* **85**, 108 (1996)

[2] FU, Y.-T., and ANDERSON, P. W., in *Lectures in the Sciences of Complexity*, D. Stein (ed.), (Addison-Wesley, 1989), p. 815.

[3] Scaling and critical phenomena have a vast literature. For a masterful summary and review, see FISHER, Michael E., *Reviews of Modern Physics* **70**, pp. 653-681 (1998).

[4] GOERDT, A., in *Proc. 7th Int. Symp. on Mathematical Foundations of Computer Science*, 264 (1992); *Journal of Computer and System Sciences*, **53**, 469 (1996)
CHVÀTAL, V., and REED,B., in *Proc. 33rd IEEE Symp. on Foundations of Computer Science*, 620 (1992)

[5] COOK,S.A., *The complexity of theorem–proving procedures* , in *Proc. 3rd Ann. ACM Symp. on Theory of Computing*, Assoc. Comput. Mach., New York, 151 (1971)

[6] GAREY, M., and JOHNSON, D.S., *Computers and Intractability; A guide to the theory of NP–completeness*, W.H. Freeman and Co., San Francisco, 1979; C. Papadimitriou, *Computational Complexity*, Addison–Wesley, 1994;

[7] ASPVALL, B, PLASS, M.F. and TARJAN,R.E. *Inf. Process. Lett.***8**, 121 (1979)

[8] KIRKPATRICK, S., and SELMAN, B., *Science* **264**, 1297 (1994)

[9] KIRKPATRICK, S., GYORGYI, G., TISHBY, N., and TROYANSKY, L., NIPS Proceedings Vol 6,

439-446 (1993).

[10] FRIEDGUT, E., preprint, presented at DIMACS 97.

[11] Above α_c proving UNSAT is still hard, with a lower bound on search cost that is exponential in N with a coefficient that decreases as a power law in α. See "On the Complexity of Unsatisfiability Proofs for Random k-CNF Formulas", P. BAUME, R. KARP, T. PITASSI and M. SAKS, *Proc. STOC-98*, 1998 561-571.

[12] MITCHELL, D., SELMAN, B., and LEVESQUE,H., "Hard and Easy Distributions of SAT problems," *Proc. of Am. Assoc. for Artif. Intell. AAAI-92* (1992), 456-465.

[13] Issue 1–2, *Artificial Intelligence* **81**, HOGG, T., HUBERMAN, B.A., and WILLIAMS, C., Eds., (1996)

[14] SELMAN, B., and KIRKPATRICK, S., *Artificial Intelligence* **81**, 273 (1996)

[15] KIRKPATRICK, S., GELATT Jr., C.D., and VECCHI, M.P., *Science* **220**, 339 (1983)

[16] HERTZ, J., KROGH, A., and PALMER, R.G., *Introduction to the theory of neural computation*, Addison-Wesley, Redwood City (CA), 1991

[17] MÉZARD, M., PARISI, G., VIRASORO M.A., *Spin Glass Theory and Beyond*, World Scientific, Singapore, 1987

[18] MONASSON, R., and ZECCHINA R., *Phys. Rev. Lett.* **76**, 3881 (1996)

[19] MONASSON, R., and ZECCHINA R., *Phys. Rev.* **E 56**, 1357 (1997)

[20] MONASSON, R., and ZECCHINA R., *J. Phys. A: Math. Gen.* **31**, 9209 (1998)

[21] MONASSON, R., and ZECCHINA R., KIRKPATRICK, S., SELMANN, B., TROYANSKY, L., "Phase Transition and Search Cost in the $2+p$ SAT Problem", proceedings of *PhysComp96*, TOFFOLI, T., BIAFORE, M., LEÃO, J., eds., Boston (1996)

[22] ACHLIOPTAS, D., KIROUSIS, L., KRANAKIS, E., and KRIZANC, D., "Rigorous Results for Random $(2+p)$–SAT", *submitted to Phys. Rev. E.* A preliminary version appeared in RALCOM 97.

[23] Broder, A.Z., Frieze, A.M., and Upfal, E., *Proc. 4th Annual ACM–SIAM Symp. on Discrete Algorithms*, 322 (1993)

[24] Monasson, R., *Phil. Mag.* **B 77**, 1515 (1998); *J. Phys.* **A 31**, 513 (1998)

[25] Talagrand, M., Huge Random Structures and mean field models for spin glasses, Doc. Math. J. DMV Extra Volume ICM I, 507 (1998)

[26] Talagrand, M. *Probability and Related Fields* **110**, 109-176 (1998).

[27] Talagrand, M. "The high temperature case of the K-sat problem". To appear in *Probability Theory and Related Fields*

[28] Gardner, E., Derrida, B. *J. Phys. A* **22**, 1983 (1989)

[29] Cover, T.M. *IEEE Trans. Electron. Comput.* **EC-14**, 326 (1965)

[30] Derrida, B. *Phys. Rev. B* **24**, 2613 (1981)

[31] Gross, D., Mézard, M. *Nuclear Phys. B* **240**, 431 (1984)

[32] Davis, M., and Putnam, H., "A computing procedure for quantification theory," *J. Assoc. Comput. Mach.*, **7** (1960), 201–215.

[33] Crawford, J.M. and Auton, L.D. (1993). "Experimental results on the cross-over point in satisfiability problems." *Proc. AAAI-93*, Washington, DC, 21–27.

[34] Van Gelder, Allen (1999). "Autarky pruning in propositional model elimination reduces failure redundancy" Journal of Automated Reasoning (to appear).

[35] Frieze, A., and Suen, S. "Analysis of two simple heuristics on a random instance of K–SAT", *Journal of Algorithms 20* (1996), 312-335.

[36] Dubois, O., and Boufkhad, Y., "A general upper bound for the satisfiability threshold of random K–SAT formulas," *Journal of Algorithms 24*, 395-420 (1997)

[37] Kirousis, L., Kranakis, E., and Krizanc, D., "Approximating the unsatisfiability threshold of random formulas," *Proceedings of the 4th European Symposium on Algorithms*, (1992), 27-38.

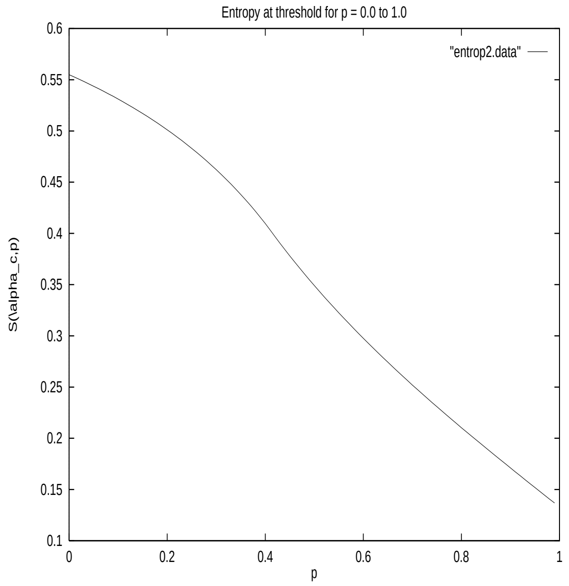
[38] Kamath, A., Motwani, R., Palem, K., and Spirakis, P., "Tail bounds for occupancy and the satisfiability threshold conjecture," *Random Structures and Algorithms* **7**, 59 (1995)

[39] Bollobás, B., "Random Graphs", (Academic Press, London, 1985).

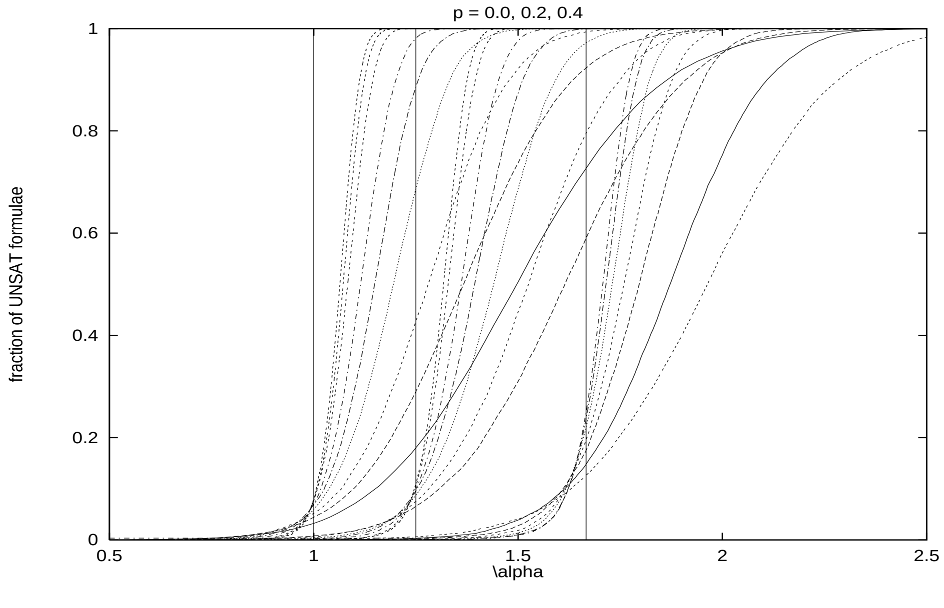
[40] Anglès d'Auriac, J.-C. and Sourlas, N., "The $3d$ random field Ising model at zero temperature," *Europhysics Lett. 39*, (1997) 473-478.

[41] Kirkpatrick, S. and Toulouse, G., "Configuration space analysis of travelling salesman problems," *Journale de Physique* **46**, (1985) 1277-1292.

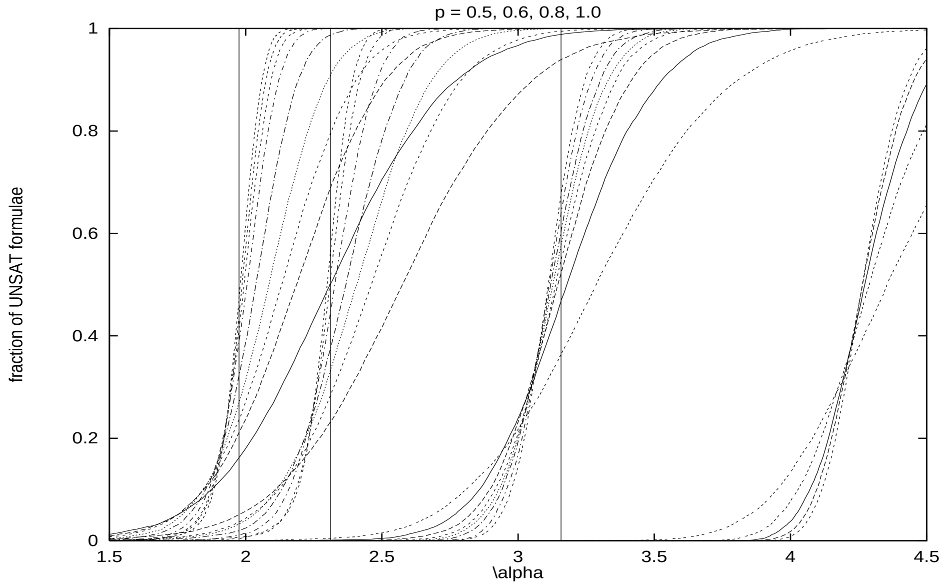
[42] Schneider, J., Froschhammer, C., Morgenstern, I., Husslein, T., Singer, J. M., "Searching for Backbones – an efficient parallel algorithm for the travelling salesman problem", *Computer Physics Communications 96* (1996) 173-188



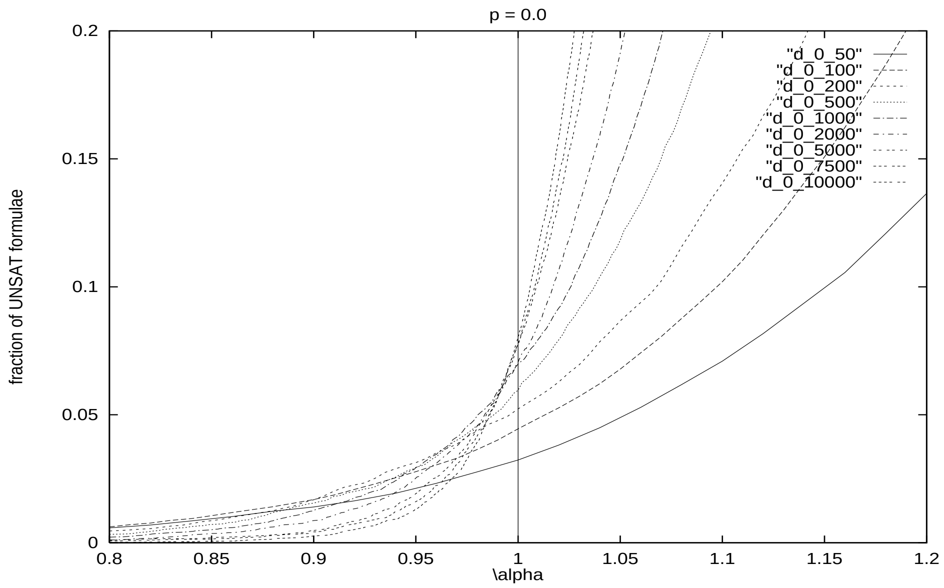
1. Ground state entropy at $\alpha_c(p)$ versus p, predicted by the RS theory of [19,20]. For $p < p_0$, $\alpha_c(p) = 1/(1-p)$. For $p > p_0$, we have used the estimates of $\alpha_c(p)$ obtained by finite-size scaling.



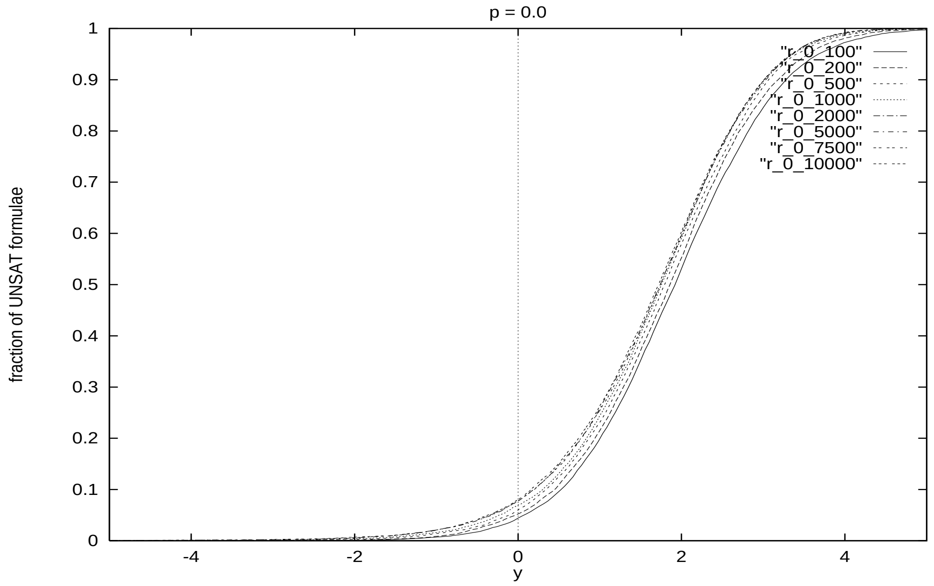
2a. Raw data used in this study for $p = 0.0$, 0.2, and 0.4. Vertical lines mark the thresholds for each value of p. For $p = 0.0 (K = 2)$, data are plotted for $N = 50$, 100, 200, 500, 1000, 2000, 5000, 7500, and 10000. For $p = 0.2$, values of N are 100, 200, 500, 1000, 2000, 5000, and 7500. For $p = 0.4$, values of N are 100, 200, 500, 1000, 2000, 3500, and 5000.



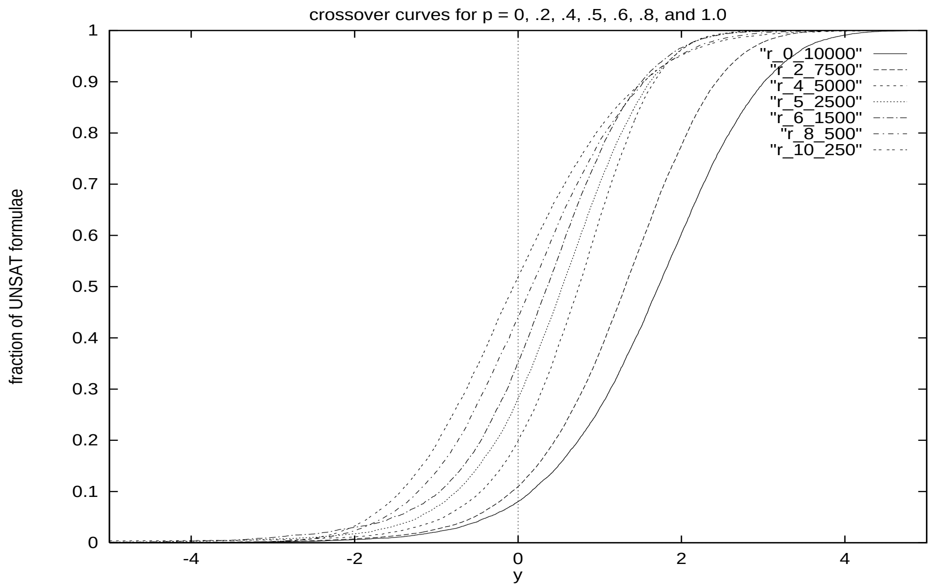
2b. Raw data for $p = 0.5$, 0.6, 0.8 and 1.0. Thresholds marked are determined from the RS theory (an overestimate). For $p = 0.5$, values of N are 50, 100, 150, 250, 500, 1000, 1500, 2000, and 2500. For $p = 0.6$, values of N are 50, 100, 150, 250, 500, 1000, and 1500. For $p = 0.8$, values of N are 50, 100, 150, 200, 250, 300, 400, and 500. For $p = 1.0 (K = 3)$, values of N are 50, 100, 150, 200, and 250.



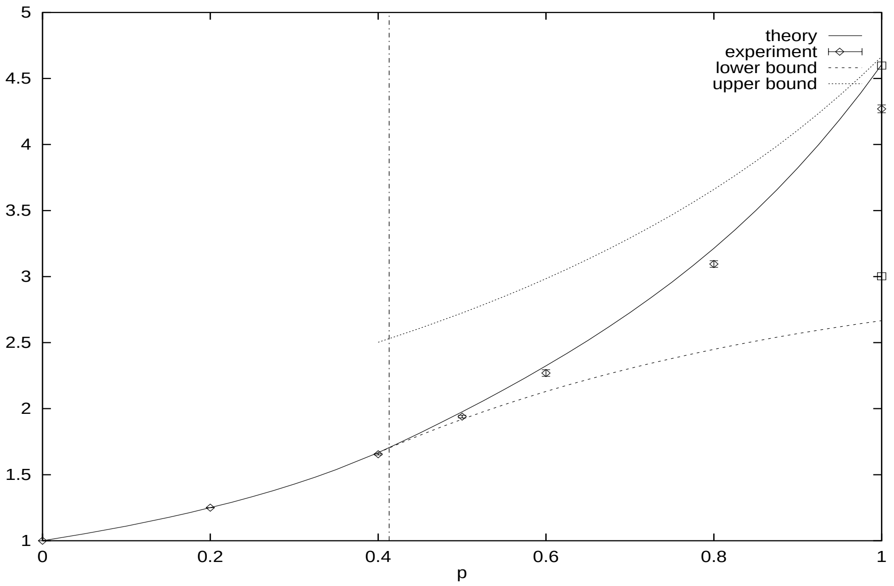
3. Plot of data obtained for $K = 2 (p = 0.0)$, shown on a finer scale to exhibit size dependence.



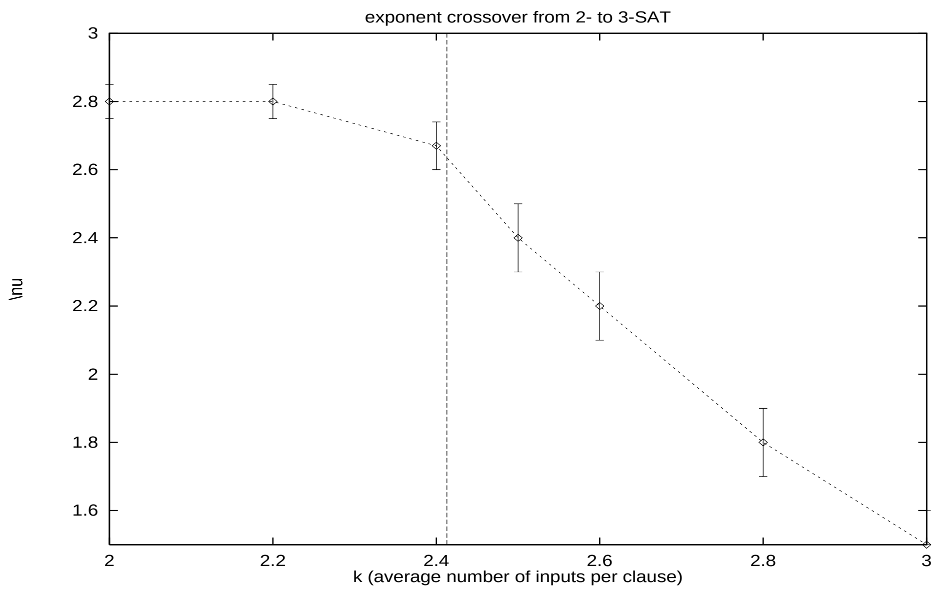
4. Rescaled data for $K = 2(p = 0.0)$ using $\nu = 2.8$. The curves for different values of N are parallel, converge to a limit for large N, and are monotonic in N at each value of y.



5. Rescaled data for all p, using the largest values of N which could be obtained, with $alpha_c$ and ν determined as described in text.



6. Theoretical and experimental results for the SAT/UNSAT transition in the 2+p-SAT model. The vertical line at p_0 separates the continuous from the discontinuous transition. The full line is the replica-symmetric theory's predicted transition, believed exact for $p < p_0$, and the diamond data points with error bars are results of computer experiment and finite-size scaling. The other two lines show upper and lower bounds obtained in [22], while the stronger upper bound due to [37], and the best known lower bound, due to [35], are indicated by square data points.



7. Crossover seen in the exponent ν governing the width of the critical regime, as K increases from 2 to 3.

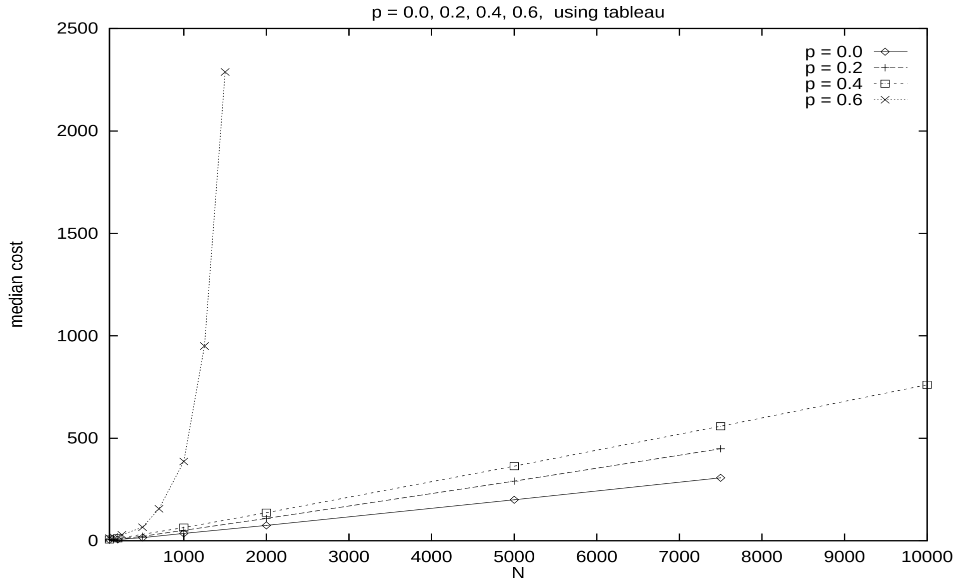


Fig. 8a: Median computational cost, linear scale.

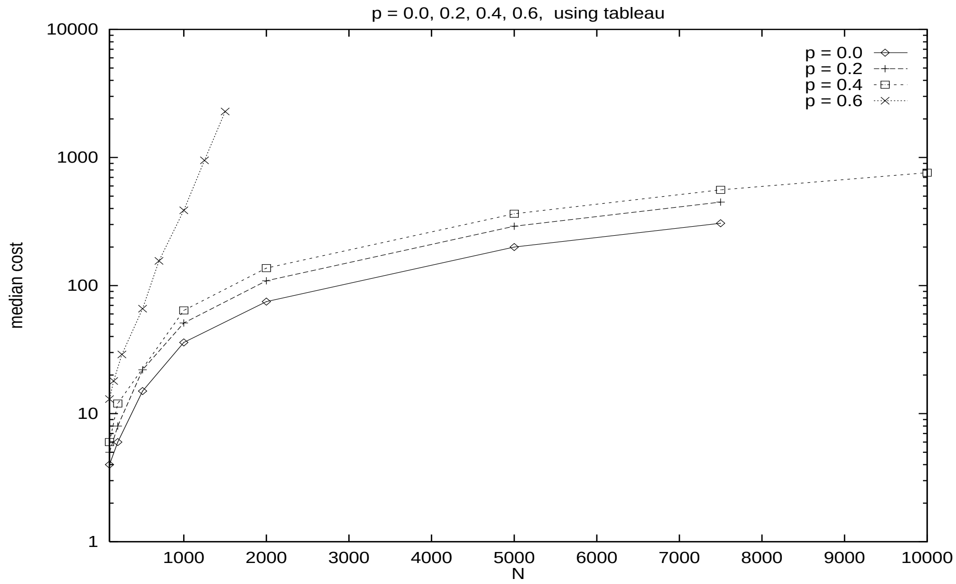


Fig. 8b: Median cost, semilog scale.

8. Median computational cost of proving a formula SAT or UNSAT using the TABLEAU search method, for p ranging from 0 to 1. The data in (a) is plotted on a linear scale, appropriate for the cases with $p < p_0$. The semi-log plots in (b) show an exponential dependence of cost on N for $p > p_0$.

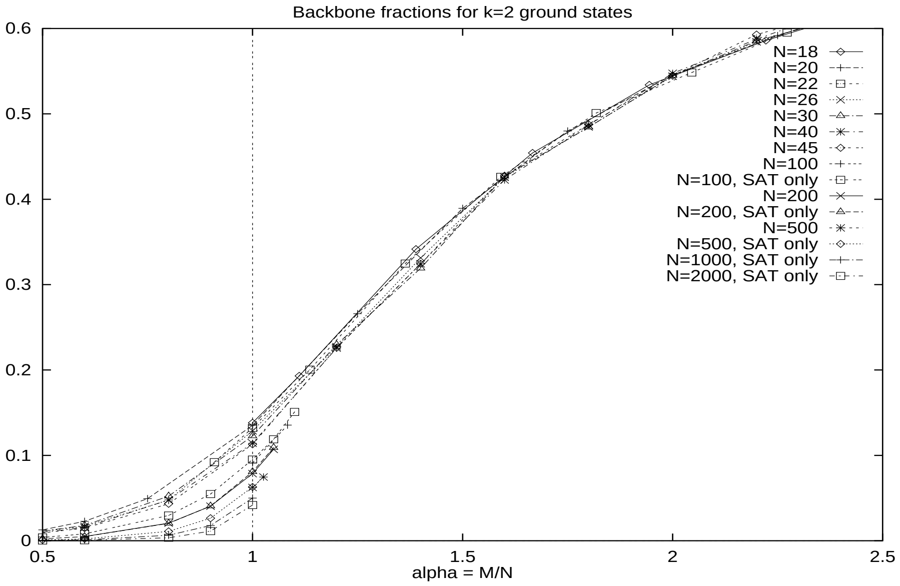


fig. 9a: Backbone fraction for $K = 2$.

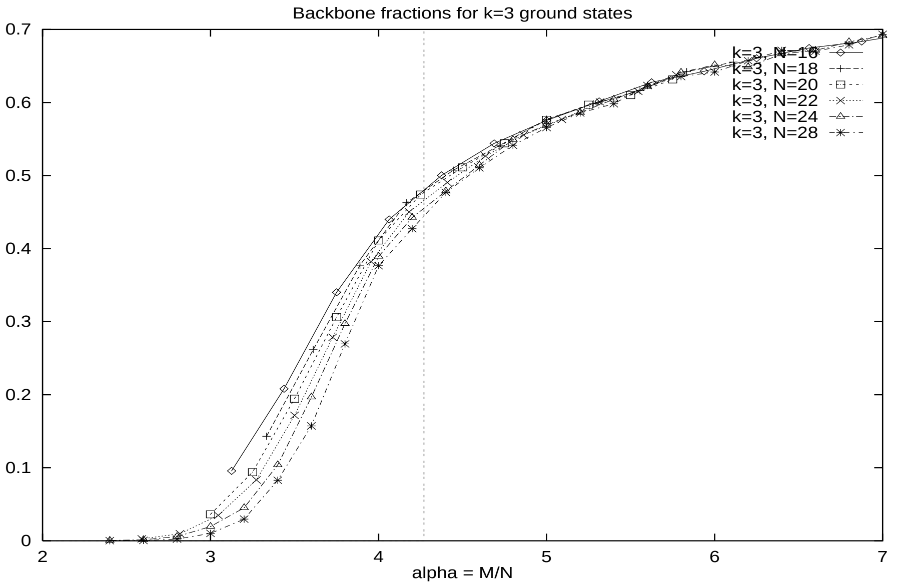


fig. 9b: Backbone fraction for $K = 3$.

9. Backbone fractions as a function of α averaged over many samples for $K = 2$ (a) cases with $N = 18$ to 45 and $K = 3$ (b) cases with $N = 18$ to 28. The vertical lines mark the SAT/UNSAT thresholds in the limit $N \to \infty$. For 2-SAT, data obtained from larger sizes $N = 100, 200, 500$ show that the backbone fraction at the threshold decreases to zero.