

AN UPPER BOUND ON THE SPACE COMPLEXITY OF RANDOM FORMULAE IN RESOLUTION*

MICHELE ZITO¹

Abstract. We prove that, with high probability, the space complexity of refuting a random unsatisfiable Boolean formula in k -CNF on n variables and $m = \Delta n$ clauses is $O\left(n \cdot \Delta^{-\frac{1}{k-2}}\right)$.

Mathematics Subject Classification. 68Q25, 03B05, 03F20.

1. INTRODUCTION

The importance of studying the complexity of (propositional) proof systems comes from its close relationship with long-standing open problems in Complexity Theory such as $NP =? Co-NP$ [8]. The complexity measure related to the classical notion of *time* is the *size* of a proof, viz. the number of *lines* used in the proof.

Recently Esteban and Torán [11] suggested a measure for the space complexity of refusing an unsatisfiable formula in a proof system called resolution (subsequent work [2] extended this notion to other proof systems). Although several results [2, 11, 18] are, by now, known on the space complexity of various classes of formulae, a precise quantitative analysis of the space needed to prove the unsatisfiability of random formulae has remained, until recently, somewhat elusive.

As a step toward the solution of this problem, we point out that a combination of a variant of the classical Davis and Putnam [10] algorithm and a linear time

Keywords and phrases: Random formulae, space complexity, satisfiability threshold.

* Partially supported by British Council grant ARC 4079. Part of this work was carried out while the author was visiting the Abteilung Theoretische Informatik, Fakultät für Informatik, University of Ulm.

¹ Department of Computer Science, University of Liverpool, Chadwick Building, Peach Street, Liverpool L69 7ZF, UK; e-mail: M.Zito@csc.liv.ac.uk

algorithm for 2-SAT outputs resolution refutations of any unsatisfiable random formula within the space bounds stated in the following theorem:

Theorem 1.1. *For each $k \geq 3$, let ϕ be an unsatisfiable random k -CNF formula on n variables and $m = \Delta n$ clauses. There is an $a = a(k) > 0$ and a $\Delta_0 = \Delta_0(k)$ such that, with probability approaching one as n goes to infinity, the space complexity of ϕ is at most $n \cdot \Delta^{-\frac{1}{k-2}} + O(1)$ for $\Delta \geq \Delta_0$.*

The remainder of this paper is organised as follows. In Section 2 we introduce all relevant notations and technical results; in Section 3 we describe the class of refutations that will be the object of our analysis; in Section 4 we give full details of the proof of Theorem 1.1; Section 5 is devoted to final remarks and open problems.

2. PRELIMINARIES

Let a finite set of variables $\mathcal{X} = \{x_1, \dots, x_n\}$ be given. A *literal* is either x^0 or x^1 for any $x \in \mathcal{X}$, but we will often follow the common practice and denote x^0 (resp. x^1) by $\neg x$ (x). We identify *clauses* with sets of literals, but we will often abuse the notation and write $x \in C$ to denote the fact that the variable x occurs in the clause C as either x^0 or x^1 . A *formula* is a (n ordered) sequence of clauses $\phi = (C_1, \dots, C_m)$. The *size* of a formula, denoted $|\phi|$, is the number of clauses it contains. A formula is in *k -conjunctive normal form* (or *k -CNF*) if $|C_i| \leq k$ for all $i \in \{1, \dots, m\}$. In all the subsequent treatment Δ will denote the *clause density* m/n of the given formula.

Let $\mathcal{C}^{k,n}$ denote the set of all clauses with exactly k literals of k distinct variables defined from \mathcal{X} . A *random formula* is obtained by selecting uniformly at random, independently and with replacement m clauses from $\mathcal{C}^{k,n}$. Let $\mathcal{F}_m^{k,n}$ denote the resulting probability space on the set of all k -CNF formulae over n variables and m clauses. We will write $\phi \sim \mathcal{F}_m^{k,n}$ to signify that ϕ is obtained by the process outlined above. In all the subsequent treatment we say that an event \mathcal{E} , depending on a parameter n , holds with high probability (w.h.p.) if it holds with probability approaching one as n tends to infinity.

A *truth-assignment* is a mapping α that assigns “false” or “true” (usually denoted by 0 or 1) to each variable in its domain $\text{Dom}(\alpha)$. We write $|\alpha|$ for $|\text{Dom}(\alpha)|$. Given a clause C , a variable x , and a value $\nu \in \{0, 1\}$, the *restriction of C to $x = \nu$* , $C|_{x=\nu}$, is C if $x \notin C$, is one if $x^\nu \in C$ and it is $C \setminus \{x^{1-\nu}\}$ otherwise. If ϕ is a formula, then $\phi|_{x=\nu}$ is the sequence $(C_1|_{x=\nu}, \dots, C_{m'}|_{x=\nu})$ for all $C \in \phi$ such that $C|_{x=\nu} \neq 1$. If α is a truth-assignment with domain $\text{Dom}(\alpha) = \{x_{j_1}, \dots, x_{j_i}\}$ then $C|_\alpha$ denotes the clause

$$\left(\dots \left(C|_{x_{j_1}=\alpha(x_{j_1})} \right) |_{x_{j_2}=\alpha(x_{j_2})} \dots \right).$$

The meaning of $\phi|_\alpha$ is defined similarly. For $i \in \{0, \dots, k\}$, let $C_i(\phi, \alpha)$ denote the set of clauses of size i in $\phi|_\alpha$. We say that a formula ϕ is *true (false) under*

the assignment α if $\phi|_\alpha$ is empty ($\{\} \in \phi|_\alpha$). A formula is *satisfiable* if there exists a truth-assignment α (also known as *satisfying* assignment) such that ϕ is true under α .

A (*resolution*) *refutation* of a formula $\phi = (C_1, \dots, C_m)$ is a sequence of clauses $\pi = (D_1, \dots, D_t)$ such that $D_t = \{\}$ and for all $i \in \{1, \dots, t-1\}$ either $D_i = C_j$ for some $j \in \{1, \dots, m\}$ or D_i is obtained from D_j and D_k (with $j, k < i$) by the *resolution rule*:

$$\{x, l_1, l_2, \dots\}, \{\neg x, t_1, t_2, \dots\} \rightarrow \{l_1, l_2, \dots, t_1, t_2, \dots\}.$$

The two clauses to the left of “ \rightarrow ” are called *premises*, the clause to the right is called *resolvent*. The *size* of refutation π , $|\pi|$ is t . Clearly, ϕ is unsatisfiable if and only if there exists a refutation of ϕ .

2.1. SPACE COMPLEXITY OF DERIVATIONS

Following [11], a k -CNF formula ϕ has a refutation bounded by space s if there exists a sequence of formulae ϕ_1, \dots, ϕ_t with

1. $\phi_1 \subseteq \phi$;
2. $|\phi_i| \leq s$, for all $i \in \{1, \dots, t\}$;
3. ϕ_{i+1} is obtained from ϕ_i by deleting (if wished) some clauses, adding the resolvent of two clauses in ϕ_i , and adding (if wished) some clauses of ϕ ;
4. $\{\} \in \phi_t$.

The *space complexity* of ϕ , denoted by $\text{space}(\phi)$ is the minimum s such that there exists a refutation of ϕ bounded by space s .

Each refutation π of a k -CNF ϕ can be represented as a directed acyclic graph (dag) $G_{\phi, \pi}$: clauses in π correspond to nodes in $G_{\phi, \pi}$, with the clauses of ϕ associated with $G_{\phi, \pi}$'s source nodes, $\{\}$ associated with the (only) sink of $G_{\phi, \pi}$, and each application of the resolution rule corresponding to an internal node of $G_{\phi, \pi}$ associated with some clause D with two incoming edges leaving the nodes associated with $D_1 \cup \{x\}$ and $D_2 \cup \{\neg x\}$ respectively, with $D = D_1 \cup D_2$. A refutation is *tree-like* if its underlying dag is a tree. Unless ambiguity arises, from now on π will refer to either a refutation of ϕ or its corresponding dag. There is a nice relationship between the space complexity of ϕ and the number of pebbles needed in a particular pebbling game \mathcal{G} played on $G_{\phi, \pi}$.

Pebbling Game \mathcal{G} . *Given a connected dag with one sink the aim of the game is to put a pebble on the sink of the graph (the only node with no outgoing edge) according to the following rules:*

1. *a pebble can be placed on any initial node (i.e. a node with no predecessor);*
2. *a pebble can be removed from any node at any time;*
3. *a pebble can be placed on any internal node provided there is a pebble on all its parents.*

The following lemma is an immediate consequence of the definitions:

Lemma 2.1 [11]. *For any formula ϕ , $\text{space}(\phi)$ coincides with the minimum number of pebbles needed to win \mathcal{G} on any graph $G_{\phi, \pi}$, where π is a refutation of ϕ .*

Using this result it is possible to analyse the space needed for refuting a formula through techniques used for bounding the number of pebbles used/needed to play \mathcal{G} . The following result is a consequence of Lemma 2.1.

Theorem 2.2 [11]. *If ϕ has a tree-like refutation of size S , then $\text{space}(\phi) \leq \lceil \log S \rceil + 1$.*

From this and the fact that any unsatisfiable formula has a tree-like refutation of size at most 2^n , it follows immediately that $\text{space}(\phi) \leq n + 1$ for any formula ϕ over n variables.

Although refutations of unsatisfiable k -CNF formulae for $k \geq 3$ may require non constant space, unsatisfiable 2-CNF formulae can always be refuted in constant space.

Theorem 2.3. $\text{space}(\phi) = O(1)$, for any 2-CNF formula ϕ .

2.2. A TECHNICAL RESULT

A fundamental conjecture about $\mathcal{F}_m^{k,n}$ states that there is a θ_k independent of n , the *unsatisfiability threshold*, such that a $\phi \sim \mathcal{F}_m^{k,n}$ is almost certainly satisfiable (resp. unsatisfiable) if $m/n < \theta_k$ ($m/n > \theta_k$). The value of such a threshold has been determined [12] for $k = 2$ and $\theta_2 = 1$ but only upper and lower bounds are known for $k \geq 3$ (see [1] and [14] for the currently available best bounds for $k = 3$). The following result, which implies $\theta_2 < 1 + \epsilon$ for any $\epsilon > 0$, will be used in the proof of the Theorem 1.1.

Lemma 2.4. *Let $\phi \sim \mathcal{F}_{cn}^{2,n}$. The probability that ϕ be satisfiable is at most $2^n(3/4)^{cn}$.*

Proof. The expression $2^n(3/4)^{cn}$ is exactly the expected number of satisfying assignments of $\phi \sim \mathcal{F}_{cn}^{2,n}$. The result follows from the Markov inequality. \square

2.3. POLYNOMIAL INEQUALITIES

The following result will be used in the proof of Theorem 1.1.

Lemma 2.5. *Let a, Δ be positive real numbers, let p be a positive integer larger than one and $q \in \{1, 2\}$. There exist two sequences of real numbers $\Delta_{p,q} = \Delta_{p,q}(a)$, and $u_{p,q} = u_{p,q}(a, \Delta)$, such that the inequality $x^p - a\Delta(x-1)^q < 0$ is satisfied if $\Delta > \Delta_{p,q}$ and $2 \leq x < u_{p,q}$.*

Proof. We will sketch the proof for $q = 2$ (the proof for $q = 1$ follows the same lines, but most of the algebraic expressions involved are simpler). For $p = 2$, the given inequality is equivalent to

$$\left(1 - \frac{1}{a\Delta}\right)x^2 - 2x + 1 > 0$$

which is satisfied for $x > \left(1 - (a\Delta)^{-\frac{1}{2}}\right)^{-1}$ if $a\Delta > 1$. Let $\Delta_{2,2} = \frac{1}{a}$, and $u_{2,2} = +\infty$.

For each $p \geq 3$ it is intuitively obvious that, provided $a\Delta$ is sufficiently large, it is possible to find an open interval $I_p \subseteq (1, +\infty)$ such that the given inequality is satisfied for $x \in I_p$. The least upper bound of the points of such an interval can be found by solving the associated equation

$$x^p - a\Delta(x - 1)^2 = 0.$$

For $p = 3$, such equation (see [20]) has the following three real solutions if $\Delta > \frac{27}{4a}$

$$\begin{aligned} x_1 &= -\frac{2}{3}\sqrt{a^2\Delta^2 - 6a\Delta} \cos \frac{\tau}{3} + \frac{a\Delta}{3} \\ x_2 &= -\frac{2}{3}\sqrt{a^2\Delta^2 - 6a\Delta} \cos \frac{\tau - 2\pi}{3} + \frac{a\Delta}{3} \\ x_3 &= -\frac{2}{3}\sqrt{a^2\Delta^2 - 6a\Delta} \cos \frac{\tau + 2\pi}{3} + \frac{a\Delta}{3} \end{aligned}$$

where $\tau = \arccos \frac{-a^3\Delta^3 + 9a^2\Delta^2 - 27a\Delta/2}{(a^2\Delta^2 - 6a\Delta)^{3/2}}$. Notice that $\tau \in [0, \pi]$. The inequalities $x_1 \leq 1 \leq x_2 \leq x_3$ follow from elementary properties of the trigonometric functions. Let $\Delta_{3,2} = \frac{27}{4a}$, and $u_{3,2} = x_3$.

For $p = 4$ the given equation has solutions [19]

$$\begin{aligned} x_1 &= \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau}{3} + \frac{a\Delta}{6}} + \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau + 2\pi}{3} + \frac{a\Delta}{6}} \\ &\quad + \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau - 2\pi}{3} + \frac{a\Delta}{6}}, \\ x_2 &= \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau}{3} + \frac{a\Delta}{6}} - \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau + 2\pi}{3} + \frac{a\Delta}{6}} \\ &\quad - \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau - 2\pi}{3} + \frac{a\Delta}{6}}, \\ x_3 &= -\sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau}{3} + \frac{a\Delta}{6}} + \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau + 2\pi}{3} + \frac{a\Delta}{6}} \\ &\quad - \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau - 2\pi}{3} + \frac{a\Delta}{6}}, \\ x_4 &= -\sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau}{3} + \frac{a\Delta}{6}} - \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau + 2\pi}{3} + \frac{a\Delta}{6}} \\ &\quad + \sqrt{\sqrt{\frac{a^2\Delta^2}{36} - \frac{a\Delta}{3}} \cos \frac{\tau - 2\pi}{3} + \frac{a\Delta}{6}} \end{aligned}$$

(where $\tau = \arccos \frac{18a^2\Delta^2 - a^3\Delta^3}{(a^2\Delta^2 - 12a\Delta)^{\frac{3}{2}}}$) which are real as long as all the functions inside the square roots are positive and the argument of the arccos function is between -1 and $+1$. This happens for $\Delta > 16/a$. If this is the case then x_1 is the largest solution and $x_1 \geq 2$.

Solutions to polynomial equations of degree at least five cannot be expressed by simple algebraic expressions involving radicals and the four arithmetic operations [17]. Although complicated expressions do exist (see [15,16]) the constraints on x obtained by solving the weaker inequality $x^p - a\Delta < 0$ will suffice for our purposes. Therefore, for $p \geq 5$, let $\Delta_{p,2} = \frac{2^p}{a}$ and $u_{p,2} = (a\Delta)^{\frac{1}{p}}$. \square

3. AN ALGORITHM

Let $Y = \{x_{j_1}, x_{j_2}, \dots, x_{j_t}\} \subseteq X$ with t to be fixed later. For any integer $b \geq 1$ the set $Y_b = \{x_{j_b}, \dots, x_{j_t}\}$ is called a *final segment* of Y , with the convention that $Y_1 = Y$ and $Y_b = \{\}$ if $b > t$. Consider the following modification of the classical Davis *et al.* [9] resolution algorithm:

Function RoughDLL (ϕ : k -CNF; Y : set of variables; α : truth-assignment):
 Boolean
 if $\phi = \{\}$ **return** true
 else if $\{\} \in \phi$ **return** false
 else if $Y = \{\}$
 return 2SAT-solver($C_2(\phi, \alpha)$)
 else
 Let x be the smallest index variable in Y ;
 $Y \leftarrow Y \setminus \{x\}$;
 return RoughDLL($\phi|_{x=0}, Y, \alpha \cup \{x=0\}$) \vee
 RoughDLL($\phi|_{x=1}, Y, \alpha \cup \{x=1\}$);

where 2SAT-solver(\dots) is a function deciding 2-SAT.

The algorithm recursively resolves a fraction $\frac{t}{n}$ of the variables of the formula ϕ , then it calls a solver for 2-SAT on any unfinished recursion branch. If the input formula ϕ is unsatisfiable then a call to RoughDLL($\phi, Y, \{\}$) will return the correct “false” answer provided either $\{\} \in \phi|_{\alpha}$ or $C_2(\phi, \alpha)$ is unsatisfiable, for every α with $\text{Dom}(\alpha) = Y$. Furthermore, the recursive calls to RoughDLL naturally induce a rooted binary tree, $T_{\phi, Y}$, whose internal nodes are labelled by the variables that are set at a particular step, with the out-edges of a node labelled by the two possible assignments to its associated variable. Each path from the root in $T_{\phi, Y}$ corresponds to a partial assignment α with $\text{Dom}(\alpha) \subseteq Y$. Each leaf is labelled by either a clause of ϕ that becomes empty or by the set $C_2(\phi, \alpha)$. If all the formulae $C_2(\phi, \alpha)$ are unsatisfiable, $T_{\phi, Y}$ can be transformed into a refutation of ϕ , by working from the clauses labelling the leaves and the refutations obtained for each $C_2(\phi, \alpha)$ towards the root of $T_{\phi, Y}$. Theorem 1.1 will be proved by showing that w.h.p. the refutations defined in this way can be pebbled with (relatively) few pebbles.

4. PROOF OF THE MAIN THEOREM

Since $\Delta > \theta_k$ the set of satisfiable formulae in $\mathcal{F}_{\Delta n}^{k,n}$ is very small. In the following we assume ϕ to be unsatisfiable. We will prove that it is possible to choose t so that RoughDLL ends with a “false” answer w.h.p. If this is the case, the refutation built using the algorithm in Section 3 is formed by joining the refutations for $C_2(\phi, \alpha)$ to the complete binary tree of depth t corresponding to the branching of RoughDLL. By Theorem 2.2 this tree can be pebbled using $t + 1$ pebbles. The result then follows from Theorem 2.3 applied to $C_2(\phi, \alpha)$ for each α .

To complete the proof note that, conditioned on the fact that $|C_2(\phi, \alpha)| = \Omega(n)$ for each α on Y , the event

“RoughDLL does not end with a ‘false’ answer”

is implied by the event

“there is an α with $\text{Dom}(\alpha) = Y$, such that $C_2(\phi, \alpha)$ is satisfiable”

and the probability of the latter is at most

$$\sum_{\alpha: \text{Dom}(\alpha)=Y} \Pr[C_2(\phi, \alpha) \in \text{SAT}].$$

Let “ $\psi \in \text{SAT}$ ” denote the event “the formula ψ is satisfiable”. For each α we can compute $\Pr[C_2(\phi, \alpha) \in \text{SAT}]$ conditioning on the size X of $C_2(\phi, \alpha)$:

$$\begin{aligned} \Pr[C_2(\phi, \alpha) \in \text{SAT}] &= \Pr[C_2(\phi, \alpha) \in \text{SAT} \mid X < dn] \Pr[X < dn] \\ &\quad + \Pr[C_2(\phi, \alpha) \in \text{SAT} \mid X \geq dn] \Pr[X \geq dn] \\ &\leq \Pr[X < dn] + \Pr[C_2(\phi, \alpha) \in \text{SAT} \mid X \geq dn] \end{aligned}$$

where $d > 0$ is some constant to be fixed later.

Since clauses in ϕ are selected independently and with replacement from $\mathcal{C}^{k,n}$, given Y and α on Y , in each of the $m = \Delta n$ trials there is a fixed probability of selecting a clause C such that $C|_{\alpha}$ is a 2-clause. This is exactly the probability of choosing $k - 2$ variables from Y with a sign fixed by the assignment α and the remaining two arbitrarily in the set $\mathcal{X} \setminus Y$. Hence, the random variable x has a binomial distribution with parameters m and $p =_{df} \Pr[C|_{\alpha}$ is a 2-clause]. Furthermore

$$\begin{aligned} p &= \frac{4 \binom{n-t}{2} \binom{t}{k-2}}{2^k \binom{n}{k}} \\ &= 2^{2-k} \binom{k}{2} (n-t)(n-t-1) \frac{t!}{(t-k+2)!} \frac{(n-k)!}{n!} \\ &= 2^{2-k} \binom{k}{2} \frac{t}{n} \left(1 - \frac{t-1}{n-1}\right) \left(1 - \frac{t-1}{n-2}\right) \prod_{i=1}^{k-3} \frac{t-i}{n-i-2} \\ &> 2^{2-k} \binom{k}{2} \frac{t}{n} \left(1 - \frac{t}{n}\right)^{2k-3} \prod_{i=1}^{k-3} \frac{t-i}{n-i-2} \end{aligned}$$

where the last inequality holds as long as $t < \frac{n}{2}$, and the product is empty for $k = 3$.

Claim. For every $k \geq 4$, if $t \geq k^2$ there exists a $c_k > 0$ such that $\prod_{i=1}^{k-3} \frac{t-i}{n-i-2} \geq c_k \left(\frac{t}{n}\right)^{k-3}$.

From the claim above it follows that

$$\Pr[C|_{\alpha} \text{ is a 2-clause}] > c_k \binom{k}{2} \left(1 - \frac{t}{n}\right)^2 \left(\frac{t}{2n}\right)^{k-2}.$$

Finally, using standard Chernoff-type bounds [13], for any $\epsilon > 0$,

$$\begin{aligned} \Pr \left[X < (1 - \epsilon)c_k \binom{k}{2} \left(1 - \frac{t}{n}\right)^2 \left(\frac{t}{2n}\right)^{k-2} \Delta n \right] &\leq \Pr[X < (1 - \epsilon)mp] \\ &\leq e^{-\frac{\epsilon^2}{2} c_k \binom{k}{2} \left(1 - \frac{t}{n}\right)^2 \left(\frac{t}{2n}\right)^{k-2} \Delta n}. \end{aligned}$$

The second probability in the expression for $\Pr[C_2(\phi, \alpha) \in \text{SAT}]$ is bounded using Lemma 2.4, since, conditioned on $|C_2(\phi, \alpha)|$ to have some known value $z \geq dn$, clauses in $C_2(\phi, \alpha)$ are distributed according to $\mathcal{F}_z^{2, n-t}$. More precisely, since the probability of satisfying a formula decreases as the number of clauses in the formula increases, the sought probability satisfies:

$$\Pr[C_2(\phi, \alpha) \in \text{SAT} \mid X \geq d(n-t)] \leq n^{O(1)} \Pr[C_2(\phi, \alpha) \in \text{SAT} \mid X = d(n-t)]$$

where $d = (1 - \epsilon)c_k \binom{k}{2} \left(1 - \frac{t}{n}\right)^2 \left(\frac{t}{2n}\right)^{k-2} \Delta$.

Finally, putting everything together, the overall error probability is at most

$$\begin{aligned} &\exp \left\{ -n \left[\frac{\epsilon^2 c_k \Delta}{2} \binom{k}{2} \left(1 - \frac{t}{n}\right)^2 \left(\frac{t}{2n}\right)^{k-2} - \frac{t \ln 2}{n} \right] \right\} \\ &+ n^{O(1)} \exp \left\{ -n \left[(1 - \epsilon) \frac{3\Delta t}{2n} \left(1 - \frac{t}{n}\right)^2 \ln \frac{4}{3} - \ln 2 \right] \right\}. \end{aligned}$$

Setting $t/n = 1/x$, the proof of the theorem can be completed by solving the following optimisation problem (where $c_3 = 1$):

$\begin{aligned} &\max \quad x \\ &\text{s.t.} \quad (1 - \epsilon)c_k \Delta \binom{k}{2} \left(1 - \frac{1}{x}\right)^2 \left(\frac{1}{2x}\right)^{k-2} - 1 > 0, \\ &\quad \frac{\epsilon^2 c_k \Delta}{2} \binom{k}{2} \left(1 - \frac{1}{x}\right)^2 \left(\frac{1}{2x}\right)^{k-2} - \frac{\ln 2}{x} > 0, \\ &\quad (1 - \epsilon) \frac{3\Delta}{2x} \left(1 - \frac{1}{x}\right)^2 \ln \frac{4}{3} - \ln 2 > 0. \end{aligned}$
--

For $k = 3$, the first constraint is weaker than the third one. By Lemma 2.5, the maximisation problem is feasible for

$$\begin{aligned} x &> \left(1 - \frac{2}{\epsilon} \sqrt{\frac{\ln 2}{3\Delta}}\right)^{-1} \\ x &> \frac{a\Delta}{3} - \frac{2}{3} \sqrt{a^2\Delta^2 - 6a\Delta} \cos \frac{\tau - 2\pi}{3} \\ x &< \frac{a\Delta}{3} - \frac{2}{3} \sqrt{a^2\Delta^2 - 6a\Delta} \cos \frac{\tau + 2\pi}{3} \end{aligned}$$

where $a = 3(1 - \epsilon) \left(1 - \frac{\ln 3}{\ln 4}\right)$, and $\tau = \arccos \frac{-a^3\Delta^3 + 9a^2\Delta^2 + 27a\Delta/2}{(a^2\Delta^2 - 6a\Delta)^{3/2}}$ provided $\Delta > \max \left\{ \frac{9}{4(1-\epsilon)\left(1 - \frac{\ln 3}{\ln 4}\right)}, \frac{4\ln 2}{3\epsilon^2} \right\}$. If $\Delta > \frac{4 \cdot 31^2 \ln 2}{3 \cdot 21^2 \epsilon^2}$, the first lower bound is smaller than the upper bound in the third line, and therefore x can be set to any sufficiently close lower bound on $\frac{a\Delta}{3} - \frac{2}{3} \sqrt{a^2\Delta^2 - 6a\Delta} \cos \frac{\tau + 2\pi}{3}$. Let ϵ be the solution to

$$\frac{9}{4(1-\epsilon)\left(1 - \frac{\ln 3}{\ln 4}\right)} = \frac{4 \cdot 31^2 \ln 2}{3 \cdot 21^2 \epsilon^2}.$$

For sufficiently small $\delta > 0$ one can set

$$\frac{t}{n} = \frac{1}{(1-\epsilon)\Delta\left(1 - \frac{\ln 3}{\ln 4}\right) \left(1 - \sqrt{1 - \frac{2}{(1-\epsilon)\Delta\left(1 - \frac{\ln 3}{\ln 4}\right)} \cos \frac{\tau + 2\pi}{3}}\right) - \delta}.$$

For $k > 3$ the first constraint is stronger than the third one: the expression on the left of the ‘>’ sign in the first constraint is smaller than that in the third constraint for

$$x > \frac{1}{2} \left[\frac{c_k k(k-1)}{12 \ln(4/3)} \right]^{\frac{1}{k-3}}.$$

The quantity on the right-hand side is less than two for $k \geq 4$.

For $k = 4$, if $\Delta > \frac{(\epsilon^2 + 2(1-\epsilon)\ln 2)^3}{3\epsilon^4 c_4 (1-\epsilon)^2 \ln 2}$ then the least upper bound on the set of points x that satisfy the first constraint is smaller than the similar quantity related to the second constraint. Thus, by Lemma 2.5, the given optimisation problem is feasible for $x < u_{3,1}$ (where $a = \frac{3}{2}(1-\epsilon)c_4$ and ϵ can be chosen so that the lower bound on Δ is minimised). For a sufficiently small $\delta > 0$, then one can set

$$\frac{t}{n} = \frac{1}{\sqrt{\frac{(1-\epsilon)c_4\Delta}{2}} \left(\cos \frac{\tau}{3} + \sqrt{3} \sin \frac{\tau}{3}\right) - \delta}$$

where $\tau = \arccos \frac{3}{\sqrt{2(1-\epsilon)c_4\Delta}}$.

A similar argument applies for $k = 5$, and $\frac{t}{n}$ can be defined again as $O(n\Delta^{-\frac{1}{k-2}})$ provided $\Delta > \frac{2(\epsilon^2 + 2(1-\epsilon)\ln 2)^4}{5\epsilon^6 c_5 (1-\epsilon)^2 \ln 2}$ and ϵ is chosen so that the bound on Δ is minimal.

For $k \geq 6$ we resort to the weaker bound given in Lemma 2.5. We can set

$$\frac{t}{n} = \left\{ \frac{1}{2} \left[\frac{(1-\epsilon)c_k\Delta}{2} \binom{k}{2} \right]^{\frac{1}{k-2}} - \delta \right\}^{-1}$$

provided ϵ is chosen so that $\frac{\epsilon^2}{2(1-\epsilon)\ln 2} = 1$ and $\Delta > \frac{2^{2k-2}}{(1-\epsilon)c_k k(k-1)}$.

5. FINAL REMARKS AND OPEN QUESTIONS

In this paper we presented a class of refutations which can be associated with high probability with any given unsatisfiable random k -CNF formula on n variables and $m = \Delta n$ clauses. A pebbling game can be played on the directed acyclic graphs corresponding to these refutations and relatively few pebbles are sufficient to win such game. As a consequence of this an upper bound can be obtained on the space complexity of refuting unsatisfiable random k -CNF formulae in resolution.

It should be said also that slightly weaker bounds can be derived, avoiding most of the technicalities in this paper, from results on the size of the refutations produced by the algorithm outlined above [3]. However the numerical precision sought in this paper, can perhaps be better understood by reading the author's original motivations. The result presented came about as part of an investigation on the unsatisfiability threshold (see Sect. 2.2) of random k -CNF formulae and its related properties. It is believed [6] that formulae with a clause to variable ratios close to the threshold point are computationally hard. The fact that w.h.p. such formulae, at least for not too large values of $\Delta > \theta_k$, have large resolution refutations (see [7] and the subsequent improvement in [3]) can be seen as supporting such a claim. The results in [4] prove that a similar phenomenon occurs if one looks at space complexity instead of size. The result in this paper gives an upper bound on the space complexity for values of Δ not much bigger than the unsatisfiability threshold.

The analysis presented might be tightened by using refined bounds [5] on the probability that a random 2-CNF formula on n variables and $m > n$ clauses be satisfiable. Moreover, "more efficient" refutations might exist. However, the lower bound $\Omega\left(n \cdot \Delta^{-\frac{1+\epsilon}{k-2-\epsilon}}\right)$ proved in [4] (for any $\epsilon \in (0, 1/2)$) rules out the possibility of a significant improvement on the results presented.

Acknowledgements. The author wishes to thank Jacobo Torán for many helpful discussions, and Nicola Galesi for informing him of reference [4]. The author is also indebted to the anonymous referees for pointing out a significant simplification and improvement of the original draft, and for bringing to his attention reference [3].

REFERENCES

- [1] D. Achlioptas and G.B. Sorkin, Optimal myopic algorithms for random 3-SAT, in *41st Annual Symposium on Foundations of Computer Science* (2000) 590-600.

- [2] M. Alekhovich, E. Ben-Sasson, A.A. Razborov and A. Wigderson, Space complexity in propositional calculus, in *Proc. of the 32nd Annual ACM Symposium on Theory of Computing*. Portland, OR (2000).
- [3] P. Beame, R. Karp, T. Pitassi and M. Saks, On the complexity of unsatisfiability proofs for random k -CNF formulas, in *Proc. of the 30th Annual ACM Symposium on Theory of Computing*. New York (1998) 561-571.
- [4] E. Ben-Sasson and N. Galesi, Space complexity of random formulae in resolution, in *Proc. of the 16th Annual IEEE Conference on Computational Complexity*. IEEE Computer Society (2001).
- [5] B. Bollobás, C. Borgs, J.T. Chayes, J.H. Kim and D.B. Wilson, The scaling window of the 2-SAT transition. *Random Structures Algorithms* **18** (2001) 201-256.
- [6] P. Cheeseman, B. Kanefsky and W.M. Taylor, Where the really hard problems are, in *Proc. of IJCAI-91*, edited by Kauffmann (1991) 331-337.
- [7] V. Chvátal and E. Szemerédi, Many hard examples for resolution. *J. ACM* **35** (1988) 759-768.
- [8] S.A. Cook and R. Reckhow, The relative efficiency of propositional proof systems. *J. Symb. Logic* **44** (1979) 36-50.
- [9] M. Davis, G. Logemann and D. Loveland, A machine program for theorem proving. *Commun. ACM* **5** (1962) 394-397.
- [10] M. Davis and H. Putnam, A computing procedure for quantification theory. *J. ACM* **7** (1960) 201-215.
- [11] J. Esteban and J.L. Toran, Space bounds for resolution, edited by C. Meinel and S. Tison, in *STACS 99: 16th Annual Symposium on Theoretical Aspects of Computer Science, Trier, Germany, March 1999, Proceedings*. Springer-Verlag, *Lecture Notes in Comput. Sci.* **1563** (1999) 551-560.
- [12] A. Goerdts, A threshold for unsatisfiability. *J. Comput. System Sci.* **53** (1996) 469-486.
- [13] T. Hagerup and C. Rüb, A guided tour of Chernoff bounds. *Inform. Process. Lett.* **33** (1989) 305-308.
- [14] A.C. Kaporis, L.M. Kirousis, Y.C. Stamatiou, M. Vamvakari and M. Zito, Coupon collectors, q -binomial coefficients and the unsatisfiability threshold, edited by A. Restivo, S. Ronchi della Rocca and L. Roversi. *Theoret. Comput. Sci.*, in *7th Italian Conference, ICTCS 2001*. Springer-Verlag, *Lecture Notes in Comput. Sci.* **2202** (2001) 328-338.
- [15] R. Bruce King, *Beyond the quartic equation*. Birkhauser, Boston, MA (1996).
- [16] R. Bruce King and E. Rodney Canfield, An algorithm for calculating the roots of a general quintic equation from its coefficients. *J. Math. Phys.* **32** (1991) 823-825.
- [17] I. Stewart, *Galois Theory*. Chapman and Hall, London (1973).
- [18] J. Toran, Lower bounds for the space used in resolution, edited by J. Flum and M. Rodríguez-Artalejo, in *Computer Science Logic: 13th International Workshop, CSL'99, 8th Annual Conference of the EACSL, Madrid, Spain, September 20-25, 1999, Proceedings*. Springer-Verlag, *Lecture Notes in Comput. Sci.* **1683** (1999).
- [19] B.L. van der Wården, *Algebra*. Frederick Ungar Publishing Co., New York (1970).
- [20] R.C. Weast, *Handbook of Tables for Mathematics*. Cleveland: The Chemical Rubber Co. (1964).

Communicated by J. Balcazar.

Received April, 2001. Accepted November, 2002.