# Statistical Physics of the Random $K$-Satisfiability Problem

*Remi Monasson and Riccardo Zecchina*

Laboratoire de Physique Théorique de l'ENS, Paris, France
International Center for Theoretical Physics, Trieste, Italy

April 6, 1998

[summary by Olivier Dubois and Remi Monasson]

The satisfaction of constrained formulae is a key issue in complexity theory. Many computational problems are shown to be NP-complete through a polynomial mapping onto the $K$-Satisfiability (SAT) problem. Recently, 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 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 $C_\ell$, $\ell = 1, \ldots, M$. The logical AND of all clauses, $\mathcal{F}$, is said to be satisfiable if there exists a logical assignment to the $x$'s evaluating $\mathcal{F}$ to true, unsatisfiable otherwise.

Numerical experiments have concentrated upon the study of the probability $P_N(\alpha, K)$ that a given $\mathcal{F}$ including $M = \alpha N$ clauses be satisfiable. For large sizes of $N$, there appears a remarkable behaviour: $P$ seems to reach unity for $\alpha < \alpha_c(K)$ and vanishes for $\alpha > \alpha_c(K)$ [6]. Such an abrupt threshold behaviour, separating a SAT phase from an UNSAT one, has indeed been rigourously confirmed for 2-SAT, which is in P, with $\alpha_c(2) = 1$ [2, 5]. For larger $K \geq 3$, $K$-SAT is in NP and much less is known. The existence of a sharp transition has not been proven yet but precise estimates of the thresholds have been found: $\alpha_c(3) \simeq 4.25$. Moreover, some lower and upper bounds to $\alpha_c(3)$ (if it exists), $\alpha_{l.b.} = 3.003$ and $\alpha_{u.b.} = 4.64$ respectively have been established [4, 3].

The classical approaches to study the SAT phenomenon threshold are both combinatorial and probabilistic. A statistical physics approach was used in [8, 9]. Such an approach allows properties to be predicted. It has been applied already to random graphs and it has led to large deviation results for the threshold phenomenon of random graphs in addition to previously known results. This approach seems therefore to be powerful. However it proves much harder to apply to the SAT threshold phenomenon. It yields in particular a surprising change concerning the proportion of variables fixed in the neighbourhood of the threshold between 2-SAT and 3-SAT. This could partly account for the complexity gap between these two problems. In order to apply the statistical physics approach, the following steps were carried out.

First, the energy function corresponding to the $K$-SAT problem is identified. The logical values of the $x$'s can be represented by $N$ binary variables $S_i$'s, 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 $C_\ell$ includes $\overline{x_i}$ (resp. $x_i$), $C_{\ell i} = 0$ otherwise. 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 $\mathbf{S}$. The minimum $E[\mathbf{C}]$ of $E[\mathbf{C}, \mathbf{S}]$, that is, the lowest number of violated clauses that can be achieved by the best possible logical assignment [8, 9], is a random variable which becomes totally concentrated around its mean value $\ll E[\mathbf{C}] \gg$ in the large size limit [1]. The latter is accessible through the

knowledge of the averaged logarithm of the generating function

$$(1) \qquad Z[\mathbf{C}] = \sum_{\mathbf{S}} \exp\left(-E[\mathbf{C}, \mathbf{S}]/T\right)$$

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. Being the minimal number of violated clauses, $\ll E[\mathbf{C}] \gg$ equals zero in the SAT region and is positive in the UNSAT phase, allowing the location of $\alpha_c(K)$.

The calculation of the average value of the logarithm of $Z$ in (1) 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 $n$ in order 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 [8]

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

It is crucial to notice that the averaged term in (2) depends on the $n \times N$ spin replicas only through the $2^n$ occupation fractions $c(\sigma)$ labelled by the vectors $\sigma$ with $n$ binary components; $c(\sigma)$ equals the number (divided by $N$) of labels $i$ such that $S_i^a = \sigma^a$, $\forall a = 1, \ldots, n$. Taking into account the combinatorial entropy of the labels $i$ at fixed occupation fractions,

$$\ll Z[\mathbf{C}]^n \gg \simeq \exp(N F_{max})$$

where $F_{max}$ is the maximum over all possible $x$s of the functional [8]

$$(3) \qquad F[\{c\}] = -\sum_{\sigma} c(\sigma) \log c(\sigma) + \alpha \log\left[\sum_{\sigma_1, \ldots, \sigma_K} c(\sigma_1) \cdots c(\sigma_K) \exp\left(-\frac{1}{T} \sum_{a=1}^{n} \sum_{\ell=1}^{K} \delta[\sigma_\ell^a + 1]\right)\right].$$

The optimisation conditions over $F[\{c\}]$ provide $2^n$ coupled equations for the $c$s. Notice that $F$ is a symmetric functional, that is, 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 $-1 \leq m \leq 1$ [8]. Once the number of replicas $n$ is sent to zero, we obtain a self-consistent functional equation for the order parameter $P(m)$ that can be solved numerically.

What is the meaning of the distribution $P(m)$? Consider a formula $\mathcal{F}$ and all the spin configurations $\mathbf{S}^j$, $j = 1, \ldots, \mathcal{N}$ attaining the minimum $E[\mathbf{C}]$ of the cost-function $E[\mathbf{C}, \mathbf{S}]$. Define then the average Boolean magnetisations of the spins

$$(4) \qquad m_i = \frac{1}{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} S_i^j,$$

over the set of optimal configurations. Call $H(\mathbf{C}, m)$ the histogram of the $m_i$s and $H(m)$ the average of $H(\mathbf{C}, m)$ over the choices of the formulae $\mathcal{F}$. $H(m)$ is a probability distribution over the interval $-1 \leq m \leq 1$ giving information about the resulting constraints on the variables induced by the clauses. It has been shown that, if the RS solution is the global maximum of (3) (and not only

a local one), $H(m)$ 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.

Of particular interest are the fully fixed variables, that is the $x_i$'s such that $m_i = \pm 1$. In the following, the fraction of fully constrained variables will be denoted by $\gamma(\alpha, K)$. Clearly, $\gamma(\alpha, K)$ vanishes in the SAT region otherwise the addition of a new clause to $\mathcal{F}$ would lead to a contradiction with a finite probability. Two kinds of scenarii arise 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 $\gamma_c \simeq 0.9$ slightly above the threshold. While $\alpha_c(2) = 1$ is correctly found, the RS prediction for $\alpha_c(3) = 4.6$ exceeds the experimental estimates by 10%. Work is currently under progress to refine the above calculation and enlarge the subspace where the global maximum is sought in.

Qualitatively speaking, however, we expect the main conclusion of this work to be correct: the SAT/UNSAT transition is accompanied by a smooth (respectively abrupt) change in the structure of the solutions of the 2-SAT (resp. 3-SAT) problem. Furthermore, we conjecture that this discrepancy is responsible for the difference of typical complexities of both models recently observed in numerical studies [10]. The typical solving time of search algorithms displays an easy-hard-easy pattern as a function of $\alpha$ with a peak of complexity close to the threshold. The peak time seems to scale polynomially with $N$ for the 2-SAT problem and exponentially with $N$ in the 3-SAT case. From an intuitive point of view, the search for solutions ought to be more time-consuming in the presence of a finite fraction of fixed variables since the exact determination of the latter necessarily requires an exhaustive enumeration of the variables. To test this conjecture, a mixed $2 + p$-model has been introduced; it includes a fraction $p$ (resp. $1 - p$) of clauses of length two (resp. three) and thus interpolates between the 2-SAT ($p = 0$) and 3-SAT ($p = 1$) problems. The RS theory predicts that the SAT/UNSAT transition becomes abrupt when $p > p_0 = 0.41$. Precise numerical simulations support the conjecture that the polynomial/exponential crossover occurs at the same critical $p_0$. An additional argument in favour of this conclusion is provided by the analysis of the finite-size effects on $P_N(\alpha, K)$ and the emergence of some universality for $p < p_0$. A detailed account of these findings may be found in [7].

## Bibliography

[1] Broder (Andrei Z.), Frieze (Alan M.), and Upfal (Eli). – On the satisfiability and maximum satisfiability of random 3-CNF formulas. In *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (Austin, TX, 1993)*. pp. 322–330. – New York, 1993.

[2] Chvátal (V.) and Reed (B.). – Miks gets some (the odds are on his side). In *Proceedings of the 33rd IEEE Symposium on Foundations of Computer Science*, pp. 620–627. – 1992.

[3] Dubois (O.) and Boufkhad (Y.). – A general upper bound for the satisfiability threshold of random $r$-SAT formulae. *Journal of Algorithms*, vol. 24, 1997, pp. 395–420.

[4] Frieze (Alan) and Suen (Stephen). – Analysis of two simple heuristics on a random instance of $k$-SAT. *Journal of Algorithms*, vol. 20, n° 2, 1996, pp. 312–355.

[5] Goerdt (Andreas). – A threshold for unsatisfiability. *Journal of Computer and System Sciences*, vol. 53, n° 3, 1996, pp. 469–486. – 1994 ACM Symposium on Parallel Algorithms and Architectures (Cape May, NJ, 1994).

[6] Mitchell (D.), Selman (B.), and Levesque (H. J.). – Hard and easy distributions of SAT problems. In *Proceedings of the American Association for Artificial Intelligence AAAI-92*, pp. 459–465. – 1992.

[7] Monasson (R.), Zecchina (R.), Kirkpatrick (S.), Selman (B.), and Troyansky (L.). – Typical-case complexity results from a new type of phase transition. – In preparation.

[8] Monasson (Rémi) and Zecchina (Riccardo). – Entropy of the $K$-satisfiability problem. *Physical Review Letters*, vol. 76, n° 21, 1996, pp. 3881–3885.

[9] Monasson (Rémi) and Zecchina (Riccardo). – Statistical mechanics of the random $K$-satisfiability model. *Physical Review E. Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics. Third Series*, vol. 56, n° 2, 1997, pp. 1357–1370.

[10] Selman (Bart) and Kirkpatrick (Scott). – Critical behavior in the computational cost of satisfiability testing. *Artificial Intelligence*, vol. 81, n° 1-2, 1996, pp. 273–295. – Frontiers in problem solving: phase transitions and complexity.