# The SAT Phase Transition

**Ian P. Gent**[1] **and Toby Walsh**[2]

**Abstract.** We describe a detailed experimental investigation of the phase transition for several different classes of randomly generated satisfiability problems. We observe a remarkable consistency of features in the phase transition despite the presence in some of the problem classes of clauses of mixed lengths. For instance, each of the problem classes considered has a sharp transition from satisfiable to unsatisfiable problems at a critical value. In addition, there is a common easy-hard-easy pattern in the median difficulty of the problems, with the hardest problems being associated with the phase transition. However, the difficulty of problems of mixed clause lengths is much more variable than that of fixed clause length. Indeed, whilst the median difficulty of random problems of mixed clause lengths can be orders of magnitude easier than that of equivalently sized problems of fixed clause length, the hardest problems of mixed clause lengths can be orders of magnitude harder than the hardest equivalently sized problems of fixed clause length. Such very hard random problems may be of considerable practical and theoretical use in analysing algorithm performance.

## 1 Introduction

Propositional satisfiability (SAT) is the problem of deciding if there is an assignment for the variables in a propositional formula that makes the formula true. SAT is of considerable practical interest as many AI tasks can be encoded quite naturally in SAT. Unfortunately, unless P=NP, SAT is intractable in the worst case as it is a NP-hard problem. There are, however, many theoretical and experimental results which show good average-case performance for certain classes of SAT problems [7]. In considering such average-case results, it is important to know whether the problems considered are hard and representative of those met in practice. Cheeseman et al [1] observed that the hard instances of NP-hard problems are often associated with a phase transition. With SAT, there is a phase transition as the ratio of the number of clauses to variables in a problem is varied. Experiments have shown that an easy-hard-easy pattern for SAT occurs as this ratio is increased and that the hard instances occur in the phase transition [15]. The phase transition for SAT is therefore of considerable practical and theoretical importance.

In this paper, we present a detailed experimental investigation of the SAT phase transition. We consider several different classes of SAT problems, some of which (like real problems) contain clauses of mixed lengths. We observe a remarkable consistency of features in the different phase transitions. For example, all the problem classes show an easy-hard-easy pattern in median problem difficulty, a region of highly variable difficulty, and a sharp transition from satisfiable to unsatisfiable at a fixed ratio of clauses to variables. Random problems of mixed clause lengths appear, however, to give much more variable behaviour. The median difficulty of random problems of mixed clause is typically orders of magnitude less than that of equivalently sized problems of fixed clause length, yet the hardest problems of mixed clause lengths can be orders of magnitude *harder* than the hardest problems of fixed clause length. With random problems of mixed clause lengths, certain key properties like the position of the phase transition also appear to be governed merely by a simple parameter, the (limiting) distribution of clause lengths in the problem class. The phase transition observed for random 3-SAT [15] thus appears to be a special case of a more general type of SAT phase transition.

## 2 Random $k$-SAT

We consider SAT problems in conjunctive normal form (CNF); a formula is in CNF iff it is a conjunction of clauses, where a clause is a disjunction of literals, and a literal is a negated or un-negated variable. A problem in random $k$-SAT consists of $L$ clauses, each of which has $k$ literals chosen uniformly from the $N$ possible variables and the $N$ possible negated variables. We use $R_k(N, L)$ to denote problems drawn from this class and $Prob(sat, X)$ to denote the probability that a problem drawn at random from the class $X$ is satisfiable.

Most recent experimental work has used the random $k$-SAT model as it has several features which makes it useful for benchmarks. In particular, there appears to be a phase transition between satisfiability and unsatisfiability as $L/N$ is varied. That is, there exists $c_k$, a critical value of $L/N$, such that:

$$\lim_{N \to \infty} Prob(sat, R_k(N, c.N)) = \begin{cases} 0 & \text{for } c > c_k \\ 1 & \text{for } c < c_k. \end{cases}$$

It is easy to show that $c_0 = c_1 = 0$. It has been shown theoretically that $c_2 = 1$ [2, 9] and $3.003 < c_3 < 4.81$. Experiments have suggested that $c_3 \approx 4.24$ [4]. The phase transition appears to be of practical value as, for a given $N$, problems with $c_k.N$ clauses seem to be the hardest problems generated for a wide variety of SAT algorithms [15]. This result needs to be treated with slight caution since it has not been shown theoretically, and since problems away from the phase transition can also be hard to solve. For example, all resolution algorithms need exponential time with probability tending to $1$ for random 3-SAT problems generated with $c > 5.6$ [3].

To determine satisfiability and problem difficulty, we use two variants of the Davis-Putnam procedure [5]. The major difference between these variants is their choice of variable upon which to branch. For small problems, we use a simple variant used in previous studies [15] which branches on the first variable in the first clause. We shall refer to this variant as "DP". For larger problems, we use ASAT [6] which branches on the variable having the greatest number of occurrences in the shortest clauses. ASAT is one of the fastest implementations of the Davis-Putnam procedure currently distributed.

[1] Department of Artificial Intelligence, University of Edinburgh 80 South Bridge, Edinburgh EH1 1HN, United Kingdom. I.P.Gent@ed.ac.uk
[2] INRIA-Lorraine, 615, rue du Jardin Botanique, 54602 Villers-les-Nancy, France. walsh@loria.fr


ECAI 94. *11th European Conference on Artificial Intelligence* Edited by A. Cohn
Published in 1994 by John Wiley & Sons, Ltd.

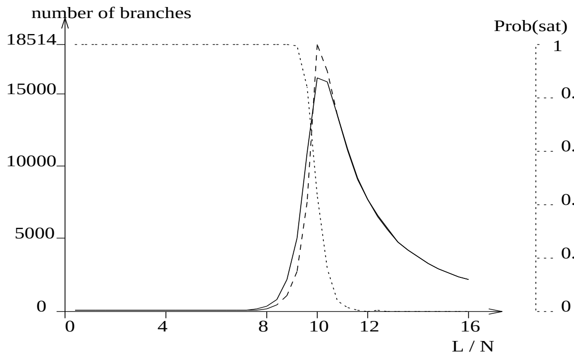



*Figure 1. Random* 4-SAT *problems, tested using ASAT, mean (solid), median (dashed) branches,* $N = 75$

Figure 1 shows a typical random $k$-SAT phase transition, like that for 3-SAT in [15], the dotted line giving probability of satisfiability at each point. For $N = 75$, we tested $1000$ random 4-SAT problems at each point from $L/N = 0$ to $16$ in steps of $0.4$. The graph of observed probability of satisfiability is similar with varying $N$, except that the transition from near $100\%$ to near $0\%$ becomes sharper with increasing $N$. There appears to be a "crossover" point at which approximately the same percentage of problems is satisfiable for all values of $N$, as previously observed in 3-SAT [13]. To examine this for 4-SAT we tested problems from $L/N = 9$ to $12$ in steps of $0.04$, for $N = 25, 50, 75$. The most consistent point appeared to be at $9.76N$ clauses at a probability of $65\%$ satisfiable.

Figure 1 shows a typical "easy-hard-easy" pattern in problem difficulty for 4-SAT using the ASAT procedure. The $y$-axis gives the mean and median (to the same scale) number of branches reported by ASAT. When $L/N$ is large, problems are usually over-constrained, and thus easily shown to be unsatisfiable. When $L/N$ is small, problems are usually under-constrained, and a satisfying assignment can be "guessed" quickly. The really hard instances tend to occur at the phase transition where the problems are finely balanced between being satisfiable and unsatisfiable. Note that mean and median behaviour are very similar.

## 3 Random mixed SAT

We now introduce a generalisation of the random $k$-SAT model, which we call "random mixed SAT". In this model, a set of clauses is generated with respect to a probability distribution $\phi$ on the integers. Each clause is generated as in random $k$-SAT. However, $k$, the length of the clause, is chosen randomly according to $\phi$. For example, if $\phi(2) = \phi(3) = \frac{1}{2}$, then clauses of length 2 and 3 appear with probability $\frac{1}{2}$, whilst if $\phi(2) = \frac{1}{3}$ and $\phi(4) = \frac{2}{3}$, clauses of length 2 appear with probability $\frac{1}{3}$ and of length 4 with probability $\frac{2}{3}$. In this paper, we will call these problem classes "2-3-SAT" and "2-4-4-SAT" respectively. The frequency of occurrence of an integer in the name reflects the frequency of occurrence of clauses of this length in the problem. Thus, for "3-4-SAT", $\phi(3) = \phi(4) = \frac{1}{2}$. Random $k$-SAT is a special case of random mixed SAT, where each clause is chosen of length $k$ with probability $1$.

The random mixed SAT model may generate problems more similar to real-world problems than random $k$-SAT. For example, many structured problem classes use clauses of mixed lengths (*eg.* scheduling problems have large numbers of binary clauses). It would be interesting to compare such problems with random mixed SAT problems with similar proportions of clauses lengths.

We write $c_\phi$ for the critical value of $L/N$ for a given random mixed SAT (if such a value exists). If $\phi(0) > 0$ or $\phi(1) > 0$, then we know that $c_\phi = 0$, as empty and unit clauses will occur, and as $c_0 = c_1 = 0$. Henceforth we assume that $\phi(0) = \phi(1) = 0$. Although we do not know the value of $c_\phi$ in other cases, it is at least simple to give an upper bound on its value if it exists. For a given $\phi$, we define the *density* of $\phi$, $d_\phi$ as:

$$d_\phi =_{def} \sum_{k=1}^{\infty} \phi(k)(1 - (\frac{1}{2})^k)$$

The density gives the mean fraction of all truth assignments that are consistent with any given clause generated by $\phi$. The expected number of models of a clause set with $N$ variables and $L$ clauses is $2^N(d_\phi)^L$. A satisfiable formula has at least one model, so for fixed $L/N$, probability of satisfiability vanishes as $N \to \infty$ if $2(d_\phi)^{L/N} < 1$. Rearranging, we see that

$$c_\phi \leq \frac{-1}{\log_2(d_\phi)}. \qquad (1)$$

For $k = 2$, the bound gives $c_2 \leq 2.41$, but it is known that $c_2 = 1$. For $k = 3$, (1) gives $c_3 \leq 5.20$ while it is believed $c_3 \approx 4.24$, and our experiments suggest $c_4 \approx 9.76$ while (1) gives $c_4 \leq 10.75$.

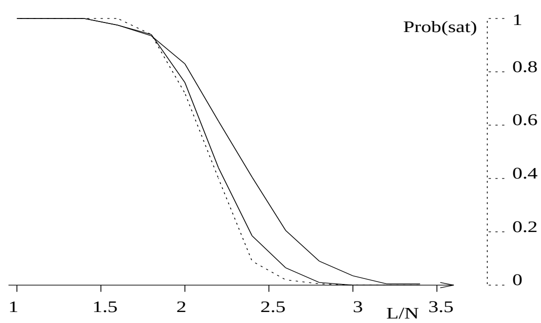


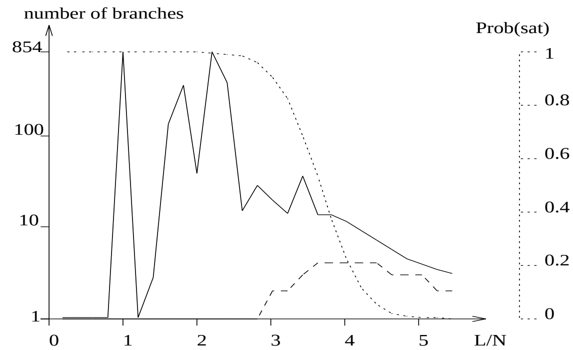

*Figure 2. Random* 2-3-SAT, $N = 50, 100, 150$

For random 2-3-SAT, we get $d_\phi = \frac{13}{16}$, giving a bound $c_\phi < 3.34$. In our exploration of the phase transition, this gives a useful upper bound on how far experiments need be performed. Figure 2 shows how the probability of satisfiability varies with the ratio of clauses to variables. For $N = 50, 100, 150$, experiments were performed with $L$ varying in steps of $0.2N$ from $0$ to $3.4N$. The higher values of $N$ are characterised by a sharper transition, the dotted line being $N = 150$. It can be seen that, as for random $k$-SAT, the probability graphs seem to be converging on some limit. There is again a crossover, but at a high percentage satisfiability. To investigate this further, we performed a more detailed investigation of the crossover region. For values of $N$ from $50$ to $150$ in steps of $10$, and from $150$ to $350$ in steps of $50$, we tested $500$ problems at values of $L$ from $1.3N$ to $2N$ in steps of at most $N/50$. Analysis of the data suggests that the crossover appears close to $94\%$ satisfiable and $L = 1.75N$, where we take the crossover to be where we observe the smallest difference in the probability of satisfiability with changing problem size. By contrast, the ratio $L/N$ where $50\%$ of problems were satisfiable declined from $2.32$ for $N = 50$ to $2.08$ for $N = 200$.

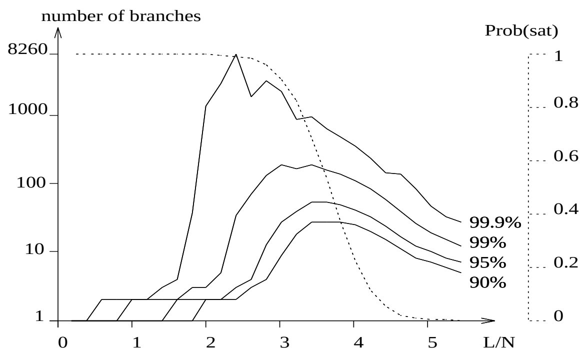
For random 2-4-4-SAT, we get $d_\phi = \frac{7}{8}$, the same as 3-SAT, giving $c_\phi < 5.20$. Once again, a similar pattern is observed when graphs of probability are plotted. However, the crossover appears at quite a different point compared to 3-SAT. For values of $N$ from $50$ to $130$ in steps of $10$, we tested $500$ problems at values of $L$ from $2.3N$ to $3.3N$ in steps of at most $N/25$. The crossover point appears to be about $L = 2.74N$ where $96\%$ of problems are satisfiable. By contrast the $50\%$ satisfiable point ranged from $3.78N$ at $N = 50$ to $3.48N$ at $N = 150$. This suggests an estimate for $c_\phi$ of $2.74$.

## 4 Problem Hardness

The identification of phase transitions is of considerable importance in the study of heuristics for NP-hard problems since the hardest instances of randomly generated problems tend to occur in the phase transition. As observed in [15] for random 3-SAT, the hardest random $k$-SAT problems appear to occur in the transition between satisfiability and unsatisfiability. This is seen very clearly for random 4-SAT in Figure 1.




*(a) mean (solid) and median (dashed) branches*




*(b) varying percentile branches*

*Figure 3. Random* 2-4-4-SAT *problems, DP,* $N = 75$

Random mixed 2-4-4-SAT gives very different behaviour to 3-SAT or 4-SAT. Figure 3(a) shows the median and mean problem difficulty for 2-4-4-SAT, using DP without branching heuristics. There is very little relation between mean and median problem difficulty. Median problem difficulty shows a slight easy-hard-easy pattern, though the peak median difficulty is only $4$ branches. Mean behaviour is exceptionally noisy, even though each data point represents the mean of $10,000$ experiments, and we have plotted the the mean on a *log* scale. This behaviour is clarified by Figure 3(b) which shows contours of the difficulty of problems representing percentiles from $90\%$ to $99.9\%$. Again a *log* plot has been used. Hard problems are still associated with the transition but the hardest problems no longer occur around the $50\%$ satisfiable point. The most difficult problems can be either satisfiable or unsatisfiable, and occur at high percentage satisfiability. Similar behaviour has been observed by Hogg and Williams for randomly generated 3-colourability problems [10].

The difference between hardness of 4-SAT and 2-4-4-SAT does not seem to be caused merely by the use of two different procedures in Figures 1 and 3. Although ASAT finds all 2-4-4-SAT problems at $N = 75$ easy, at $N = 300$ it needed more than $3$ million branches for one satisfiable problem in a region of high percentage satisfiability while needing less than $100$ branches for almost all other problems. Experiments with other random mixed SAT problems and other procedures also show that the hardest problems with mixed clause lengths can be orders of magnitude harder than the hardest problems with a fixed clause length, and that these hard problems tend to occur in regions of high percentage satisfiability.

## 5 Constant Probability Model

Another common problem class which gives clauses of mixed lengths is the constant probability model. This class has also been the subject of much theoretical attention. In the constant probability model, clauses with $N$ variables and $L$ clauses are generated according to a parameter $p$, $0 < p \leq 1$. For each clause, each literal (that is, a variable or the negation of a variable) is included with probability $p$, independently of the inclusion of other literals. In particular, the empty clause is allowed. Our experiments use a variant of the constant probability model proposed in [11] in which if a clause is generated containing either *no* literals or only *one* literal, it is discarded and another clause generated in its place. This is because the inclusion of empty or unit clauses typically makes problems easier. We shall call this the "CP" model.

This model cannot strictly be seen as an example of a random mixed SAT, because the probability of a given clause length being chosen varies with $N$. However, we observe very similar effects to those seen with random mixed SAT, provided that we omit empty and unit clauses, and provided that we fix the value $2Np$ and so vary $p$ as $1/N$. This keeps the expected clause length nearly constant. Indeed, for any given value of $2Np$, this gives a limiting distribution of clause lengths determined by the Poisson distribution with parameter $2Np$ (adjusted for the omission of clauses of length $0$ and $1$). For comparatively small values such as $2Np = 3$, we get quite fast convergence to the true distribution of clause lengths. For example, for $N = 25$, the true probability of length $3$ clauses is $0.285$, while the Poisson model gives a probability of $0.280$.

For the Poisson approximation to the constant probability model with parameters $N$ and $p$, we have $\phi(k) = e^{-2Np}(2Np)^k/k!$ by definition. We can derive an expression for the density $d_\phi$. If empty and unit clauses are allowed, then we can derive $1 - d_\phi = e^{-Np}$. Allowing for the omission of empty and unit clauses, if we choose $2Np = 3$, we get $d_\phi \approx 0.877$, a value close to the density of 3-SAT, $0.875$.

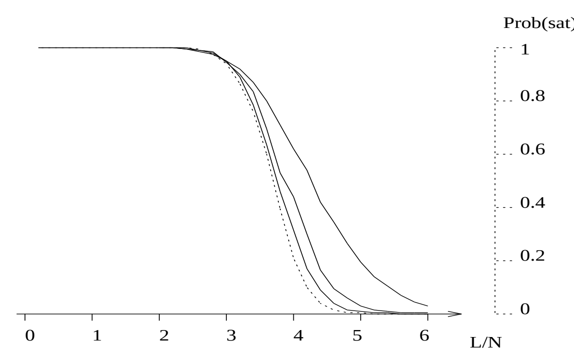



*Figure 4. CP,* $2Np = 3$, $N = 25, 50, 75, 100$

Our observations in §3 and §4 seem to apply to CP. In particular, if $2Np$ is kept constant as in Figure 4, graphs of probability of satisfiability show the same features, with a crossover at about $2.80$. Very interestingly, the distribution of problem difficulty in Figure 5 is similar to the very variable distribution seen in Figure 3(b). Again, we plot different percentile branches on a *log* scale, with each data point representing the result over $5000$ problems. The hardest problems in the CP model can be three orders of magnitude harder than the hardest equivalently sized problems of fixed clause length [8].

These hard problems again tend to occur in regions of high percentage satisfiability. By comparison, the median displays a simple easy-hard-easy pattern. In [8], we give further experimental analysis to show that this highly variable and hard behaviour in CP cannot be eliminated by the use of better heuristics. We conjecture that this hardness arises from hard unsatisfiable problems in a region of otherwise satisfiable problems, or satisfiable problems which give hard unsatisfiable subproblems.

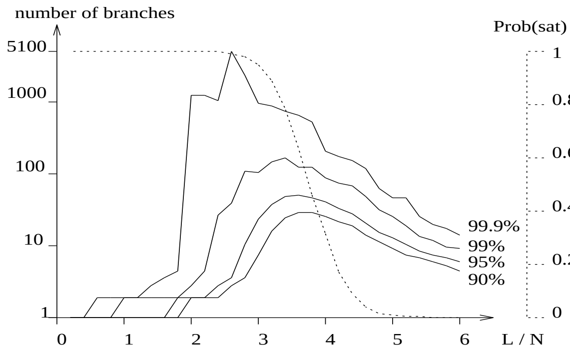



*Figure 5. CP, percentile branches,* $2Np = 3$, $N = 75$, *DP*

## 6 Conjectures about the Crossover

We have presented experimental evidence to show that many features of the phase transition for random $k$-SAT are also present in random mixed SAT and in the CP model. Our experiments suggest that for any distribution $\phi$ of clause lengths, there is a critical value $c_\phi$. In §3, we gave an upper bound on $c_\phi$ in terms of $d_\phi$, the density. In this section, we derive further bounds and conjecture an estimate for $c_\phi$ in terms of $c_k$.

For any $k$, if we have $c.\phi(k) > c_k$, then certainly $c > c_\phi$. This immediately gives us that, $c_\phi \leq \max_{k=1}^{\infty} \frac{c_k}{\phi(k)}$ For example, for 2-3-SAT, this gives us $c_\phi \leq c_2/\frac{1}{2} = 2$ and for 2-4-4-SAT, $c_\phi \leq c_2/\frac{1}{3} = 3$. These are tighter bounds than those obtained earlier by consideration of the density. Furthermore, they show that the value of $c_\phi$ cannot be determined by $d_\phi$ alone, as 3-SAT and 2-4-4-SAT have the same density, but it is known theoretically that $c_3 > 3$, whilst $c_\phi \leq 3$ for 2-4-4-SAT. We can also observe that $c_\phi \geq \min_{\phi(k)>0} c_k$. This bound takes no account of $\phi(k)$ except in being non-zero. Thus for both 2-3-SAT and 2-4-4-SAT we derive only $c_\phi \geq c_2 = 1$.

Neither of the above bounds, or the bound obtained earlier, give good predictions for $c_\phi$ in the cases we have tested. We would like to have a formula for $c_\phi$ in terms of $\phi$ and its constituent $c_k$. An intuitively appealing possibility is that $c_\phi$ is given by the weighted parallel sum of the constituent $c_k$. The intuition behind this conjecture is that $c_k.N$ clauses of length $k$ are "equivalent" in terms of preventing a truth assignment being a model as $c_2.N$ binary clauses. Thus, as $c_2 = 1$, a clause of length $k$ is $c_k$ times less effective at filtering out models as a binary clause. At the crossover, there are $c_\phi.N$ clauses of which $\phi(k).c_\phi.N$ are of length $k$. These clauses contribute to defeating possible models the same as $\phi(k).c_\phi.N/c_k$ binary clauses. In total, we need an effective contribution from all clauses which is the same as the combined effect of $N$ binary clauses. Thus, $\frac{\phi(2).c_\phi.N}{c_2} + \frac{\phi(3).c_\phi.N}{c_3} + \frac{\phi(4).c_\phi.N}{c_4} + \ldots = N$ or

$$\frac{1}{c_\phi} = \phi(2) + \frac{\phi(3)}{c_3} + \frac{\phi(4)}{c_4} + \ldots \qquad (2)$$

A physical analogy is that of electrical resistance. A set of clauses of length $k$ offer some "resistance" to whether a truth assignment is a model. The total resistance of a mixed SAT problem is the "sum" of the resistances of the sets of clauses of different lengths. We take the parallel sum since a truth assignment can be defeated by any of the sets independently.

This conjecture obeys all the firm bounds derived in this paper, very easily the two bounds derived in this section. Some tedious manipulation shows that the parallel sum also obeys the bound given by (1). Furthermore, the values it predicts seem to be close to those derived earlier in this paper from our experimental analysis. Using the values, $c_2 = 1, c_3 \approx 4.24$, $c_4 \approx 9.76$, we get the following results. For calculating $c_\phi$ for CP, we approximate $c_k$ for $k \geq 5$ using (1). The slight error between observed and predicted $c_\phi$ might be explained by interactions between clauses of different lengths.

| | Predicted $c_\phi$ | Observed $c_\phi$ |
|---|---|---|
| 2-3-SAT | $1.62$ | $1.76$ |
| 3-4-SAT | $5.91$ | $5.88$ |
| 2-4-4-SAT | $2.49$ | $2.74$ |
| CP | $2.67$ | $2.80$ |

## 7 Conjectures about Scaling

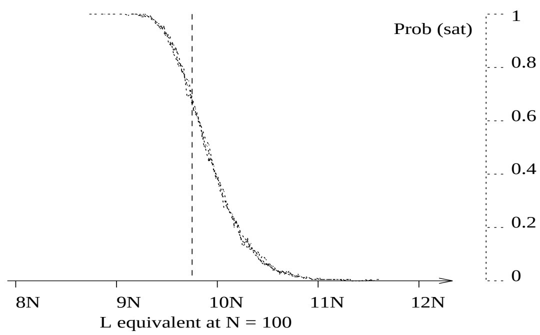



*Figure 6. Scaling of* 4-SAT, $\alpha_c = 9.76$, $v = 1.25$

Phase transitions occur frequently in natural systems. One of the most unusual and theoretically interesting systems is that of spin glasses. Each of the $N$ atoms in a spin glass has a magnetic spin which can have only one of two values, 'up' or 'down'. The system therefore has $2^N$ possible configurations. Macroscopic properties of a configuration (*eg.* the energy, entropy) depend only on interactions between the spins of nearest neighbours. Due to the differences in separation of the atoms, some of these interactions are ferromagnetic (promoting alignment of spins) whilst others are anti-ferromagnetic (promoting opposite spins). The net effect is a random force leading to a large number of equilibrium configurations. An analogy can be made between such spin glasses and random $k$-SAT. Each of the $N$ variables in a truth assignment has one of two values, 'True' or 'False'. The system therefore has $2^N$ possible configurations. Macroscopic properties (*eg.* satisfiability) depend only on the interaction between variables in each clause. Due to the random polarities of these variables, the net effect on a variable is a random "preference" towards 'True' or 'False'. [14] have used this analogy to suggest a fascinating scaling result for random $k$-SAT. They propose that for random $k$-SAT, there is a fundamental function $f$, and values $\alpha_c$ and $v$, such that

$$Prob(sat, R_k(N, L)) = f((L/N - \alpha_c)N^{1/v}) \qquad (3)$$

$f$ can be estimated experimentally and used to give accurate predictions of the value $Prob(sat, R_k(N, L))$. For 3-SAT, [14] report $\alpha_c = 4.15$, $v = 1.5$. Their value for $\alpha_c$ is slightly lower than recent experimental values of $c_3$ [4].

A restatement of (3) is that all graphs of $Prob(sat, R_k(N, L))$ will be identical if the $x$-ordinate used is $(L/N - \alpha_c)N^{1/v}$. Figure 6 shows our experimental data for 4-SAT for $N = 25, 50, 75$, scaled in this way. For convenience, we have multiplied the $x$-ordinate by $100^{-1/v}$, and added $\alpha_c$ so that the values on the $x$-axis give the equivalent value of $L$ at $N = 100$. The dashed line gives the point $\alpha_c$. By trial and error, we found a good fit using $\alpha_c = 9.76$, our experimentally observed value of $c_4$, and $v = 1.25$. Indeed, the curves are never more than $0.08N$ apart. We have also looked at problems generated from the random mixed SAT model. Figure 7 shows data for 2-3-SAT for $N = 50, 100, 150, 200$ using $\alpha_c = 1.76$ and $v = 2.5$. We tested $500$ problems in steps of $N/50$ clauses. The curves are never more than $0.06N$ apart. For 2-4-4-SAT, for $N$ up to $150$, we observed a good fit using $\alpha = 2.74$, $v = 3.5$. These graphs suggest that Kirkpatrick et al's conjecture can be extended from random $k$-SAT to random mixed SAT.

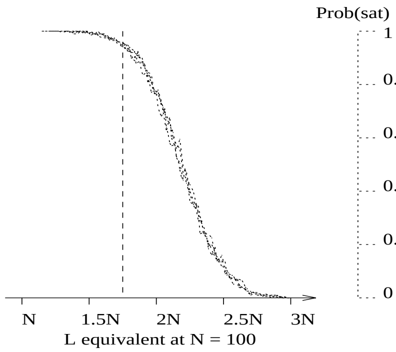



*Figure 7. Scaling of* 2-3-SAT, $\alpha_c = 1.76$, $v = 2.5$

## 8 Related Work

Phase transitions are attracting increasing attention in AI. Huberman and Hogg [12] predict that many large scale systems will undergo sudden phase transitions that affect computational performance. They show, for example, that a simple model of heuristic search changes from linear to exponential behaviour at a phase boundary. Cheeseman *el al.* [1] observed that many NP-hard problems have an order parameter, that hard problems occur at a critical value of this parameter, and that this value separates a region of underconstrained, typically soluble, problems from a region of overconstrained, typically insoluble, problems. Mitchell *et al.* [15] demonstrated that for random 3-SAT the order parameter is $L/N$, the ratio of clauses to variables, and that median performance of DP has an easy-hard-easy pattern with the hardest median instance occurring at the phase transition. Crawford and Auton [4] accurately identified the position of the phase transition for random 3-SAT as $L/N = 4.24$. [15] also report that the CP model has an easy-hard-easy pattern for median performance, but prematurely dismiss CP as being too easy compared with random $k$-SAT. Although median performance is easy, worst case performance is not [8]. The hardest CP problems can be orders of magnitude harder than the hardest comparably sized random $k$-SAT problems.

## 9 Conclusions

We have performed a detailed experimental investigation of the phase transition for randomly generated SAT problems. The sharp change from satisfiable to unsatisfiable problems previously observed at a critical value in random $k$-SAT problems is also present in the more general class of random mixed SAT problems, as well as in random problems generated by the constant probability model. We have used our experimental results to conjecture the critical value for these problem classes in terms of the critical values of the constituent $k$-SAT classes. Furthermore, we have been able to extend Kirkpatrick et al's conjecture on the scaling of transition behaviour to random mixed SAT.

As with random $k$-SAT, we have observed an easy-hard-easy pattern in the median difficulty for random mixed SAT problems, with the hardest problems being associated with the phase transition. However, the difficulty of problems of mixed clause lengths is much more variable than that of random $k$-SAT problems. Indeed, the hardest problems of mixed clause lengths can be orders of magnitude harder than comparably sized problems of fixed clause length. Such very hard problems tend to occur in otherwise easy regions where most problems are satisfiable. These results are of considerable value both to experimental and theoretical AI since the empirical comparison of algorithms for NP-hard problems and average-case analyses requires the identification of hard instances of randomly generated problems.

## Acknowledgements

This first author was supported by a SERC postdoctoral fellowship and the second by a HCM postdoctoral fellowship. We thank Alan Bundy, Pierre Lescanne, Bob Constable, and the members of the Mathematical Reasoning Group, Edinburgh (supported by SERC grant GR/H 23610), the Eureca group, INRIA-Lorraine, and the Department of Computer Science, Cornell University, for their constructive comments and for rather a lot of CPU cycles.

## REFERENCES


[1] P. Cheeseman, B. Kanefsky, and W.M. Taylor. Where the really hard problems are. In *Proceedings of the 12th IJCAI*, pages 163–169, 1991.

[2] V. Chvatal and B. Reed. Mick gets some (the odds are on his side). In *Proc. of the 33rd Annual Symposium on Foundations of Computer Science*, pages 620–627. IEEE, 1992.

[3] V. Chvatal and E. Szemeredi. Many hard examples for resolution. *JACM*, 35(4):759–768, 1988.

[4] J.M. Crawford and L.D. Auton. Experimental results on the crossover point in satisfiability problems. In *Proceedings of the 11th National Conference on AI*, pages 21–27, 1993.

[5] M. Davis and H. Putnam. A computing procedure for quantification theory. *JACM*, 7:201–215, 1960.

[6] O. Dubois, P. Andre, Y. Boufkhad, and J. Carlier. SAT versus UNSAT. In *Proc. of the Second DIMACS Challenge*, 1993.

[7] J. Franco and M. Paull. Probabilistic analysis of the Davis Putnam procedure for solving the satisfiability problem. *Discrete Applied Mathematics*, 5:77–87, 1983.

[8] I. P. Gent and T. Walsh. Easy problems are sometimes hard. Research paper 642, Dept. of AI, Edinburgh University, 1993.

[9] A. Goerdt. A threshold for unsatisfiability. In I. Havel and V. Koubek, editors, *Mathematical Foundations of Computer Science*, pages 264–274. Springer Verlag, 1992.

[10] T. Hogg and C. Williams. The Hardest Constraint Problems: A Double Phase Transition. Technical report, Xerox Palo Alto Research Center, 1993. PARC Preprint.

[11] J. N. Hooker and C. Fedjki. Branch-and-cut solution of inference problems in propositional logic. *Annals of Mathematics and Artificial Intelligence*, 1:123–139, 1990.

[12] B.A. Huberman and T. Hogg. Phase Transitions in Artificial Intelligence Systems. *Artificial Intelligence*, 33:155–171, 1987.

[13] T. Larrabee and Y. Tsuji. Evidence for a Satisfiability Threshold for Random 3CNF Formulas, Technical report, University of California, Santa Cruz, UCSC-CRL-92-42, 1992.

[14] S. Kirkpatrick, G. Györgyi, N. Tishby, L. Troyansky. The Statistical Mechanics of k-Satisfaction. *Advances in Neural Information Processing Systems*, 6, 1993.

[15] D. Mitchell, B. Selman, and H. Levesque. Hard and easy distributions of SAT problems. In *Proceedings, 10th National Conference on AI*, pages 459-465, 1992.