

Critical Behavior in the Satisfiability of Random Boolean Formulae

Scott Kirkpatrick*

IBM TJ Watson Research Center
Yorktown Heights, NY 10598
kirk@watson.ibm.com

Bart Selman

AT&T Bell Laboratories
Murray Hill, NJ 07974
selman@research.att.com

Abstract

The satisfiability of randomly generated Boolean formulae with k variables per clause is a popular testbed for the performance of search algorithms in artificial intelligence and computer science. For $k = 2$, formulae are almost always satisfiable when the ratio of clauses to variables is less than 1; for ratios larger than 1, the formulae are almost never satisfiable. We present data showing a similar threshold behavior for higher values of k . We also show how finite-size scaling, a method from statistical physics, can be used to characterize size dependent effects near the threshold. Finally, we comment on the relationship between thresholds and computational complexity.

*This work was done during a 1993 sabbatical at the Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem, 91904 Israel. A portion of the work was carried out at the Salk Institute, with support from the McDonnell-Pew Foundation.

Properties of randomly generated combinatorial structures often exhibit sharp threshold phenomena. The simplest such structure is a random graph, introduced by Erdős and Rényi (1), which is simply a list of N vertices and M edges, each edge connecting one pair of vertices, selected at random. Many properties of the connected clusters (groups of vertices connected by the edges) can be predicted with high precision for almost every graph constructed this way, if we make N and M large but hold their ratio, $\alpha \equiv M/N$, constant. Two regimes can be identified: when $\alpha < 1/2$, many small isolated clusters of maximum size $\ln N$ are found; when $\alpha > 1/2$, a single “giant component” with size αN absorbs many of the clusters. At the boundary between the two regimes, when α has its critical value, $\alpha_c = 1/2$, the largest clusters are of size $N^{2/3}$. Subsequent work (2, 3) has made precise the sharpness of the threshold: its characteristics persist across a range of α of order $N^{-1/3}$ about $\alpha_c = 1/2$. This is now recognized as the prototype of “percolation” thresholds observed in disordered materials (4). We shall use this well-understood model to test techniques for identifying critical phenomena in less-understood combinatoric structures.

Recently, threshold phenomena have also been observed in randomly generated Boolean formulae. Mitchell *et al.* (5) consider the k -satisfiability problem (k -SAT). An instance of k -SAT is a Boolean formula in conjunctive normal form (CNF), *i.e.*, a conjunction (logical AND) of M clauses, each consisting of the logical OR of k Boolean variables. Each variable is selected at random from a set of N “input” variables or their negations. The problem is to determine whether there is an assignment to the variables such that all clauses evaluate to true. M , N , and their ratio, $\alpha \equiv M/N$, define the scale and natural parameters of this problem just as in the random graph model. For example, an instance of $k = 2, N = 2, M = 3$ might be $(x \wedge y) \vee (\neg x \wedge \neg y) \vee (x \wedge \neg y)$, which is satisfied if x is true and y is false.

For randomly generated 2-SAT instances, it has been shown analytically that for large N , when the ratio $\alpha < 1$ the instances are almost all satisfiable, whereas for $\alpha > 1$, almost all instances are unsatisfiable (6, 7). For $k \geq 3$, a rigorous analysis has proven to be elusive. Loose upper and lower bounds have been obtained, but there is yet no rigorous proof of the existence of a threshold (8). Experimental evidence, however, strongly suggests a threshold with $\alpha \approx 4.2$ for 3SAT (5,9,10).

One of the main reasons for studying randomly generated CNF formulae is for their use in the empirical evaluation of combinatorial search algorithms (10, 11, 12). Formulae with $k = 3$ (3CNF) are good candidates for the evaluation of such algorithms because determining their satisfiability is an NP-complete problem (13). It is therefore widely believed that there does not exist an efficient (polynomial time) algorithm for determining the satisfiability of 3CNF formulae. This also holds for larger values of k . For $k = 1$ or 2, the satisfiability problem can be solved efficiently (14).

One has to be careful however in the use of randomly generated formulae: simple heuristic methods can often quite easily determine the satisfiability of most such formulae. This has led to some overly strong claims in the literature about handling very large formulae. Computationally challenging test instances can be obtained with high probability by generating formulae at or near the threshold (5). Cheeseman *et al.* (15) have made a similar observation of increased computational cost for heuristic search at a boundary between two distinct phases or behaviors of a combinatorial model. We will provide a precise characterization of the dependence on N of the threshold phenomena for k -SAT with k ranging from

2 to 6.

Our analysis will show that the threshold in k -SAT closely resembles the phase transition studied in spin glasses (16). We use finite-size scaling (4,17), a method from statistical physics in which observing how the width of a transition narrows with increasing sample size gives direct evidence for critical behavior at a phase transition. We first illustrate the finite-size scaling approach on random graphs.

The empirical observation behind phenomenological scaling is that sufficiently close to a threshold or critical point, systems of all sizes are indistinguishable except for an overall change of scale. In the random graph ensemble, the clusters of size $N^{2/3}$ that occur close to its threshold in a random graph with $N = 100$ should simply look like coarse versions of the clusters found in a graph with $N = 10000$. However, to make the comparison, the narrow threshold observed for very large graphs must be expanded in scale to compare it with the broader threshold seen in small graphs. In the random graph ensemble, we know exactly how to do this. Correcting for the known $N^{-1/3}$ dependence of the width, and normalizing α to its threshold value, α_c , we define a rescaled parameter, $y \equiv N^{1/3}2(\alpha - 1/2)$, against which to plot data for a graphical analysis.

But what to plot? A well-behaved quantity for this problem proves to be the sizes of the larger clusters, normalized to the size of the largest cluster in the same graph. When α is small, all clusters should be of the same general magnitude, so these ratios tend to unity. For $\alpha \gg 1/2$, normalizing by the size of the giant cluster makes all the ratios tend to zero. Let L_k denote the size of the k -th largest cluster found in a particular graph, and $\langle \rangle$ represent averaging over many samples of graphs with the same M and N . In Fig. 1a, we show $\langle L_2/L_1 \rangle$ for graphs with $N = 100, 1000, 10000$, and 40000 , plotted against α . 10000 samples were analyzed for each data point, giving roughly 1% accuracy throughout. The curves show the sharpening of the transition with increasing N and all intersect at the critical point, $\alpha = 1/2$. Plotting them against y in Fig. 1b, we find that the averaged normalized cluster sizes follow universal forms through the critical regions, only separating when $2|\alpha - 1/2| > N^{-1/3}$. The values of the size ratios at the critical point are insensitive to N , and appear to be new results. These are: $\langle L_2/L_1 \rangle = 0.538 \pm .0015$, $\langle L_3/L_1 \rangle = 0.382 \pm .001$, $\langle L_4/L_1 \rangle = 0.302 \pm .0006$, $\langle L_5/L_1 \rangle = 0.252 \pm .0006$ (averaged over 160,000 samples at each N).

It is somewhat surprising that finite-size scaling works here, since the standard heuristic derivations (4,17) explain the size dependence of a crossover between two phases as a measurement of a correlation length, ξ , which diverges at the critical point in an infinite system. If two points in such a system are separated by more than ξ , they are independent. Combinatoric problems like k -SAT have no lengths, and there is no geometric criterion for separating them into independent sub-problems. Yet it appears that power-law behavior in the size, N , replaces scaling with respect to a length. This is consistent with renormalization-group derivations of finite-size scaling (18), in which lengths occur only through the volume of the system, which here is measured by N .

We shall now apply this procedure to k -SAT, determining the critical concentration by exact calculation if possible, by observation of the measured properties if not. For example, the intersection of all the lines in Fig. 1a identifies the critical point even if α_c were not known to be $1/2$.

We generated extensive data on the satisfiability of randomly generated k -CNF formulae with k ranging from 2 to 6. Fig. 2 shows the fraction of random k -SAT formulae that is *unsatisfiable* as a function of the ratio, α . For example, the left-most curve in Fig. 2 shows the fraction of formulae that is unsatisfiable for random 2CNF formulae with 50 variables over a range of values of α . Each data point was generated using 10000 randomly generated formulae, giving roughly 1% accuracy. We used a highly optimized implementation (9) of the Davis-Putnam procedure (19). This procedure performs a backtrack search through the space of possible truth assignments, and is the fastest known complete procedure for satisfiability testing on many classes of formulas (20).

Fig. 2 shows a clear threshold for each value of k . Except for the case $k = 2$, the curves cross at a single point and sharpen up with increasing N . For $k = 2$, the intersections between the curves for the largest values of N seem to be converging to a single point as well, although the curves for smaller N deviate. The thresholds move rapidly to the right with increasing k . This is because a clause with k distinct variables prohibits only one of the 2^k truth assignments to the k variables, thus the constraints get weaker as k increases.

We can estimate the behavior of the threshold and crossover for large values of k by neglecting the overlap between clauses. This is called an annealed estimate, by analogy with annealed theories of materials (16) which average independently over sources of disorder. Each of the M clauses reduces the expected number of satisfying input configurations from 2^N by a factor of $(2^k - 1)/(2^k) = (1 - 2^{-k}) \equiv \gamma_k$. We get a plausible estimate for the threshold by asking when, on average, only one satisfying configuring configuration survives. Since $2^N \gamma_k^M = 2^{N(1 + \alpha \log_2 \gamma_k)}$, this occurs at $\alpha_{ann} = -1/\log_2 \gamma_k \approx 2^k \ln 2$. α_{ann} is identical to the upper bound described in (6). We have marked $\alpha_{ann}(k)$ with arrows in the figures, and tabulate it in Table 1.

The annealed estimate can be extended (21) to describe the crossover for large k . The probability that there are no satisfying configurations is $(1 - \gamma_k^{\alpha N})^{2^N}$. This can be transformed into

$$\begin{aligned} (1 - 2^{N(\alpha/\alpha_c)})^{2^N} &= (1 - 2^{-N(\alpha - \alpha_c)/\alpha_c} / 2^N)^{2^N} \\ &\sim e^{-2^{-y_{ann}}}, \end{aligned}$$

where

$$y_{ann} = N(\alpha - \alpha_c)/\alpha_c .$$

The dimensionless rescaled parameter to use in finite-size scaling is just a slight generalization of this (4,17):

$$y = N^{1/\nu}(\alpha - \alpha_c)/\alpha_c .$$

Fig. 3a shows the threshold in more detail over a range of sample sizes for random 3CNF formulae. Both the threshold shift and the increasing slope in the curves of Fig. 3a can be accounted for by finite-size scaling. Values for α_c and ν must be derived from the experimental data. First α_c is determined as the crossing point of the curves for large N in Fig. 3a. Then ν is determined to make the slopes match up through the critical region. In Fig. 3b we find that these two parameters capture both the threshold shift and the steepening of the curves, using $\alpha_c = 4.17$ and $\nu = 1.5$. Rescaling the data for random 4SAT, in Fig. 3c, leads to a tighter fit to a single curve. In Table 1 we give the critical parameters obtained from this analysis for values of k from 2 to 6. The error bars show the range of each parameter

k	α_{ann}	α_c	y_{50}	ν
2	2.41	1.0	2.25	$2.6 \pm .2$
3	5.19	$4.17 \pm .03$	0.74	$1.5 \pm .1$
4	10.74	$9.75 \pm .05$	0.67	$1.25 \pm .05$
5	21.83	$20.9 \pm .1$	0.71	$1.1 \pm .05$
6	44.01	$43.2 \pm .2$	0.69	$1.05 \pm .05$

Table 1: Critical parameters for random k -SAT.

over which the best fits were obtained. Note that ν appears to be tending to 1, the annealed limit, and α_{ann} becomes an increasingly good approximation to α_c as k increases.

The rescaled curves in Figs. 3b and 3c are similar in form. Combining the rescaled curves for all values of k in Figure 4, we find that the curves for $k \geq 3$ all roughly coincide. As $k \rightarrow \infty$, the curves approach the annealed limit derived above. The curve for $k = 2$ is also similar, but shifted to the right from the others.

From the perspective of performance evaluation for search algorithms, the point where 50% of the formulae are unsatisfiable is thought to be where the computationally hardest problems are found (5, 15). Note that the 50% point lies somewhat to the right of the scale-invariant point (the point where the curves cross each other), and shifts with N . Since the fraction of unsatisfiable formulae is given by $f_k(y)$, where the invariant function, f_k , is that graphed in Fig. 4 for each k , a description of the 50% threshold shift follows immediately. If we define y_{50} by $f_k(y_{50}) = 0.5$, then $\alpha_{50} = \alpha_c(1 + y_{50}N^{-1/\nu})$. From Fig. 3b or Table 1, we find that, for $k = 3$, $\alpha_{50} \approx 4.17 + 3.1N^{-2/3}$. Crawford and Auton (9) fit their data on the 50% point as a function of N by arbitrarily assuming that the leading correction will be $O(1/N)$. They obtain $\alpha_{50} = 4.24 + 6/N$. However, the two expressions differ by only a few percent as N ranges from 10 to ∞ . For $k = 2$, the difference between the scaling expression and a $1/N$ extrapolation will be greater.

We have shown how finite-size scaling methods from statistical physics can be used to model the threshold phenomenon in randomly generated k -SAT problems. Given the good fit of our scaling analysis, we conjecture that this method can also be of use in characterizing phase transitions in other combinatorial problems of interest to the computer science community.

Several authors have attempted to relate computational complexity to the characteristics of phase transitions in models of disordered systems. Fu and Anderson (22) have proposed spin glasses (magnets with 2-spin interactions of random sign) as having inherent exponential complexity. Also, computationally hard graph coloring problems and 3SAT problems were found at or near phase transitions (5, 15). Huberman and colleagues (23) were first to focus on the diverging correlation length seen at continuous phase transitions as the root of computational complexity. In fact, both effects can play important roles, but are not sufficient and may not even be necessary.

There are NP-complete problems (*e.g.*, travelling salesman or max-clique) which lack a clear phase boundary at which “hard problems” cluster. Percolation thresholds are phase

transitions, yet the cost of finding the largest cluster never exceeds N steps. Similarly, finding a satisfying assignment, or proving its non-existence, for any 2-SAT instance, can be done in linear time. The spin glass Hamiltonians studied in (22) are similar to our 2-SAT formulae, but the questions studied are different and computationally much harder. Finding an assignment which falsifies the minimum number of clauses is like finding the ground state in a spin glass phase, and does not reduce to a single search on the directed graph of Aspvall (14). For 2-SAT, in fact, finding such “ground states” is NP-hard. Therefore, if both diverging correlations (diverging in size if no lengths are defined) and a “spin-glass” phase occurs, we expect search to be exponentially difficult.

References

1. P. Erdős and A. Rényi, *Publ. of the Math. Inst. of the Hung. Acad. Sci.*, **5**, 17 (1960).
2. B. Bollobás, *Random Graphs* (Academic Press, London, 1985), and J. Spencer, *Ten Lectures on the Probabilistic Method* (Soc. for Ind. and Appl. Math., Philad., 1987).
3. S. Janson, D.E. Knuth, T. Łuczak, B. Pittel, *Random Structures and Algorithms*, **4**, 231 (1993).
4. D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1992).
5. D. Mitchell, B. Selman, and H.J. Levesque, *Proc. 10th Natl. Conf. on Artif. Intel.* (AAAI Press, Menlo Park, 1992), p. 459.
6. V. Chvátal and B. Reed, *Proc. Ann. ACM Symp. on the Theo. of Comput.* (Assoc. for Comput. Mach., New York, 1992), p. 620.
7. A. Goerdt, *Proc. 17th Int. Symp. on the Math. Found. of Comp. Sc.* (Springer-Verlag, Berlin, 1990).
8. A.Z. Broder, A.M. Frieze, E. Upfal, *Proc. 4th Ann. ACM-SIAM Symp. on Discr. Alg.*, (Soc. for Ind. and Appl. Math., Philad., 1993), p. 322.
9. J.M. Crawford and L.D. Auton, *Proc. 11th Natl. Conf. on Artif. Intel.* (AAAI Press, Menlo Park, 1993), p. 21.
10. T. Larrabee and Y. Tsuji, *Proc. AAAI Symp. on AI and NP-hard problems*, 112 (1993).
11. B. Selman, H.J. Levesque, D. Mitchell, *Proc. 10th Natl. Conf. on Artif. Intel.* (AAAI Press, Menlo Park, CA, 1992), p. 440.
12. A.P. Kamath, N.K. Karmarkar, K.G. Ramakrishan, M.G.C. Resende, *Annals of O.R.*, **25**, 43 (1990).
13. S.A. Cook, *Proc. 3rd Ann. ACM Symp. on the Theo. of Comput.*, 151 (1971). (Assoc. for Comput. Mach., New York, 1971), p. 151.
14. B. Aspvall, M.F. Plass, R.E. Tarjan, *Inform. Process. Let.*, **8**, 121 (1979).
15. P. Cheeseman, B. Kanefsky, W.M. Taylor, *Proc. 12th Int. Conf. on Artif. Intel.* (Morgan Kaufmann, San Mateo, CA, 1991), p. 331.
16. M. Mézard, G. Parisi, M.A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1986).
17. S. Kirkpatrick, and R.H. Swendsen, *Comm. ACM*, **28**, 363 (1985).

18. M.N. Barber, in *Phase Transitions and Critical Phenomena*, **8**, C. Domb and J. L. Lebowitz (Eds.) (Academic Press, London, 1983), p. 145.
19. M. Davis and H. Putnam, *J. Assoc. Comput. Mach.*, **7**, 201 (1960).
20. O. Dubois, P. Andre, Y. Boufkhad, J. Carlier, *Proc. 2nd DIMACS Implementation Challenge*, Piscataway, NJ (1993).
21. S. Kirkpatrick, G. Györgyi, N. Tishby, L. Troyansky, to appear in *Advances in Neural Information Processing Systems* **6**, 1993, and unpublished work.
22. Y. Fu, in *Lectures in the Sciences of Complexity*, D. Stein (Ed.) (Addison-Wesley, 1989), p. 815.
23. S.H. Clearwater, B.A. Huberman, T. Hogg, *Science*, **254**, 1181 (1991).

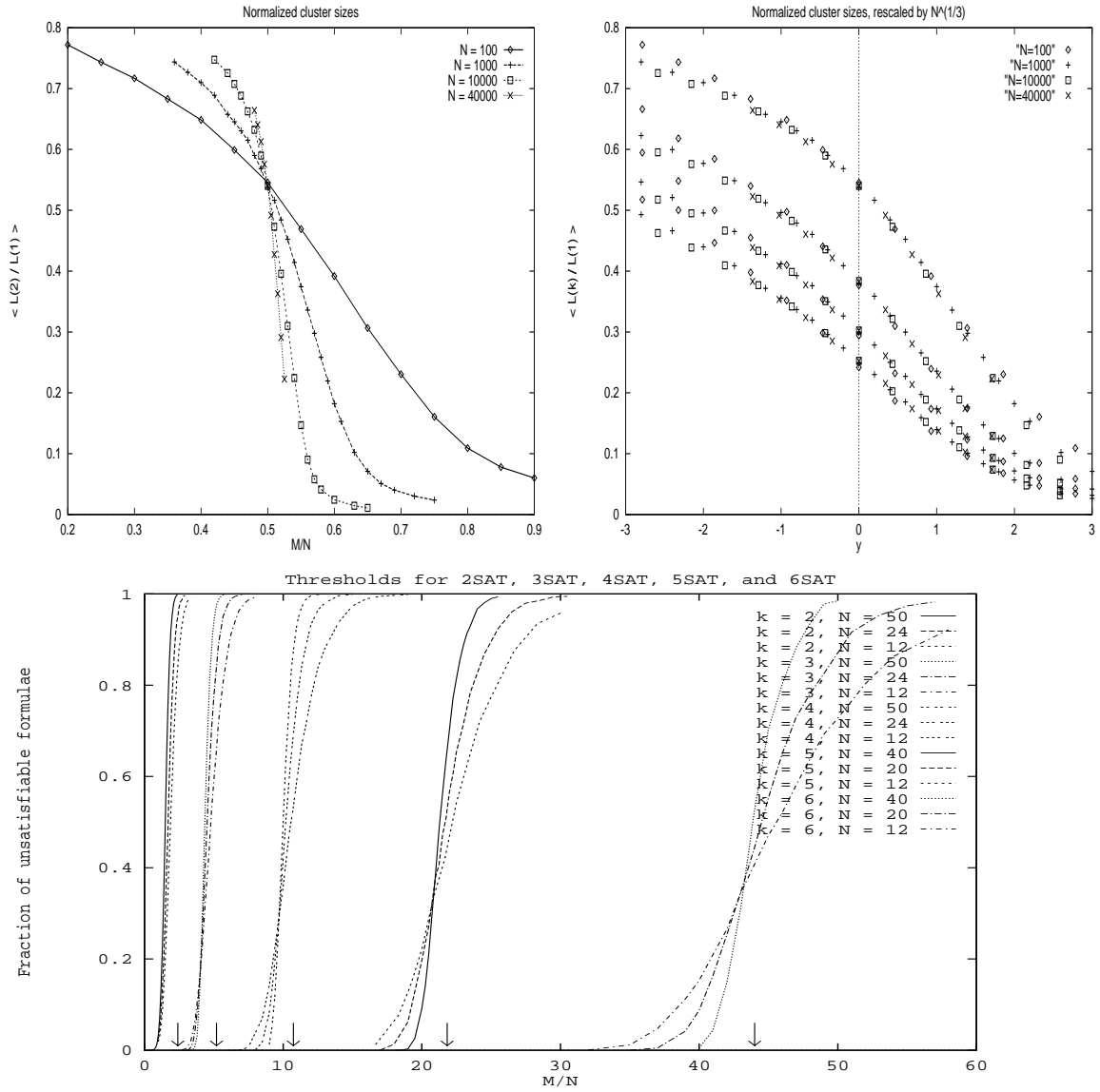


Fig. 1a: The ratio of the size of second largest cluster to that of the largest cluster.
 Fig. 1b: Rescaled cluster data. ($(k-1)^{\text{st}}$ curve from top gives rescaling for $\langle L_k/L_1 \rangle$ data.)
 Fig. 2: Fraction of unsatisfiable formulae for (left to right) 2-, 3-, 4-, 5-, and 6-SAT.

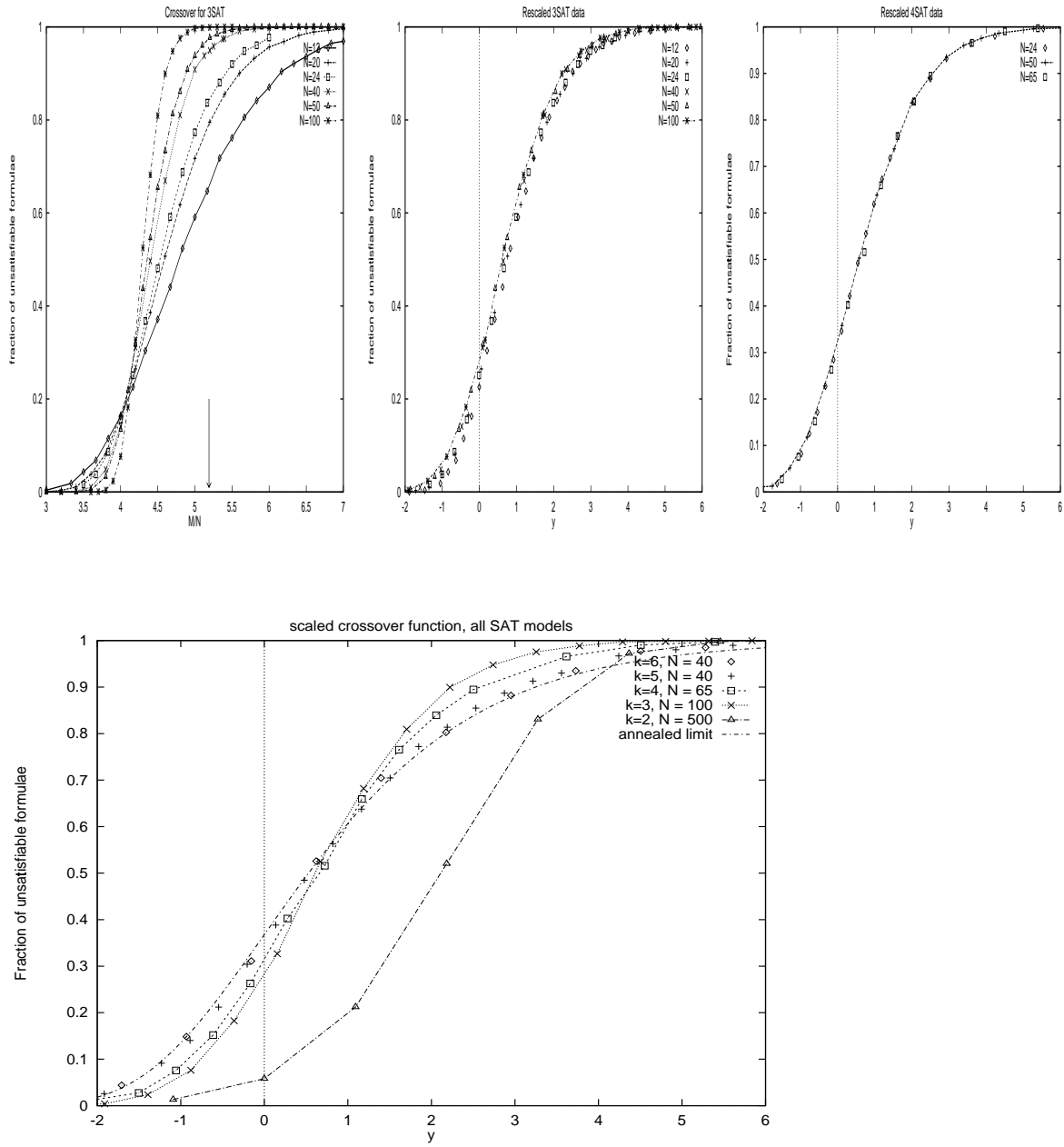


Fig. 3a: Threshold data for 3-SAT, $N = 12$ to 65.

Fig. 3b: Rescaled 3-SAT data using $\alpha_c = 4.17$, $\nu = 1.5$.

Fig. 3c: Rescaled 4-SAT data using $\alpha_c = 9.75$, $\nu = 1.25$.

Fig. 4: Rescaled crossover functions for $k = 2, 3, 4$, and $\langle k \rangle = 4$ (see (5)).