The unsatisfiability threshold conjecture: the techniques behind upper bound improvements

Lefteris M. Kirousis *   Yannis C. Stamatiou *†   Michele Zito ‡

Abstract

One of the most challenging problems in probability and complexity theory concerns the establishment and the determination of the *satisfiability threshold* for random Boolean formulas consisting of clauses with exactly k literals, or k-SAT formulas with emphasis on the case $k = 3$, or 3-SAT. According to many experimental observations, there exists a *critical* value r_k of the number of clauses to the number of variables ratio $r = m/n$ such that almost all randomly generated formulas with $r > r_k$ are unsatisfiable while almost all randomly generated formulas with $r < r_k$ are satisfiable. The statement that such a crossover point really exists is called the "satisfiability threshold conjecture". While experiments hint at such a direction, as far as theoretical work is concerned, progress has been difficult. Up to now, there are rigorous proofs of only successively better upper and lower bounds for the value of the (conjectured) threshold although, in an important advance, Friedgut proved that the phase transition is sharp (without showing the existence of a *fixed* transition point). In this work, our goal is to review the series of improvements of the upper bounds for 3-SAT and the techniques from which the improvements resulted. We give only a passing reference to the improvements of the lower bounds, as they rely on significantly different techniques that would require much more space to present.

1 Introduction

Let ϕ be a random 3-SAT (in general k-SAT) formula constructed by selecting uniformly and with replacement m clauses from the set of all possible clauses with three (generally k) literals (no variable repetitions are allowed within a clause) over n variables. It has been experimentally observed that as the numbers n, m of variables and clauses respectively tend to infinity while the ratio m/n (the *density* of the formula) is fixed to a constant r, the property of satisfiability exhibits a phase transition: if r is greater than a number which has been experimentally determined to be approximately 4.25, then almost all random 3-SAT formulas are unsatisfiable (i.e. the fraction of unsatisfiable formulas tends to 1) while the opposite is true if $r < 4.25$. Analogous

*University of Patras, Department of Computer Engineering and Informatics, University Campus, GR-265 04 Patras, Greece. Research supported by the University of Patras, Research Committee (Project C. Carathéodory no. 2445).
e-mail: {kirousis,stamatiu}@ceid.upatras.gr

†Computer Technology Institute, Kolokotroni 3, 263 21, Patras, Greece.

‡Department of Computer Science, University of Liverpool, UK. e-mail: M.Zito@csc.liv.ac.uk

phenomena have been observed for k-SAT with $k > 3$ (the experimentally determined threshold point is increasing with k). The experiments that led to the above conclusions were initiated by the work of Cheeseman *et al.* [10]. For detailed numerical results see [13, 40]). For $k = 2$, it has been *rigorously* established independently by Chvátal and Reed [11], Goerdt [26, 27], and Fernandez de la Vega [18] that a transition from almost certain satisfiability to almost certain unsatisfiability takes place at a clause to variable ratio equal to 1.

For $k \geq 3$, finding the exact value of the threshold point where this transition occurs or even proving that such a threshold exists is still an open problem. Friedgut [22] has shown that for k-SAT the transition is sharp, i.e. in the large n limit, the probability of satisfiability changes from arbitrarily close to 1 to arbitrarily close to 0, as the density r moves along arbitrarily short intervals. However, it is not known whether these intervals converge to a fixed point. Also Istrate [29] has shown that the transition is *first order*, i.e. as r moves along along these intervals of asymptotically zero length, the value of a certain combinatorial parameter defined in terms of the random formula jumps from zero to a non-zero multiple of n. Such parameters are known in Statistical Physics as *order parameters*. The specific one used in 3-SAT is the *spine* of the formula (the spine is defined to be the cardinality of the set of variables x for which a sub-formula ψ of the given random formula can be found so that x is false in every truth assignment that satisfies ψ). Furthermore, recent theoretical work in Statistical Physics has supplied additional and almost conclusive evidence (but not a formal proof in the mathematical sense) for the existence of the threshold point (see Mézard *et al.* [39]).

Apart from the above results, much effort has been put into rigorously establishing upper and lower bounds for the region where the transition for k-SAT occurs. These efforts resulted in interesting and novel probabilistic techniques. In this paper we will mainly concentrate on presenting the upper-bound results and the techniques that lead to them (see also the review by Dubois [16]).

2 Generating random 3-SAT formulas

Let Ω denote the set of all $2^3\binom{n}{3}$ possible 3-SAT clauses. A random 3-SAT formula ϕ on $m = rn$ clauses, can be formed using one of the following frequently employed probability models:

1. Select the m clauses of ϕ by drawing them uniformly, independently of each other and with replacement from Ω – model G_{mm}.

2. As above, but with no replacement – model G_m.

3. Each clause in Ω gets in ϕ independently of the others and with probability p – model G_p.

4. There are $3rn$ possible clause positions for literals (rn clauses, each having 3 literals) that can be filled at random with a literal chosen at random from within the set of $2n$ possible literals over n variables – a model that actually does not have a universally accepted name but we will simply call $G_{3,rn}$ for our purposes.

All these models are variations of the, so-called, *fixed clause length* model that was introduced by Franco and Paull in [21]. This model is an adaptation of a classical model for random graphs introduced by Erdös and Rényi in a series of seminal papers published in the 1960's (see the book of Bollobás [6] for the historical development of the field of random graphs).

The fixed clause length model is in sharp contrast with the *constant density* model introduced by Goldberg *et al.* in [28] in order to study the average time complexity of satisfiability algorithms. According to the constant density model, each of a *fixed* number of clauses is formed by letting all the literals decide independently of each other to enter the clause with some probability. This model has the disadvantage of inducing a probability distribution on the set of all boolean formulas on a given number of variables that favours easy instances, a feature not shared by the fixed clause length model that allows the manipulation of the instance hardness by means of the *clause density* parameter r (clause to variable ratio).

Each of these models has its own distinct advantages and disadvantages. G_{mm}usually leads to tighter results than G_p. The latter on the other hand, has the important property of independence for events involving *non-intersecting* sets of clauses while such events may be dependent in G_{mm}. Finally $G_{3,rn}$, as we will see later, enables one to study as well us manipulate individual literal appearances in a formula. This fact leads to a finer description of the formula than the detail the other models can achieve. This (as we will see in Section 6) may lead to better upper bound values as one usually applies the techniques we will examine on a more limited and well-defined set of formulas. However, it is well known that if a threshold exists in anyone of the above models, it exists in all of them and its value is equal in all of them (however, as already said, bounds obtained by the same method may differ from model to model).

In the sections where we examine the rigorous techniques that have been used in order to bound from above the satisfiability threshold, we will see examples where the advantages and disadvantages mentioned above arise and they are exploited or circumvented, respectively.

3 The successive threshold approximations

The basic mathematical tool employed for bounding from above the transition region of k-SAT (conjectured to be a single point in the limit, which for reasons of notational convenience we will assume that it has been shown to exist and denote it by r_k) is a probabilistic technique known as the *first moment method*. This method makes use of *Markov's inequality*: let X be a nonnegative integer random variable and let $\mathbf{E}[X]$ be the expectation of X, then $\mathbf{Pr}[X \geq 1] \leq \mathbf{E}[X]$. Actually, Markov's inequality is applied on a sequence of random variables $X = X_n, n = 0, 1, \ldots$ which depends on certain parameters sometimes referred to as *control parameters*. Now, if one finds a condition on the control parameters that forces $\mathbf{E}[X]$ to approach zero as n approaches infinity, then the probability that X is nonzero also converges to 0 in the large n limit, whenever the condition holds. Despite its simplicity, the first moment method is a powerful tool that quickly provides us with some condition (most often not the tightest possible) for proving that some random variable is almost certainly zero, asymptotically.

The connection of the first moment method with the satisfiability threshold conjecture was observed by a number of researchers, including Franco and Paull [21], Simon *et al.* [44], Chvátal and Szemerédi [12]. Let ϕ be a random 3-SAT formula on n variables generated according to G_{mm} and let $X(\phi)$ be the random variable that counts the number of satisfying truth assignments of ϕ. The probability that a truth assignment satisfies a clause is $\frac{7\binom{n}{3}}{8\binom{n}{3}} = \frac{7}{8}$. Therefore, the expected value of $X(\phi)$ is $2^n(7/8)^m$. Since $\mathbf{Pr}[\phi \text{ is satisfiable}] = \mathbf{Pr}[X(\phi) \geq 1]$, from Markov's inequality we get that

$$\mathbf{Pr}[\phi \text{ is satisfiable }] \leq 2^n \left(\frac{7}{8}\right)^{rn} \tag{1}$$

If by r_M we denote the exact solution of the equation $2(7/8)^r = 1$ (where r_M is approximately 5.19), then we observe that under the condition $r > r_M$, the right-hand side of Equality (1) tends to zero, which establishes the value r_M as an upper bound for the critical value r_3.

It is perhaps instructive at this point to provide the Markov inequality computations for model G_p as an example of the difference in accuracy that can be obtained using various random models. In G_p, the probability that a truth assignment satisfies a random formula is $(1-p)^{\binom{n}{3}}$, which is the probability that none of the $\binom{n}{3}$ clauses not satisfied by the assignment is part of the formula. Furthermore we will set $p = \frac{6r}{8n^2}$. For such a choice of the selection probability it holds that if the event "ϕ is satisfiable" has a vanishingly small probability in the G_p model, then the probability of this event is also small in G_m and G_{mm} for $m = rn$ and in $G_{3,rn}$. By Markov's inequality in G_p we have

$$\mathbf{Pr}[\phi \text{ is satisfiable }] \leq \mathbf{E}[X(\phi)] = 2^n\,(1-p(n))^{\binom{n}{3}} = 2^n\left(1-\frac{6r}{8n^2}\right)^{\binom{n}{3}} \leq 2^n e^{\frac{n}{8}r} \tag{2}$$

leading to the inequality $r_3 < 5.54$. Equations (1) and (2) provide a simple example of a frequently occurring trade-off among the various probabilistic models: accuracy of results vs. easiness in handling complicated situations, like the computation of the probability of conjunctions of events.

The first observation that the inequality $r_3 < 5.19$ is not the best possible, came from Broder, Frieze, and Upfal in [7], who pointed out that the condition $r > r_M - 10^{-7}$ is sufficient to guarantee that $\mathbf{Pr}[\phi \text{ is satisfiable}]$ tends to zero. El Maftouhi and Fernandez de la Vega obtained a further improvement, by showing in [17] that the condition can be relaxed to $r > 5.08$. Then Kamath *et al.* in [32] obtained the improved condition $r > 4.758$ using a numerical computation while they also gave an analytical proof of the condition $r > 4.87$. Using a refinement of Markov's inequality based on the definition of a restricted class of satisfying truth assignments, Kirousis, Kranakis, and Krizanc ([35]) proved an upper bound value of 4.667. Using the same class of satisfying truth assignments, after more accurate but lengthier computations, Dubois and Boufkhad [14] independently obtained the upper bound 4.642. Also, Kirousis *et al.* [37] give the bound 4.602 by what they call "the method of local maxima". Later, in [31], Janson, Stamatiou, and Vamvakari lowered this value to 4.596 through two different approaches: by viewing a formula as a physical spin system and then using an optimization technique from statistical physics to compute an asymptotic expression for its energy and by obtaining an improved upper bound to the Rogers-Szegö polynomials. In Zito's doctoral thesis [47] the upper bound was further improved to about 4.58 while in [34], Kaporis *et al.* obtained the value 4.571 using a new upper bound for the q-binomial coefficients obtained in [38]. Finally Dubois, Boufkhad, and Mandler [15] give an upper bound of 4.506 using an approach involving formulas with "typical" number of appearances of signed occurrences of their variables.

For general k, Franco and Paull [21] used the first moment method and derived an upper bound for the value of the satisfiability threshold of k-SAT equal to $2^k \ln 2$ while the same derivation was also observed by Simon *et al.* [44] and Chvátal and Szemerédi [12]. Kirousis, Kranakis, Krizanc, and Stamatiou [37] and, independently, Dubois and Boufkhad [14] gave techniques that improved this general upper bound without, however, improving the leading term that, in both approaches, is again equal to $2^k \ln 2$.

On the lower bounds side, Chao and Franco [8, 9] were the first to analyze the asymptotic behaviour of some algorithms that apply a heuristic in order to iteratively assign a truth value to all the variables of a formula. One of the algorithms they analyzed applied the UNIT CLAUSE heuristic and they showed, using a technique relying on differential equations in order to model the workings of their algorithm, that the algorithm succeeds with *positive* probability (but *not* almost certainly) for formulas with clause to variable ratio less than 2.9. Following this, the first lower bound for 3-SAT was established by Franco [19] who analyzed an algorithm that satisfies only literals whose complements do not appear in the formula ("pure literals"). He showed that for clause to variable ratios below 1, the algorithm succeeds almost certainly in finding a satisfying truth assignment to all the variables. Then Broder, Frieze, and Upfal [7] showed that the "pure literal" heuristic actually succeeds in satisfying a formula almost certainly if the ratio is smaller than 1.63. Frieze and Suen [23] improved the lower bound to 3.003 by analyzing the GENERALIZED UNIT CLAUSE heuristic (GUC) with limited backtracking and showing that it succeeds almost certainly for ratios lower than 3.003. Finally, using the differential equations method developed by Wormald [46] for approximating the evolution of discrete random processes, Achlioptas [1] and Achlioptas and Sorkin [4] reached the values 3.143 and 3.26 respectively. They developed a framework for a special class of algorithms they call "myopic" and they showed that no algorithm in this class can succeed in satisfying almost certainly formulas with clause to variable ratios larger than 3.26. Recently, Kaporis, Kirousis, and Lalas [33] analyzed a greedy and simple heuristic, where the literal that is selected to be satisfied at each step is one with maximum number of occurrences in the formula. They obtained the upper bound of 3.42. This was the first time that a heuristic that makes use of information related to the number of appearances of literals in a random formula (*degree sequence*) is analyzed. With a little more complicated greedy heuristics that at each step satisfy a literal with a large degree but whose negation has a small degree, an upper bound of more than 3.5 can be attained. This is the currently best value.

The best currently known *general* lower bound for k-SAT, for any fixed value of k, is given by a recent result by Achlioptas and Moore [3] who showed that $r_k \geq \frac{\ln 2}{2} 2^k - c$, for some constant $c > 0$ independent of k. This result essentially bridged the asymptotic gap between the $2^k \ln 2$ general upper bound and the $1.817(2^k/k)$ previously best general lower bound obtained by Frieze and Suen [23]. Moreover, Frieze and Wormald [24] proved that r_k is asymptotic to $2^k \ln 2$ if k is a function of n and $k - \log_2 n \to \infty$. Both results are the first successful efforts (to the best of our knowledge) in applying the *second moment method* in order to prove a lower bound to the satisfiability threshold, something that up to now was feasible only through the probabilistic analysis of satisfiability algorithms relying on specific heuristics in order to satisfy a random formula (see our discussion above on lower bounds). Using a technique known in physics as the *replica method*, Monasson and Zecchina predict in [41] that the asymptotic (in k) expression for the threshold is equal to $2^k \ln 2$ although their approach is not a rigorous one.

4 Approaches based on the harmonic mean

Although simple to apply, the first moment method does not lead to the best possible upper bounds for our problem. For values of the clause to variable ratio smaller than r_M as defined in the previous section, the expected number of satisfying truth assignments of a random formula tends to infinity although the empirical evidence suggests that most of such formulas have no satisfying truth assignment at all. This is due to the fact that there exist very rare formulas which are *satisfiable* and have a large number of satisfying assignments. El Maftouhi and Fernandez de la Vega [17] and, independently, Kamath *et al.* [32] studied this situation in detail. They resorted to the harmonic mean formula, first introduced (or formalized) by Aldous [5] to take this problem into better account.

Aldous result. Let $(B_i \in I)$ be a finite family of events in a probability space. For a permutation π of I, call (B_i) *invariant under* π if:

$$\mathbf{Pr}[B_{i_1} \cap B_{i_2} \ldots \cap B_{i_r}] = \mathbf{Pr}[B_{\pi(i_1)} \cap B_{\pi(i_2)} \ldots \cap B_{\pi(i_r)}]$$

for all $r \geq 1$ and $i_1, \ldots, i_r \in I$. Call the family (B_i) *transitive invariant* if for each $i_1, i_2 \in I$ there exists π such that $\pi(i_1) = i_2$ and (B_i) invariant under π. In particular transitive invariance implies that $\mathbf{Pr}[B_i] = p$ is actually independent of i.

Let N be the random variable counting the number of B_i's which occur. Then, if $(B_i : i \in I)$ is a transitively invariant family of events (with $p = \mathbf{Pr}[B_i]$, independent of i)

$$\mathbf{Pr}[\bigcup_{i \in I} B_i] = p \cdot |I| \cdot \mathbf{E}[N^{-1} | B_{i_0}].$$

The method gives an (alternative) expression for $\mathbf{Pr}[\phi \text{ is satisfiable}]$, if one interprets the B_i as the events "ϕ is satisfied by assignment A_i". Notice that $X(\phi)$ (i.e. the number of satisfying truth assignments of ϕ) is denoted by $\#F$ in [32] and $|\mathcal{M}od(\mathcal{F})|$ in [17]. Let T_i be the set of formulas which are satisfied by the "ith" truth assignment (say, w.r.t. reverse lexicographic order). Thus T_1 corresponds to the assignment setting all variables to TRUE. The following is a restatement of Aldous result, in the context 3-SAT formulas:

$$\mathbf{Pr}[\phi \text{ is satisfiable }] = \mathbf{E}[X(\phi)] \times \sum_{\psi \in T_1} \frac{1}{X(\psi) \cdot |T_1|}.$$

Notice that an expression equivalent to the equation above is the following (this is the one proved explicitly in [17, Equation (1)]):

$$\mathbf{E}_{\phi \in SAT}[X(\phi)] = \frac{|T_1|}{\sum_{\psi \in T_1} \frac{1}{X(\psi)}}$$

where $\mathbf{E}_{\phi \in SAT}[X(\phi)]$ is the expectation of $X(\phi)$ w.r.t. all satisfiable formulas on n variables and m clauses. The authors in [17] prove that it is possible to define a class of formulas $T_1^* \subseteq T_1$ of size at least $(1 - 2^{-\delta n})|T_1|$, where δ is a constant, such that each formula in T_1^* has at least $2^{\delta n}$ satisfying truth assignments. This implies: $\mathbf{E}_{\phi \in SAT}[X(\phi)] \geq 2^{\delta n - 1}$ (see Subsection 4.1). Therefore the probability that a random 3-SAT formula be satisfiable is at most

$$2^n \left(\frac{7}{8}\right)^{rn} 2^{1-\delta n}$$

The authors, using a simple random experiment, manage to compute $\delta = 0.02137$ and the above inequality gives the improved result $r_3 \leq 5.0798$. The different quality of the bounds derived in [17] and [32] is due not only to the use of coarse upper bounds, instead of exact asymptotics, for the estimate of the proportion of "interesting" formulas with a particular structure performed in [17] (in fact it can be proved that the difference between the two is vanishingly small) but also to the different experiment used to count this proportion. In the following sections we report, briefly, the results in the two papers. The careful reader will be able to pick up the similarities and the differences in the two approaches.

4.1 Accounting for rare formulas with many satisfying truth assignments: dispensable variables

In [17], El Maftouhi and Fernandez de la Vega define a subset T_1^* of T_1 such that $|T_1^*| \geq (1 - 2^{-\delta n})|T_1|$ and all formulas in T_1^* have at least $2^{\delta n}$ satisfying assignments. Notice that this can be rewritten as $\mathbf{Pr}[X(\phi) \geq 2^{\delta n} \mid \phi \in T_1] \geq 1 - 2^{-\delta n}$. Hence one can write (since $X(\phi) \geq 1$ for any $\phi \in T_1$):

$$\textstyle\sum_{\psi \in T_1} \frac{1}{X(\psi)} = \sum_{\psi \in T_1^*} \frac{1}{X(\psi)} + \sum_{\psi \in T_1 \setminus T_1^*} \frac{1}{X(\psi)} \leq \frac{|T_1^*|}{2^{\delta n}} + (|T_1| - |T_1^*|) \leq \frac{|T_1|}{2^{\delta n}} + \frac{|T_1|}{2^{\delta n}}.$$

Therefore $\mathbf{E}_{\phi \in SAT}[X(\phi)] \geq \frac{|T_1|}{\frac{2|T_1|}{2^{\delta n}}} = 2^{\delta n - 1}$.

Let $\mathcal{C}_{3-i} = \mathcal{C}_{3-i}(\psi)$ be the set of clauses in ψ containing exactly i positive (i.e. non negated) variables. We first estimate $|\mathcal{C}_i|$ under the assumption that $\psi \in T_1$. Notice that no formula in T_1 can contain a clause with only negated variables, therefore $|\mathcal{C}_3| = 0$. Furthermore, for formulas in T_1 with n variables and θn clauses, it is fairly easy to compute the asymptotic distribution of the formulas with $|\mathcal{C}_i| = n_{3-i}$ for each $i \in \{0, 1, 2\}$ (with $\theta n = n_1 + n_2 + n_3$):

$$\mathbf{Pr}[n_1, n_2, n_3] = \frac{\binom{\theta n}{n_1, n_2, n_3} \left(n\binom{n}{2}\right)^{n_1 + n_2} \binom{n}{3}^{n_3}}{\binom{2n}{3} - \binom{n}{3}}.$$

Using Stirling's approximation for the various factorials involved and setting $\gamma_i = n_i/n$ it is easy to prove that $\mathbf{Pr}[n_1, n_2, n_3]^{\frac{1}{n}}$ is asymptotic to $\frac{(6\theta/7)^\theta}{\gamma_1^{\gamma_1}\gamma_2^{\gamma_2} 2^{\gamma_1+\gamma_2}(6\gamma_3)^{\gamma_1}}$, assuming all n_i's tend to infinity. Such expression, considered as a function of γ_1, γ_2, and γ_3 reaches its maximum (equal to one) for $\gamma_1 = \gamma_2 = \frac{3\theta}{7}$ and $\gamma_3 = \frac{\theta}{7}$. Let $T_1^* \subseteq T_1$ be the set of all those formulas in T_1 with $\gamma_1 \leq 2.37$, $\gamma_2 \leq 2.37$, and $\gamma_3 \leq 0.87$ and satisfying $\gamma_1 + \gamma_2 + \gamma_3 = 5.08$. The above inequalities hold with probability at least $1 - e^{-0.01483n}$, implying that the size of this set is at least $(1 - \mathrm{e}^{-0.01483n})|T_1|$. It remains to prove that the formulas in T_1^* have at least $2^{0.02137n}$ satisfying truth assignments with sufficiently high probability. To this end, the authors introduce the notion of a *dispensable* variable x in a formula ψ w.r.t. a given satisfying assignment A. A variable x is called dispensable if the assignment A' resulting from A by changing the value assigned to the variable x also satisfies ψ. The authors prove that, with high probability, the formulas in T_1^* contain many dispensable variables. More precisely, let $D(\psi)$ be the set of dispensable variables in ψ with respect to the assignment that sets all variables to TRUE. Clearly $X(\psi) \geq 2^{|D(\psi)|}$. The paper is completed by the analysis of the size of the set of dispensable variables returned by the following "greedy" algorithm:

- Determine the set I_1 of the variables which are isolated. That is, for each variable $x \in I_1$ there is at least one clause in the set $\mathcal{C}_2$ of clauses which contains x. Select at random one variable from those clauses in $\mathcal{C}_1$ whose positive variables are both out from I_1. Let I_2 the set of pairwise distinct selected variables. Set $J_2 = I_1 \cup I_2$.

- Select at random one literal from each clause in $\mathcal{C}_0$ whose 3 positive literals are all out of J_2. Call I_3 the set of pairwise distinct selected literals. Set $J_3 = J_2 \cup I_3$.

The estimate on the size of J_3 is obtained by finding upper bounds on $|I_1|$ conditioned on γ_i taking up a particular range of values, on $|I_2|$ conditioned on $|I_1|$ and the values of γ_i, and on $|I_3|$ conditioned on $|I_2|$, $|I_1|$, and the values of γ_i (these two probabilities are both binomial where the number of trials are $\gamma_2 n$ and $\gamma_3 n$ respectively, the success probabilities can be computed quite easily and success means "not being in I_2" and "not being in I_3", respectively). This argument shows that J_3 cannot be too large and, therefore, the remaining variables (they are all dispensable) are a fixed proportion of V.

In order to estimate the number $n_1 = \gamma_1 n$, the authors resort to the *occupancy problem*. In this problem, we uniformly throw at random μn balls (μ is a constant) into n boxes and we ask for the distribution of the random variable Y that counts the number of *non-empty boxes*. Then for any $\epsilon > 0$, $r = r(\epsilon) = (1+\epsilon)(1 - e^{-\frac{\mu}{1+\epsilon}})$, $s = s(\epsilon) = 1 - r(\epsilon)$ the following is established:

$$\frac{1}{n} \log \mathbf{Pr}[Y \geq \lfloor r(\epsilon) n \rfloor] \sim \log \frac{(s+\epsilon)^{s+\epsilon}(1+\epsilon)^{\mu-1-\epsilon}}{s^s}.$$

4.2 Sharper estimation of occupancy probabilities: independent variables

Kamath *et al.* performed in [32] a similar investigation of the structure of the typical $\psi \in T_1$. A variable x *covers* a clause C if x occurs unnegated in C. A set of variables V covers a set of clauses F if every clause in F is covered by at least a variable V. In the formula (represented by the sequence of sets of literals forming individual clauses in it)

$\psi(x_1, x_2, x_3, x_4, x_5) =$
$\mathcal{C}_0$ $\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_3, x_4, x_5\},$
$\mathcal{C}_1$ $\{\overline{x}_2, x_3, x_4\}, \{x_1, x_4, \overline{x}_5\}, \{x_1, \overline{x}_2, x_5\}, \{x_1, \overline{x}_2, x_4\}, \{x_1, \overline{x}_3, x_5\}, \{x_3, x_4, \overline{x}_5\}, \{x_2, x_4, \overline{x}_5\},$
$\{x_2, \overline{x}_3, x_4\}, \{\overline{x}_3, x_4, x_5\}, \{x_1, x_3, \overline{x}_4\},$
$\mathcal{C}_2$ $\{x_2, \overline{x}_3, \overline{x}_5\}, \{\overline{x}_1, x_4, \overline{x}_5\}, \{\overline{x}_1, \overline{x}_3, x_5\}, \{\overline{x}_1, \overline{x}_2, x_5\}, \{\overline{x}_1, x_3, \overline{x}_4\}, \{\overline{x}_3, x_4, \overline{x}_5\}, \{\overline{x}_2, x_4, \overline{x}_5\},$
$\{\overline{x}_1, x_4, \overline{x}_5\}, \{x_1, \overline{x}_3, \overline{x}_5\},$
$\mathcal{C}_3$ $\{\overline{x}_1, \overline{x}_2, \overline{x}_3\}, \{\overline{x}_1, \overline{x}_4, \overline{x}_5\}, \{\overline{x}_2, \overline{x}_3, \overline{x}_4\}, \{\overline{x}_1, \overline{x}_2, \overline{x}_4\}, \{\overline{x}_1, \overline{x}_3, \overline{x}_5\}.$

The variable x_1 covers the clauses:

$$\{x_1, x_2, x_3\}, \{x_1, x_4, \overline{x}_5\}, \{x_1, \overline{x}_2, x_5\}, \{x_1, \overline{x}_2, x_4\}, \{x_1, \overline{x}_3, x_5\}, \{x_1, x_3, \overline{x}_4\}, \{x_1, \overline{x}_3, \overline{x}_5\}.$$

Figure 1 shows a formula with the set of clauses in $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}_2$, a set of variables X_2 covering $\mathcal{C}_2$, a second set of variables needed to cover the uncovered portions of $\mathcal{C}_0$ and $\mathcal{C}_1$ and the remaining *independent set* $\mathcal{I}$. Notice that by setting the variables in a cover to TRUE, the whole formula is satisfied and all variables in the independent set can be set arbitrarily. In other words, if ψ has a cover of size s then it has 2^{n-s} different satisfying assignments (at least).

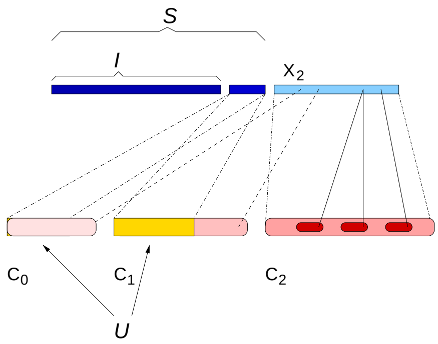

Figure 1: A set of clauses and the corresponding covers.

To estimate $\sum_{\psi_j \in T_1} \frac{1}{X(\psi_j)\cdot|T_1|}$ we partition T_1 in "slices" containing formulas with minimal cover of size s and use:

$$\sum_{\psi_j \in T_1} \frac{1}{X(\psi_j)\cdot|T_1|} \leq \sum_{\text{cover}(\psi_j)=s} \frac{\mathbf{Pr}[\text{cover}(\psi_j)=s]}{X(\psi_j)} \leq \sum_{\text{cover}(\psi_j)=s} 2^{s-n}\mathbf{Pr}[\text{cover}(\psi_j)=s].$$

The problem reduces to the estimation of $\mathbf{Pr}[\text{cover}(\psi_j) = s]$, as accurately as possible, using asymptotic expressions for binomial tails and occupancy probabilities.

- We first fix the size n_1 of $\mathcal{C}_2$ (these contain a single positive variable). The probability $\mathbf{Pr}[\alpha]$ that this number exceeds by more than α its mean can be shown to be small using binomial tails.
- Then we determine the set I_1 of the variables which are isolated (denoted by $\mathcal{X} \setminus \mathcal{S}$ in [32]). Conditioned on n_1 being within α of its mean, the probability $\mathbf{Pr}[\beta|\alpha]$ that $|I_1|$ is within β times its mean is estimated using the occupancy asymptotics (again we are throwing clauses "into" variables, the empty bins correspond to the variables in $V \setminus I_1$).
- The third step is to compute, conditioned on $|I_1|$ and n_1, the number of clauses in $\mathcal{C}_1 \cup \mathcal{C}_0$ that are *not* covered by I_1 (set $\mathcal{U}$). The probability that this number exceeds by more than γ its mean, $\mathbf{Pr}[\gamma|\alpha, \beta]$, is a binomial tail again.
- Finally we bound the size of the variables needed to cover $\mathcal{U}$.

The improvement in [32] comes from the estimation of the size of $\mathcal{U}$ and from adding to the cover only variables needed to cover $\mathcal{U}$, not the whole set $\mathcal{C}_1 \cup \mathcal{C}_0$ (as done in the other paper, in the second and third selection step).

5 Special classes of satisfying truth assignments

The next improvements resulted from a different kind of exploitation of rare formulas with many satisfying truth assignments.

Since the main disadvantage of Markov's inequality can be attributed to rare formulas having a large number of satisfying truth assignments, a plausible approach for improvement is to use in the inequality *not* the expected cardinality of the set of satisfying truth assignments of a random formula but the expected cardinality of a *smaller* set. Of course one needs to prove that the expectation of the new random set still bounds from above the probability that ϕ is satisfiable.

This idea was introduced by Kirousis, Kranakis, and Krizanc in [35] ("single flips") and then, independently, by Dubois and Boufkhad in [14]. In the remainder of this section we describe these approaches and several others derived from them and we report the resulting improvements on the estimation of r_3.

The following formula

$$\phi(x_1, x_2, x_3, x_4) = \{x_1, \overline{x}_2, x_4\}, \{\overline{x}_1, \overline{x}_2, \overline{x}_4\}, \{x_1, \overline{x}_3, \overline{x}_4\}, \{\overline{x}_1, x_2, \overline{x}_3\}$$

whose satisfying assignments are

	x_1	x_2	x_3	x_4
A_1	FALSE	FALSE	FALSE	FALSE
A_2	FALSE	FALSE	FALSE	TRUE
A_3	FALSE	FALSE	TRUE	FALSE
A_4	FALSE	TRUE	FALSE	TRUE
A_5	TRUE	FALSE	FALSE	FALSE
A_6	TRUE	FALSE	FALSE	TRUE
A_7	TRUE	TRUE	FALSE	FALSE
A_8	TRUE	TRUE	TRUE	FALSE

will be used occasionally to convey a better understanding of the various results.

5.1 Single flips

The strategy described here was introduced in [35]. In what follows it may be convenient to identify classes of truth assignments on n variables with sets of lexicographically ordered sequences over a two letter alphabet (say the number 0 and 1 with $0 < 1$).

Given a random formula ϕ, the set $\mathcal{A}_n^1$ is defined as the class of truth assignments A such that the following two conditions hold: (i) A satisfies ϕ, and (ii) any assignment obtained from A by changing exactly one FALSE value to TRUE does not satisfy ϕ. Such a change is called a *single flip*. The class $\mathcal{A}_n^1$ contains the elements of $\mathcal{A}_n$ that are *local maxima* with respect to single flips. In other words, a truth assignment belongs to $\mathcal{A}_n^1$ if it satisfies ϕ and if no other possible satisfying truth assignment can be obtained from it by changing a single FALSE value to TRUE (i.e. by performing all possible single flips in isolation).

In our example the first truth assignment in $\mathcal{A}_n$ does not belong to $\mathcal{A}_n^1$: if we change the value assigned to x_1 the resulting truth assignment is still in $\mathcal{A}_n$. However, the fourth truth assignment does belong to $\mathcal{A}_n^1$: changing the value assigned to x_1 produces the assignment $A' \notin \mathcal{A}_n$ which sets all variables except x_3 to TRUE. Similarly changing the value of x_3 leads to an assignment that does not satisfy ϕ. Therefore, since for all possible transformations which complement a FALSE value result in truth assignments which do not satisfy ϕ, the assignment A_4 is in $\mathcal{A}_n^1$. It can be easily checked that the set $\mathcal{A}_n^1$ for ϕ is formed by the assignments A_3, A_4, A_6, and A_8.

Since $\mathcal{A}_n^1 \subseteq \mathcal{A}_n$, it is true that $\mathbf{E}[|\mathcal{A}_n^1|] \leq \mathbf{E}[|\mathcal{A}_n|]$. Thus, to relax Markov's inequality we only need to establish that $\mathbf{E}[|\mathcal{A}_n^1|]$ still bounds from above $\mathbf{Pr}[\phi \text{ is satisfiable}]$. This can be easily seen as follows. Let I_ϕ be the random indicator for the property "ϕ is satisfiable". Clearly $I_\phi \leq |\mathcal{A}_n^1|$. The sought inequality follows immediately from this remark and the fact that we can write

$$\mathbf{Pr}[\phi \text{ is satisfiable }] = \textstyle\sum_\phi I_\phi \mathbf{Pr}[\phi].$$

Thus exploit this technique one needs to prove a bound, asymptotically smaller than $2^n(7/8)^{rn}$ on $\mathbf{E}[|\mathcal{A}_n^1|]$.

Using a correlation inequality for the computation of the probability that a single flip results in an assignment that does not satisfy ϕ, it can be proved that, in the random formula model G_{mm}, the expected size of class $\mathcal{A}_n^1$ is at most $(7/8)^{rn}(2 - e^{-3r/7} + o(1))^n$, Therefore, the unique positive solution of the equation $(7/8)^r(2-e^{-3r/7}) = 1$ gives an upper bound for the satisfiability threshold critical value r_3. This solution is approximately 4.667.

On the other hand, the application of Markov's inequality in G_p leads to the solution of the equation $e^{-r/8}(2 - e^{-3r/7}) = 1$. This expression, unfortunately, gives an upper bound equal to 5.07, which is worse than the bound given within the context of G_{mm}. However, it avoids the computation of probabilities of conjunctions of *dependent* events.

5.2 The set of negatively prime solutions (NPS)

Independently from Kirousis *et al.*, Dubois and Boufkhad introduced a class of satisfying truth assignments that they called *Negatively Prime Solutions* (NPS) [14]. This class actually coincides with the class $\mathcal{A}_n^1$ described in Subsection 5.1.

Dubois and Boufkhad proved the following exact expression for the expected cardinality of NPS.

$$\mathbf{E}[|\text{NPS}|] = \sum_{i=0}^{n}\sum_{j=i}^{m} 2^{-km}2^i \binom{n}{i}\binom{m}{j} S_2(j,i)i! \left(\frac{k}{n}\right)^j (2^k - 1 - k)^{m-j} \quad (3)$$

where $S_2(j,i)$ are the *Stirling numbers of the second kind* that count the number of ways of partitioning a j element set into i *non-empty* subsets. Then the authors performed a series of asymptotic manipulations and arrived at a closed form upper bound for (3), showing that it converges to 0 for values of the clause to variable ratio greater than 4.642.

5.3 Restricting further the class of satisfying truth assignments: double flips

In [37], Kirousis *et al.* define as a *double flip* the change of exactly two variables x_i and x_j, with $i < j$, where x_i is changed from FALSE to TRUE and x_j from TRUE to FALSE. Notice that the restriction $i < j$ implies that a double flip always leads to a lexicographically *larger* assignment. Let A^{df} denote the truth assignment that results from A after the application of the double flip *df*. Let $\mathcal{A}_n^{2\sharp}$ be the set of truth assignments A that have the following three properties: (i) $A \models \phi$, (ii) for all possible single flips *sf* of A, it holds $A^{sf} \not\models \phi$, and (iii) for all possible double flips *df* of A, it holds $A^{df} \not\models \phi$.

It can be proved (see [37]), that the following inequality holds:

$$\mathbf{Pr}[\phi \text{ is satisfiable}] \leq \mathbf{E}[|\mathcal{A}_n^{2\sharp}|] = (7/8)^{rn}\textstyle\sum_{A\in S}\mathbf{Pr}[A \in \mathcal{A}_n^1 \mid A \models \phi] \cdot \mathbf{Pr}[A \in \mathcal{A}_n^{2\sharp} \mid A \in \mathcal{A}_n^1]. \quad (4)$$

Therefore, to get an upper bound on r_3 it suffices to find the smallest possible value for the clause to variable ratio for which the right-hand side of the above inequality tends to 0. In what follows, $sf(A)$ denotes the total number of single flips of A (this is just the number of variables set to FALSE by A) and $df(A)$ is the number of double flips.

It can be proved that $\mathbf{Pr}[A \in \mathcal{A}_n^1 \mid A \models \phi]$ can be bounded from above by an expression of the form $X^{sf(A)}$ where X depends only on the clause to variable ratio r (this is exactly the expression used in the derivation of the improved upper bound in Subsection 5.1). The hard part is, of course, the computation of the second probability, involving the realization of the double flips events conditional upon the realization of the single flips events. As it turns out, the computation of this probability in the model G_{mm}would involve dependencies among events of a very complicated nature. On the other hand, in model G_p at least there would be no dependencies ensuing from the fundamental requirement of G_{mm} that the size of the formula is fixed. Thus, what remains is the dependencies arising from the fact that some of the double flip events involve double flips that share a particular FALSE variable. The computation of an upper bound to this probability was made possible with the use of a version of Suen's correlation inequality [45] proved by Janson in [30]. This bound has the form $Y^{df(A)}$ with Y dependent on n and r. The reader may consult [37] for the derivation of this bound. Inequality (4) may then be rewritten as follows:

$$\mathbf{Pr}[\phi \text{ is satisfiable}] \leq 3m^{1/2}(7/8)^{rn} \sum_{A \in \mathcal{A}_n} X^{sf(A)} Y^{df(A)}, \tag{5}$$

where the polynomial factor $3m^{1/2}$ is a byproduct of the change of the model from $G_{mm}$to G_p (see [6]). To complete the derivation of the improved bound the authors noticed the following combinatorial identity (which can be proved by induction on n)

$$\sum_{A \in \mathcal{A}_n} X^{sf(A)} Y^{df(A)} = \sum_{k=0}^{n} \binom{n}{k}_Y X^k,$$

with $\binom{n}{k}_q$ the *q-binomial* or *Gaussian* coefficients (see [25]), for $0 \leq k \leq n$ and $q \neq 1$, and then, using this connection with q-binomial coefficients, the inequality

$$\sum_{A \in \mathcal{A}_n} X^{sf(A)} Y^{df(A)} \leq \prod_{k=0}^{n-1} (1 + XY^{k/2}), \qquad 0 \leq X^2 \leq Y \leq 1. \tag{6}$$

Thus, Equation (5) becomes

$$\mathbf{Pr}[\phi \text{ is satisfiable}] \leq 3m^{1/2}(7/8)^{rn} \prod_{k=0}^{n-1} (1 + XY^{k/2}). \tag{7}$$

and the product on the right hand side can be estimated using a connection between the above expression, *hypergeometric series* (see also [25]) and an inequality derived in [36]. This eventually leads to the proof of the inequality $r_3 < 4.602$.

5.4 Occupancy bounds and q-binomial coefficients

As we saw in the previous subsection, a key step to the improvement of the upper bound to 4.602 was the derivation of an upper bound to the sum that appears in (5) through its connection with the q-binomial coefficients. As it turned out, this bound could be further improved. In [31], Janson, Stamatiou, and Vamvakari gave two approaches for improving the upper bound of Inequality (6). Both approaches result in the same upper bound, namely 4.596. However, these approaches are interesting in their own right. The first approach links the problem of determining upper bounds to the satisfiability threshold with the study of *Ising spin systems* in *Statistical Mechanics* while the second approach links this problem with the branch of mathematics dealing with q-hypergeometric series and their generating functions.

More specifically, in the first approach the sum in (5) is written as

$$\sum_{\varepsilon_1,\ldots,\varepsilon_n\in\{0,1\}} \exp\left(a\sum_{i=1}^{n}\varepsilon_i + \frac{b}{n}\sum\sum_{1\le i<j\le n}\varepsilon_i(1-\varepsilon_j)\right), \tag{8}$$

with the sum ranging over all the 2^n sequences $\varepsilon_1,\ldots,\varepsilon_n$ of 0's and 1's, where each of them codes a truth assignment A but with 0 and 1 representing TRUE and FALSE respectively. This sum is indeed equal to the sum of (5) with $a=\ln(X)$ and $b=n\ln(Y)$.

This form of the sum enabled the application an optimization technique common in statistical physics that results in an asymptotic expression for the sum. In statistical physics, this particular form of the sum defines the *partition function* for a system with n spin sites, each having a spin value $\varepsilon_i\in\{0,1\}$ and with an energy function equal to $H=-a\sum_{i=1}^{n}\varepsilon_i-\frac{b}{n}\sum\sum_{1\le i<j\le n}\varepsilon_i(1-\varepsilon_j)$. In this expression, the first term corresponds to an external field that acts on all the spins of the system while the second sum to an interaction acting between pairs of sites of arbitrary distance with the left site having spin 1 and the right site having spin 0. The energy function can easily be rewritten into a more conventional form. In fact, $H=\sum_{i=1}^{n}(-a-b+b\frac{i}{n})\varepsilon_i+\frac{b}{n}\sum\sum_{i<j}\varepsilon_i\varepsilon_j$, or, substituting $\varepsilon_i=(1+s_i)/2$ to have more traditional (and symmetrical) spins with values ±1, $H=-\frac{a}{2}n-\frac{b}{8}(n-1)+\sum_{i=1}^{n}(-\frac{a}{2}+b\frac{i-(n+1)/2}{2n})s_i+\frac{b}{4n}\sum\sum_{i<j}s_is_j$, which finally demonstrates that in Physics the system can be interpreted as a *mean-field Ising model with an inhomogeneous (linear) external field*.

In the second approach, a sharp upper bound for the Rogers–Szegö polynomials is derived using their *Eulerian* generating function and a technique that is stated in Lemma 8.1 in [43]. Using this technique, for any t, $0<t<\min(1,1/x)$, the following upper bound is obtained (see [31]):

$$F_{n,q}(x)\le t^{-n}\exp\left[-\frac{1}{\log q}(\mathrm{Li}_2(tx)+\mathrm{Li}_2(t)+\mathrm{Li}_2(q^n)-\mathrm{Li}_2(q))\right]\frac{1}{(1-t)(1-tx)}$$

where $\mathrm{Li}_2(y)=\mathrm{dilog}(1-y)=\mathrm{Polylog}(2,y)=\sum_{k\ge1}\frac{y^k}{k^2}$ is the *dilogarithm* function. By optimizing for t this upper bound and plugging it into (5), we obtain the upper bound 4.596.

5.5 Balls and bins

The calculations described in the previous sections have their limits. The single flips idea, although it led to a good improvement of the upper bound, only takes into account a very limited range of locality around a given satisfying assignment. The results in [37] exploit wider locality ranges, but because of the complexity of the resulting numerical expressions, the authors were forced to use weak bounds on $\mathbf{Pr}[A \in \mathcal{A}_n^1 \mid A \models \phi]$ and the overall $\mathbf{E}[|\mathcal{A}_n^{2\sharp}|]$. Finally, the approach in [31] reached the limits of what can be exploited from the upper bound shown in (5). Thus, in order to obtain further improvements, one should step backwards in the derivation of (5) and attempt to effect improvements on the probabilities involved.

In [34], Kaporis *et al.* achieved an improved bound of 4.571 using sharp estimates for certain probabilities related to the classical occupancy problem. For a given satisfying assignment A, the probability that no single flip satisfies ϕ is best computed (up to polynomial factors) by noticing the following "structural" condition imposed on ϕ by the given event: for each variable x set to false by A the formula must contain a *critical* clause $\{\overline{x}, \ell_1, \ell_2\}$ with $A(\ell_1) = A(\ell_2) = \text{FALSE}$.

If the assignment A sets k variables to false, the event "$\forall sf\, A^{sf} \not\models \phi$" conditioned on "$A \models \phi$" occurs when: (i) some $l \geq k$ clauses out of $m = rn$ are critical (the remaining being consistent with A), and (ii) the l clauses can be seen as a sequence of balls that are dropped into k distinct bins (corresponding to different single flips) in a way that leaves no bin completely empty.

Asymptotic estimates on the resulting occupancy probabilities [32] as well as a change of models from G_{mm} to G_m in order to be able to model our problem in the balls and bins framework lead to a sharper bound of $\mathbf{Pr}[A \in \mathcal{A}_n^1 \mid A \models \phi]$ (Zito [47] performed a similar analysis using coupon collector probabilities instead, deriving a bound of about 4.58). The analysis in [34] improves on the previous results also for another reason. The overall bound on $\mathbf{E}[|\mathcal{A}_n^{2\sharp}|]$ is tightened by means of a more direct estimation of the q-binomial coefficient involved. Using simple generating function inequalities (and elementary calculus) it is possible to bound the term $\binom{n}{k}_Y$ directly and avoid the use of the not entirely immediate relationship between $\sum_{k=0}^{n} \binom{n}{k}_Y X^k$ and the *Rogers-Szegö* polynomials.

Finally to obtain their results the authors had to establish a relationship between the probabilities of the event "$A \in \mathcal{A}_n^{2\sharp}$" conditioned to $\in \mathcal{A}_n^1$ in the models G_{mm} and G_p. Using results described in Bollobás[6, Chap. II], it is easy to prove the desired relationship for *unconditional* events when the average length of a random formula constructed in the model G_p equals m. However, the conditioning here may bias the expected length of the formula to higher values. The authors show how to adjust p appropriately as to obtain the above inequality.

6 Typical formulas

All the techniques outlined above relied on what could be termed as a *semantic* approach, as they work by placing restrictions on the *assignments* (*semantics*, in some sense) satisfying a formula and for which the expectation which is required by the first moment method is computed. As remarked in Subsection 5.2, the application of this technique can provably not result in any upper bound improvement, at least for the simplest such case of assignment restrictions described in Subsections 5.1, 5.2, and 5.3. Dubois, Boufkhad, and Mandler considered in [15] random formulas with the special characteristic that the numbers of appearances of their literals fall within certain ranges that are "typical" for randomly generated formulas. In this way, they disallowed the rare formulas that seem to be the reason why Markov's inequality does not give an upper bound close to the experimentally determined value. By computing the expected number of negative prime solutions for these formulas only, they achieved an upper bound improvement to 4.506. In contrast to other methods relying on restricting the set of truth assignments taking part in the application of the first moment method, this approach can be characterized as *syntactic* as it focuses on restricting the *form* or *syntax* of the set of formulas participating in the first moment method computations. However, the method also restricts the possible truth assignments using the restricted sets defined in Sections 5.1 and 5.2. Without going into details, Dubois, Boufkhad, and Mandler give an expression for the expected number of negative prime solutions for these formulas, obtaining as an upper bound to the threshold the value 4.506.

References

[1] D. Achlioptas, "Setting two variables at a time yields a new lower bound for random 3-SAT," *Proc. 32nd Annual ACM Symposium on Theory of Computing (STOC '00)*, pp. 28–37, 2000.

[2] D. Achlioptas, "Lower bounds for random 3-SAT via differential equations," *Theoretical Computer Science* **265**, pp. 159–185, 2001.

[3] D. Achlioptas and C. Moore, "The asymptotic order of the random k-SAT threshold," to appear in *Proc. 43rd Annual Symposium on Foundations of Computer Science*, 2002.

[4] D. Achlioptas and G.B Sorkin, "Optimal myopic algorithms for random 3-SAT," *Proc. 41st Annual Symposium on Foundations of Computer Science*, pp. 590–600, 2000.

[5] D. Aldous, "The harmonic mean formula for probabilities of unions: applications to sparse random graphs," *Discrete Mathematics* **76**, pp. 167–176, 1989.

[6] B. Bollobás, *Random Graphs*, Second Edition, Cambridge University Press, 2001.

[7] A.Z. Broder, A. Frieze, and E. Upfal, "On the satisfiability and maximum satisfiability of random 3-CNF formulas," in *Proc. 4th Symposium on Discrete Algorithms (SODA)*, pp. 322–330, 1993.

[8] M. T. Chao and J. Franco, "Probabilistic analysis of two heuristics for the 3-satisfiability problem," *SIAM J. Comput.* **15**, pp. 1106–1118, 1986.

[9] M.T. Chao and J. Franco, "Probabilistic analysis of a generalization of the unit-clause litteral selection heuristics for the k-satisfiability problem," *Inform. Sci.* **51**, pp. 289–314, 1990.

[10] P. Cheeseman, B. Kanefsky, and W.M. Taylor, "Where the really hard problems are," *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 331–337, 1991.

[11] V. Chvátal and B. Reed, "Mick gets some (the odds are on his side)," *Proc. 33rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 620–627, 1992.

[12] V. Chvátal and E. Szemerédi, "Many hard examples for resolution," *Journal of the Association for Computing Machinery* **35**, pp. 759–768, 1988.

[13] J.M. Crawford and L.D. Auton, "Experimental results on the crossover point in random 3-SAT," *Artificial Intelligence* **81** (1-2), pp. 31–57, 1996.

[14] O. Dubois and Y. Boufkhad, "A general upper bound for the satisfiability threshold of random r-SAT," *J. of Algorithms* 24, pp. 395–420, 1997. Originally appeared in: *Proc. RR-2992 of the INRIA Algorithms Seminar*, B. Salvy (ed.), pp. 137–139, 1995–1996 and as a short abstract titled "Restrictive structures for characterizing the satisfiability," in *Proc. 4th International Symposium on Artificial Intelligence and Mathematics*, Fort Lauderdale, Florida, pp. 176, 1996.

[15] O. Dubois, Y. Boufkhad, and J. Mandler, "Typical random 3-SAT formulae and the satisfiability threshold," in: *Proc. 11th Symposium on Discrete Algorithms (SODA)*, pp. 126–127, 2000.

[16] O. Dubois, "Upper bounds on the satisfiability threshold," *Theoretical Computer Science* **265**, pp. 187–197, 2001.

[17] M. El Maftouhi and W. Fernandez de la Vega, "On random 3-SAT," *Combinatorics, Probability and Computing* **4**, pp. 190–195, 1995.

[18] W. Fernandez de la Vega, "On random 2-SAT," unpublished manuscript, 1992.

[19] J. Franco, "Probabilistic analysis of the pure literal heuristic for the satisfiability problem," *Ann. Oper. Res.* **1**, pp. 273–289, 1984.

[20] J. Franco, "Results related to threshold phenomena research in satisfiability: Lower Bounds," *Theoretical Computer Science* **265**, pp. 147–157, 2000.

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

[22] E. Friedgut, appendix by J. Bourgain, "Sharp thresholds of graph properties and the k-sat problem," *J. Amer. Math. Soc.* **12**, pp 1017–1054, 1999.

[23] A. Frieze and S. Suen, "Analysis of two simple heuristics on a random instance of k-SAT," *Journal of Algorithms* **20**, pp. 312–355, 1996.

[24] A. Frieze and N. Wormald, "k-SAT: a tight threshold for moderately growing k," in *Proc. 5th International Symposium on the Theory and Applications of Satisfiability Testing*, pp. 1–6, 2002.

[25] G. Gasper and M. Rahman, *Basic Hypergeometric Series*, *Encyclopedia of Mathematics and its Applications* **35**, Cambridge University Press, Cambridge, 1990.

[26] A. Goerdt, "A threshold for unsatisfiability," in *Proc. of the 17th Symposium on Mathematical Foundations of Computer Science (MFCS)*, (I.M. Havel and V. Koubek Eds), pp. 264–274, 1992.

[27] A. Goerdt, "A threshold for unsatisfiability," *J. Comput. System Sci.* **53**, pp. 469–486, 1996.

[28] A. Goldberg, P.W. Purdom, and C.A. Brown, "Average time analyses of simplified Davis-Putnam procedures," *Information Processing Letters* **15**, pp. 72–75, 1982.

[29] G. Istrate, "Phase transitions and *all that*," manuscript 2002 (www.satlive.org/hits.jsp?id=262).

[30] S. Janson, "New versions of Suen's correlation inequality," *Random Structures and Algorithms* **13**, 467–483, 1998.

[31] S. Janson, Y.C. Stamatiou, and M. Vamvakari, "Bounding the unsatisfiability threshold of random 3-SAT," *Random Structures and Algorithms* **17**, pp. 103–116, 2000.

[32] A. Kamath, R. Motwani, K. Palem, and P. Spirakis, "Tail bounds for occupancy and the satisfiability threshold conjecture," *Random Structures and Algorithms* **7**, pp. 59–80, 1995.

[33] A.C. Kaporis, L.M. Kirousis, and E.G. Lalas, "The probabilistic analysis of a greedy satisfiability algorithm," *Proc. of 10th Annual European Symposium on Algorithms (ESA' 02), Track A,*, pp. 574–584, 2002.

[34] A.C. Kaporis, L.M. Kirousis, Y.C. Stamatiou, M. Vamvakari, and M. Zito, "The unsatisfiability threshold revisited," to appear in *Discrete Mathematics.*

[35] L.M. Kirousis, E. Kranakis, and D. Krizanc, "A better upper bound for the unsatisfiability threshold," in: D. Du, J. Gu, and P.M. Pardalos (eds.), *Satisfiability Problem: Theory and Applications, DIMACS Workshop, March 11–13, 1996,* DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 35, pp. 643–648, 1997. Originally appeared as Technical Report TR-95-26, Carleton University, School of Computer Science, Canada, 1995.

[36] L. Kirousis, E. Kranakis, and D. Krizanc, "An upper bound for a basic hypergeometric series," *Tech. Rep.* TR-96-07, Carleton University, School of Computer Science, Canada, 1996.

[37] L.M. Kirousis, E. Kranakis, D. Krizanc, and Y.C. Stamatiou, "Approximating the unsatisfiability threshold of random formulas," *Random Structures and Algorithms* **12**, pp. 253–269, 1998.

[38] L.M. Kirousis, Y.C. Stamatiou, and M. Vamvakari, "Upper bounds and asymptotics for the q-binomial coefficients," *Studies in Applied Mathematics* **107**, pp. 43–62, 2001.

[39] Mézard and R. Zecchina, "The random K-satisfiability problem: from an analytic solution to an efficient algorithm," *Phys. Rev. E*, to appear (www.ictp.trieste.it/ zecchina).

[40] D. Mitchell, B. Selman, and H. Levesque, "Hard and easy distribution of SAT problems," in *Proc. of the 10th National Conference on Artificial Intelligence (AAAI '92)*, pp. 459–465, 1992.

[41] R. Monasson and R. Zecchina, "Statistical mechanics of the random k-Sat problem," *Phys. Rev. E 56*, pp. 1357–1361, 1997.

[42] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, "Computational complexity from "characteristic" phase transitions," *Nature* **400**, pp. 133–137, 1999.

[43] A.M. Odlyzko, *Handbook of Combinatorics*, Chapter 22 *Asymptotic Enumeration Methods.* Eds. R.L. Graham, M. Grötschel, and L. Lovász, pp. 1063–1229. Elsevier Science Publishers, Amsterdam, 1995.

[44] J.C. Simon, J. Carlier, O. Dubois, and O. Moulines, "Étude statistique de l' existence de solutions de problèmes SAT, application aux systèmes-experts," *C.R. Acad. Sci. Paris. Sér. I Math.* 302, pp. 283–286, 1986.

[45] W.C. Suen, "A correlation inequality and a Poisson limit theorem for nonovelapping balanced subgraphs of a random graph," *Random Structures and Algorithms* **1**, pp. 231–242, 1990.

[46] N.C. Wormald, "The differential equation method for random graph processes and greedy algorithms," *Ann. Appl. Probab.* **5**, 1217–1235, 1995. Also in: M. Karonski and H.J. Proemel, eds. *Lectures on Approximation and Randomized Algorithms*, PWN, Warsaw, pp. 73–155, 1999.

[47] M. Zito, Randomised Techniques in Combinatorial Algorithmics, PhD Thesis, Department of Computer Science, University of Warwick, November 1999.