comput. complex. **19** (2010), 501–519
1016-3328/10/040501-19, *published online* 11 August 2010
DOI 10.1007/s00037-010-0293-1



computational complexity

# RANDOM CNF'S ARE HARD FOR THE POLYNOMIAL CALCULUS

Eli Ben-Sasson and Russell Impagliazzo

**Abstract.** We prove linear lower bounds on the Polynomial Calculus (PC) refutation-degree of random CNF whenever the underlying field has characteristic greater than 2. Our proof follows by showing the PC *refutation-degree* of a unsatisfiable system of linear equations modulo 2 is equivalent to its *Gaussian width*, a concept defined by the late Mikhail Alekhnovich.
The equivalence of refutation-degree and Gaussian width which is the main contribution of this paper, allows us to also simplify the refutation-degree lower bounds of Buss *et al.* (2001) and additionally prove non-trivial upper bounds on the resolution and PC complexity of refuting unsatisfiable systems of linear equations.

**Keywords.** Propositional proof complexity, polynomial calculus, Groebner basis, random CNF formulae.

**Subject classification.** 03F20, 03B05, 68Q17.

## 1. Introduction

The seminal paper of Chvátal & Szemerédi (1988) showed that almost every unsatisfiable 3-CNF formula with $n$ variables and $cn$ clauses (for $c$ a large enough constant), is extremely hard for Resolution to refute, i.e. the refutation size is exponential in $n$. This result came shortly after lower bounds for concrete contradictions were shown, most notably for the unsatisfiable CNF encoding the pigeonhole principle (Haken 1985) and for Tseitin contradictions over expander graphs (Urquhart 1987), and showed that the weakness of Resolution is not limited to specially tailored formulas. On the contrary, the formulas that are easy to refute are the exception.

In the 1990's several algebraic proof systems have entered the proof complexity scene, the Polynomial Calculus (PC) drawing most attention, because of its simplicity and partial automatizability (Clegg *et al.* 1996; Pitassi 1996).

The main complexity measure of PC is the minimal refutation-degree, which was shown by Clegg *et al.* (1996) to be closely related to minimal PC refutation size. Several refutation-degree lower bounds have been established for concrete contradictions such as the pigeonhole principle (Impagliazzo *et al.* 1999a; Razborov 1998), Tseitin contradictions and other counting principles (Buss *et al.* 2001). A major open question has been whether PC displays a weakness for random inputs, similar to that of Resolution. The main result of this paper is a positive resolution of this question for all fields of characteristic greater than 2. The case of characteristic 2 was resolved by Alekhnovich & Razborov (2003) via different techniques shortly after our initial report was published. Our proof technique allows us to derive the bounds of Buss *et al.* (2001) in a simpler manner and to prove nontrivial upper bounds on the refutation complexity of systems of linear equations modulo 2 in Resolution and PC. Next we describe in detail our main results, followed by an overview of the proof technique.

### 1.1. Main results.

**Lower bounds for random CNF's.** It is well-known since the work of Chvátal & Szemerédi (1988) that with high probability a random $k$-CNF with $n$ variables and $O(n)$ clauses, demands exponential size resolution refutations. Stated simply: Resolution is usually no better than exhaustively checking all possible $2^n$ assignments. The simplest explanation for this inefficiency is that Resolution performs badly on inputs that have high expansion (according to Definition 3.2 below) and random CNFs do have large expansion. Our first main result, stated next, is that PC does not fare significantly better on random CNF and for similar reasons. The main complexity measure studied in the context of the polynomial calculus is that of *refutation degree* (see Section 2 for a formal definition). The refutation degree of a CNF $\mathcal{F}$, formulated as a set of polynomials, is the minimal degree $d$ such that there exists a polynomial calculus refutation of $\mathcal{F}$ in which each line (which is a polynomial) has degree at most $d$.

Definition 1.1 (Random $k$-CNF's). *Let $\mathcal{F} \sim \mathcal{F}_k^{n,\Delta}$ denote that $\mathcal{F}$ is a random $k$-CNF formula on $n$ variables and $m = \Delta \cdot n$ clauses, chosen at random by picking $\Delta \cdot n$ clauses i.i.d from the set of all $\binom{n}{k} \cdot 2^k$ clauses, with repetitions. $\Delta$ is called the* clause density.

THEOREM 1.2 (Random CNFs are hard for PC). *For integer $k \geq 3$ and $\Delta > 0$ there exists a constant $\epsilon = \epsilon(k, \Delta) > 0$ such that the following holds. For $F$ a field of characteristic greater than 2, with high probability the refutation-degree of $\mathcal{F} \sim \mathcal{F}_k^{n,\Delta}$ over $F$ is at least $\epsilon n$.*

As shown in (Impagliazzo *et al.* 1999b, Section 6), the minimal number of monomials appearing in a polynomial calculus refutation of a 3-CNF $\mathcal{F}$ over $n$ variables is exponential $d^2/n$ where $d$ denotes the refutation degree of $\mathcal{F}$. Thus, Theorem 1.2 also implies exponential lower bounds on the size of polynomial calculus refutations of a random CNF.

**Simplified lower bounds for Tseitin and related contradictions.** The work reported in Buss *et al.* (2001) proved linear lower bounds on the refutation-degree of Tseitin contradictions and certain counting principles. We provide a different proof of these statements and as an example provide the proof for Tseitin contradictions. Recall that a linear equation modulo 2 of the form $\sum_{i=1}^{k} x_i \equiv b \pmod 2$ for $b \in \{0,1\}$ can be expressed in a unique way as a $k$-CNF formula with $2^{k-1}$ clauses of size $k$ each. We call this CNF the *defining CNF* of the linear equation and the clauses of this CNF are referred to as the *defining clauses* of the linear equation.

DEFINITION 1.3 (Tseitin contradiction). *Let $G = (V, E)$ be an undirected simple graph over $n$ vertices. Let $b : V \to \{0,1\}$ be an odd-labeling of $V$, i.e., a labeling satisfying $\sum_{v \in V} b(v) \equiv 1 \pmod 2$. The* Tseitin contradiction *over $G$ and $b$, denoted $\mathcal{T}(G, b)$, is the CNF over variable set $\{x_e : e \in E\}$ that is the conjunction of the defining CNFs of the following linear constraints*

$$\sum_{e \ni v} x_e \equiv b(v) \pmod 2, \quad v \in V\,.$$

It is well-known since the work of Tseitin (1968); Urquhart (1987) that $\mathcal{T}(G, b)$ is unsatisfiable and furthermore, requires exponentially long proofs when $G$ is an expander graph. This result was extended to the case of PC refutation-degree in Buss *et al.* (2001) who proved the following result, for which our techniques provide an arguably simpler proof (see Section 3).

THEOREM 1.4 (Buss *et al.* 2001). *Let $G = (V, E)$ be a connected graph and $b : V \to \{0,1\}$ be an odd-labeling of $V$. For $\mathcal{T}(G, b)$ the Tseitin contradiction over $G$ and $b$, and $F$ a field of characteristic greater than 2, the refutation-*

*degree of $\mathcal{T}(G, b)$ over $F$ is at least half the* expansion *of $G$, defined as*

$$e(G) = \min\left\{|E(S, V \setminus S)| : \frac{|V|}{3} \leq |S| \leq \frac{2|V|}{3}\right\}.$$

**Upper bounds on refutation-degree.** Our final set of results use the refutation system for unsatisfiable systems of linear equations introduced in Section 2 under the name *Gaussian calculus* to give nontrivial upper bounds on the resolution proof size and PC refutation-degree of systems of linear systems modulo 2. In particular, we prove that every such system has a resolution refutation of size at most $2^{n/2+o(n)}$ and a PC refutation of degree at most $n/4 + o(n)$. This is in contrast to work reported by Pudlák & Impagliazzo (2000) which showed that tree-like resolution of such systems can require size $2^{(1-\epsilon)n}$ for arbitrary $\epsilon > 0$, thus approaching exhaustive search.

**1.2. Proof techniques.** All results reported in this paper come from considering unsatisfiable systems of linear equations modulo 2. Let $Ax = b$ denote such a system, where $A$ is a $m \times n$ matrix with $\{0, 1\}$ values, $b \in \{0, 1\}^m$ and $x = x_1, \ldots, x_n$ is a set of variables. Basic linear algebra shows that $Ax = b$ is unsatisfiable if and only if there exists $y \in \{0, 1\}^m$ such that $y^\top A = (0, 0, \ldots, 0)^\top$ and $y^\top b = 1$, where all computations are carried out modulo 2. Thus, the vector $y$ can be viewed as a refutation of $Ax = b$.

The following analogous view of this refutation, suggested by the late Mikhail Alekhnovich [personal communication, 1999], will be crucial for obtaining our results. View the $i^{\text{th}}$ row of $A_i$ and the $i^{\text{th}}$ element of $b$ as a constraint on $x$. A *Gaussian refutation* of $Ax = b$ is a sequence of lines, where each line is either a constraint $A_i x = b_i$ or a linear combination of two previous lines and the final line is $(0, 0, \ldots, 0)^\top x = 1$. The *Gaussian width* of such a refutation is the maximal number of variables appearing in a line in the refutation and the *Gaussian refutation width* of the system is the minimal Gaussian width taken over all Gaussian refutations.

The observation of Alekhnovich is that the Gaussian width of $Ax = b$ is essentially equal to the resolution refutation width of the CNF defining $Ax = b$. Our main observation is that the Gaussian width is also equivalent to half the refutation-degree of the related CNF when working in a field of characteristic $> 2$. With this observation we can apply the width-method of Ben-Sasson & Wigderson (2001) to obtain linear lower bounds on the refutation-degree of various *expanding* sets of constraints. Informally, a set of constraints is said to be expanding if every relatively small subset of constraints mentions many distinct variables.

As an application of our proof technique we now sketch the proof of the lower bound on refutation-degree of random 3CNFs. A set of linear constraints modulo 2, where each constraint mentions at most 3 variables, can be encoded as a 3CNF in a natural way. The converse is also true, namely, a 3CNF can be viewed as a subset of clauses that define a linear system modulo 2 with at most 3 variables per constraint. Furthermore, if the 3CNF is expanding, so is the underlying linear system. Random 3CNFs are well-known to be expanding, implying their underlying systems are expanding. By our key observation such linear systems have large refutation-degree over fields of characteristic $> 2$. This also implies that the 3CNF which induced them must have large refutation-degree, because this CNF can be derived from the linear system using constant degree. This completes the sketch of the proof of Theorem 1.2.

**1.3. Subsequent developments.** The main open question left by our work was to understand the PC refutation-degree of random CNFs over fields of characteristic 2. This problem was resolved via different methods than reported here by Alekhnovich & Razborov (2003).

**Organization of the rest of the paper.** In Section 2 we define Gaussian width and show it is equivalent to PC refutation-degree for linear systems modulo 2. Section 3 presents lower bounds for Gaussian width in terms of the expansion of the input system. Section 4 proves Theorem 1.2 by reducing random linear equations to random CNF's. In Section 5 we conclude with upper bounds for Resolution width and PC refutation-degree for systems of linear equations.

## 2. Gaussian width and refutation degree

In this section we define *Gaussian width* and prove that it is equivalent to PC refutation-degree for systems of linear equations modulo 2 that are formulated over fields of characteristic greater than 2.

**Notation.** The letter $F$ will denote a field and $F_q$ a finite field of size $q$. Calligraphic letters denote sets of formulas: $\mathcal{P}$ will be a set of polynomials, and $\mathcal{L}$ will be reserved for sets of linear equations.

We shall start by dealing with unsatisfiable sets of linear equations mod $q$, a prime (most of the paper sets $q = 2$). Let $\mathcal{L} = \{\ell_i\}_{i=1}^m$ be such a set over $n$ variables $x_1, \ldots, x_n$. For simplicity let us assume that $\mathcal{L}$ is *minimal* unsatisfiable, i.e. every proper subset of it is satisfiable. For reasons that will become

clear later on, we allow variables to appear on both sides of the equality sign, so formally $\ell_i$ has the following syntactic structure:

$$\sum_{j \in S_{i_1}} a_j \cdot x_j = c + \sum_{j \in S_{i_2}} b_j \cdot x_j \tag{2.1}$$

where $a_j, b_j \in F_q \backslash \{0\}$ and $c \in F_q$. We say that two lines are *equivalent modulo q* if subtracting one from the other modulo $q$ gives the equation $0 = 0$.

Perhaps the most natural way to prove that $\mathcal{L}$ is unsatisfiable is by Gaussian elimination in the field $F_q$. We start with the lines of $\mathcal{L}$ as axioms and produce new linear equations by addition of existing lines and scalar multiplication. We prove that $\mathcal{L}$ is unsatisfiable by deriving the unsatisfiable linear equation $1 = 0$. The number of lines needed to refute $\mathcal{L}$ is at most $m$. Let us call such a refutation a *Gaussian Calculus refutation*, or simply, a Gaussian refutation. Let the *Gaussian width* of refuting $\mathcal{L}$, denoted $w_G(\mathcal{L})$, be the maximal number of variables appearing in a line of the Gaussian refutation.

Suppose we wish to show that $\mathcal{L}$ is unsatisfiable, but we work within a field $F$ with characteristic $p \neq q$, which has a primitive $q$'th root of unity called $\omega$. The multiplicative group generated by $\omega$ is isomorphic to the additive group $Z_q$, and we use this isomorphism to translate $\mathcal{L}$ into a set of polynomials over $F$ such that there is a bijection between the set of assignments satisfying $\mathcal{L}$ and the set of common roots of the corresponding set of polynomials (such a set is called an algebraic set or a variety).

Definition 2.2 (Linear equations modulo $q \neq p$ as a set of polynomials). *Let $\mathcal{L} = \{\ell_i\}_{i=1}^m$ be a minimal unsatisfiable set of linear equations mod $q$, where $\ell_i$ is as described in* (2.1). *Let $F$ be a field of characteristic $p \neq q$, having a primitive root of unity of order $q$, denoted $\omega$. The* polynomial translation *of the equation $\ell_i$ to $F$, denoted $P_F(\ell_i)$ is the following polynomial:*

$$\prod_{j \in S_{i_1}} y_j^{a_j} - \omega^c \cdot \prod_{j \in S_{i_2}} y_j^{b_j} \,. \tag{2.3}$$

*The translation of $\mathcal{L}$ to $F$, denoted $\mathcal{P}_F(\mathcal{L})$ is the set of polynomials that is the union of:*

*(i)* $\{y_i^q - 1\}_{i=1}^n$. *This set ensures the roots of the system of polynomials are contained in* $\{1, \omega, \omega^2, \ldots, \omega^{q-1}\}^n$.

*(ii)* $\{P_F(\ell_i)\}_{i=1}^m$.

REMARK 2.4. *(i) Notice $\mathcal{P}_F(\mathcal{L})$ is a set of* binomials, *i.e. polynomials with two monomials. This simple observation will play a crucial role in our investigation.*

*(ii) The transformation described in the previous definition is invertible, i.e. every binomial $P$ which has structure as in* (2.3) *where $a_j, b_j, c \in \{0, 1, \ldots, q-1\}$, corresponds to a linear equation $\ell$. Moreover, $\ell$ is satisfied by an assignment $\alpha = (\alpha_1, \ldots, \alpha_n) \in \{0, \ldots, q-1\}^n$ iff $(\omega^{\alpha_1}, \ldots, \omega^{\alpha_n})$ is a root of $P_F(\ell)$.*

*(iii) Suppose $P_F(\ell_1), P_F(\ell_2)$ are polynomial translations of $\ell_1, \ell_2$, respectively. Let $I$ denote the ideal generated by $\{y_i^q - 1\}_{i=1}^n$. Then $\ell_1$ is equivalent to $\ell_2$ modulo $q$ if and only if $P_F(\ell_1) \equiv P_F(\ell_2) \pmod{I}$.*

If $\mathcal{L}$ is unsatisfiable then $\mathcal{P}_F(\mathcal{L})$ is unsatisfiable in $F$, i.e., it has no common roots, but Gaussian elimination is not sufficient for proving this fact. In order to reach a contradiction, one must allow multiplication of lines by variables, working within the framework of the Polynomial Calculus, introduced in Clegg *et al.* (1996) and further developed in Beame *et al.* (1994); Razborov (1998).

DEFINITION 2.5 (The Polynomial Calculus). *Let $F$ be a fixed field, and $\mathcal{P}$ a set of polynomials over $F$. A* polynomial calculus derivation, *or simply, derivation, of a polynomial $Q$ from $\mathcal{P}$ is a sequence of polynomials $\pi = \{P_1, \ldots, P_S\}$ such that $P_S = Q$, and each $P_i$ is either an axiom from $\mathcal{P}$ or derived from previous polynomials by addition or by multiplication by scalars and variables. A* refutation *is a derivation of the polynomial* 1 *(that clearly has no roots). The set $\mathcal{P}$ is said to be* unsatisfiable *or* contradictory *iff its set of roots is empty, which happens iff $\mathcal{P}$ has a refutation.*

The natural complexity measure for the Polynomial Calculus is its refutation-degree, defined next.

DEFINITION 2.6 (Refutation-degree). *The* degree *of a Polynomial Calculus refutation is the* maximal degree *of a polynomial appearing in the refutation.*

*The* refutation-degree *of an unsatisfiable set of polynomials $\mathcal{P}$, denoted $d(\mathcal{P})$, is the minimal degree of a refutation of $\mathcal{P}$.*

The main theorem of this section is stated next. The rest of this section is devoted to its proof.

THEOREM 2.7. *Let $\mathcal{L}$ be an unsatisfiable set of linear equations modulo 2 of width at most $k$, let $F$ be a field of characteristic $p \neq 2$, and let $\mathcal{P} = \mathcal{P}(\mathcal{L})$. Then,*

$$\max\left\{k, \frac{1}{2}w_G(\mathcal{L})\right\} \leq d(\mathcal{P}) \leq \max\left\{k, \frac{1}{2}w_G(\mathcal{L})\right\} + 1\,.$$

In our proof of Theorem 2.7 we shall require a couple of lemmas, stated next. Their proof follows that of Theorem 2.7. The first lemma, stated here for $q = 2$, can be generalized to systems of linear equations mod $q \neq 2$ using the same proof strategy as below.

LEMMA 2.8. *Let $P_F(\ell_1), P_F(\ell_2)$ be the translations of two equivalent linear equations mod 2, denoted $\ell_1, \ell_2$, to a field $F$ with characteristic greater than 2. Then there is a PC derivation of $P_F(\ell_2)$ from $P_F(\ell_1) \cup \{y_i^2 - 1\}_{i=1}^n$ with degree at most $1 + \max\{d(P_F(\ell_1)), d(P_F(\ell_2))\}$.*

The next lemma can be proved either by use of the Laurent relation described in Buss *et al.* (2001) or by means of the Groebner Basis algorithm described in Clegg *et al.* (1996). We take the latter approach. In what follows, a *binomial refutation* is a refutation in which all polynomials are binomials.

LEMMA 2.9. *If $\mathcal{P}$ is a set of binomials, then there is a binomial refutation of $\mathcal{P}$ of minimal degree. Moreover, if all coefficients of $\mathcal{P}$ are in $\{1, -1\}$, then there is binomial refutation of $\mathcal{P}$ of minimal degree, in which all monomials have coefficients in $\{1, -1\}$.*

PROOF OF THEOREM 2.7. To see that $d(\mathcal{P}) \leq \lceil \max\{k, \frac{1}{2}w_G(\mathcal{L})\rceil\} + 1$, take any Gaussian refutation of width $w$, and let $d = \lceil w/2 \rceil + 1$. Without loss of generality a line $\ell$ in the refutation is equivalent to $\sum_{i\in S} x_i = b$ for some $S \subseteq \{1, \ldots, n\}, |S| \leq w$ and $b \in \{0, 1\}$. We shall now prove, by induction on the number of lines in the Gaussian refutation, that any polynomial $\prod_{i\in S_1} y_i - (-1)^b \prod_{i\in S_2} y_i$ with $S_1, S_2$ a partition of $S$ satisfying $|S_1|, |S_2| \leq \lceil w/2 \rceil$ has a PC derivation from $\mathcal{P}$ of degree at most $\max\{k + 1, d\}$.

The base case of $\ell \in \mathcal{L}$ follows from Lemma 2.8 because $P_F(\ell) \in \mathcal{P}$ is of degree at most $k$. Next, assume $\ell$ is the sum of the two lines $\ell_1 := \sum_{i\in S} x_i = b$ and $\ell_2 := \sum_{i\in T} x_i = b'$ and $\ell$ has the form $\sum_{i\in S\oplus T} x_i = (b + b' \pmod 2)$. Let $P_1, P_2$ be corresponding polynomials, such that $P_1$ is equivalent to $P_F(\ell_1)$ modulo $\{y_i^2 - 1\}_{i=1}^n$, and similarly for $P_2$. Define $R = S \cap T, r = |R|, s = |S - R|, t = |T - R|$. Note that $s + t = |S \oplus T| \leq w$.

Suppose $r \leq \lceil w/2 \rceil$. Assume wlog $s \leq t$. Let $T' \subseteq T - R$ be an arbitrary set of size $t' = \min\{t, \lceil w/2 \rceil\}$, and let $T'' = (T - R) - T'$ and $t'' = |T''|$. We

claim that $t'' + s \leq \lceil w/2 \rceil$. To see this notice that if $t' = \lceil w/2 \rceil$ then

$$t'' + s = (t - t') + s = t + s - \lceil w/2 \rceil \leq w - \lceil w/2 \rceil \leq \lceil w/2 \rceil .$$

On the other hand, if $t' = t < \lceil w/2 \rceil$ then $t'' = 0$ hence $t'' + s = s \leq \lceil w/2 \rceil$ by our assumption that $s \leq t$ and $s + t \leq w$. We have shown that $t'' + s \leq \lceil w/2 \rceil$ as claimed. Additionally, we have $r + t'' \leq \lceil w/2 \rceil$. To see this, assuming $t' = \lceil w/2 \rceil$ we have

$$r + t'' = |T| - t' \leq w - \lceil w/2 \rceil \leq \lceil w/2 \rceil ,$$

whereas assuming $t' = t < \lceil w/2 \rceil$ we get $t'' = 0$ so $r + t'' = r \leq \lceil w/2 \rceil$ with the last inequality following from the supposition made at the beginning of this paragraph.

Use $P_2$ to derive

$$P_3 := \prod_{i \in R \cup T''} y_i - (-1)^{b'} \prod_{i \in T'} y_i .$$

Use $P_1$ to derive

$$P_4 := \prod_{i \in R} y_i - (-1)^{b} \prod_{i \in S - R} y_i$$

and multiply $P_4$ by $\prod_{i \in T''} y_i$ to derive

$$P_5 := \prod_{i \in R \cup T''} y_i - (-1)^{b'} \prod_{i \in (S-R) \cup T''} y_i .$$

Subtract $P_3$ from $P_5$ to derive

$$P_6 := \prod_{i \in (S-R) \cup T''} y_i - (-1)^{bb'} \prod_{i \in T'} y_i .$$

Inspection reveals that the maximal degree of $\{P_1, \ldots, P_6\}$ is at most $d$, and by use of Lemma 2.8 this completes the proof of the case $r \leq \lceil w/2 \rceil$.

Next we deal with the case of $r > \lceil w/2 \rceil$. Let $R' \subset R$ be an arbitrary set of size $|R'| = \lceil w/2 \rceil$ and set $R'' = R - R'$. Use $P_1$ to derive

$$P_3 := \prod_{i \in R'} y_i - (-1)^{b} \prod_{i \in S - R'} y_i .$$

Use $P_2$ to derive

$$P_4 := \prod_{i \in R'} y_i - (-1)^{b'} \prod_{i \in T - R'} y_i .$$

Subtract the two to get the binomial

$$P_5 := \prod_{i \in S - R'} y_i - (-1)^{bb'} \prod_{i \in T - R'} y_i \,.$$

Notice that the polynomial modulo 2 corresponding to this polynomial is equivalent to the polynomial modulo 2 corresponding to the polynomial

$$P_6 := \prod_{i \in S - R} y_i - (-1)^{bb'} \prod_{i \in T - R} y_i \,.$$

Thus, by Lemma 2.8 we conclude $P_6$ can be derived from $P_5$ in degree at most $d$ as claimed.

In the other direction, assume $\mathcal{L}_p$ has a PC refutation of degree $d$, hence, by Lemma 2.9 it has a refutation in which each line is a multilinear binomial and all coefficients are either 1 or $-1$. We may construct inductively a corresponding Gaussian refutation with Gaussian width at most $2d$. To see this, notice that the PC refutation has no scalar multiplications by constants other than $-1$. A multiplication of $P$ by the variable $y_j$ corresponds to changing $\sum_{i \in S_1} x_i = b + \sum_{i \in S_2} x_i$ into $x_j + \sum_{i \in S_1} x_i = b + x_j + \sum_{i \in S_2} x_i$, and for any addition of $P_1$ and $P_2$, one of the monomials in $P_1, P_2$ must be identical (otherwise the output would not be a binomial), and hence adding the corresponding linear equations would yield the proper result.

If we get a single monomial by adding two binomials in the PC refutation, then both sides of the above polynomials involve the same sets $S_1, S_2$ of variables, but with different constants $b$. Then adding the corresponding linear equations will give us a nonzero constant equaling 0, and we have obtained a contradiction. Thus, $w_G(\mathcal{L}) \leq 2d(\mathcal{P})$ as claimed and the proof is complete. □

Proof of Lemma 2.8. $\ell_1, \ell_2$ are equivalent iff $\ell_1$ is $\sum_{i \in S_1} x_i = c + \sum_{i \in T_1} x_i$ and $\sum_{i \in S_2} x_i = c + \sum_{i \in T_2} x_i$ where $S_1 \oplus T_1 = S_2 \oplus T_2$ and $\oplus$ denotes symmetric difference. We argue the case of $S_1 = S_2 \cup \{x_1\}, T_2 = T_1 \cup \{x_1\}$. The full case then follows by induction. We have that $P_F(\ell_1)$ is

$$\prod_{i \in S_1} y_i - (-1)^c \cdot \prod_{i \in T_1} y_i = y_1 \prod_{i \in S_2} y_i - (-1)^c \cdot \prod_{i \in T_1} y_i \,,$$

where $1 \notin T_1$. We multiply the above equation by $y_1$ yielding

$$y_1^2 \prod_{i \in S_2} y_i - (-1)^c \cdot \prod_{i \in T_2} y_i \,.$$

Next, we multiply the axiom $y_1^2 - 1$ by $\prod_{i \in S_2} y_i$ and subtract this from the previous equation, yielding $P_F(\ell_2)$. Notice the degree of this derivation is at most $1 + \max\{d(P_F(\ell_1)), d(P_F(\ell_2))\}$ and this completes the proof. □

Proof of Lemma 2.9. In Clegg *et al.* (1996) it is shown that if a system of equations has a degree $d$ refutation, then such a refutation is generated by the Groebner basis algorithm (using any degree-respecting order on terms). Thus, it suffices to prove our claim for this particular refutation, which we do by induction on the number of steps in the Greobner basis refutation.

Inspection of the Groebner basis algorithm shows that the refutation so generated has the following property: each time a polynomial $P$ is generated as a linear combination of two previous polynomials $P = aP_1 + bP_2$, then $a = 1$ and $b$ is chosen such that some monomial in $P_1$ is canceled by this sum, meaning $P_1, P_2$ can be written respectively as $P_1 = m_1 + P_1', P_2 = m_2 + P_2$ where $bm_2 + m_1 = 0$. Thus, if $P_1, P_2$ are binomials with $\{-1, 1\}$-coefficients then so is $P$. Furthermore, multiplying a binomial with $\{-1, 1\}$-coefficients by a variable results in a binomial with $\{-1, 1\}$-coefficients. This completes the proof of the inductive step and shows that all lines in the Groebner refutation (which has minimal degree) are binomials with $\{-1, 1\}$-coefficients, as claimed. □

## 3. High expansion yields high degree

In this section we show that the Gaussian width and, by Theorem 2.7, the refutation-degree, of a system of linear equations is directly connected to its *expansion*, defined next. To define expansion we require the notion of a boundary.

Definition 3.1 (Boundary). *For $f$ a function and $x$ a variable, we say that $f$* depends *on $x$ if there is some assignment to all variables other than $x$, such that the value of $f$ is not fixed. let $Vars(f)$ be the set of all variables that $f$ depends on and let the* width *of $f$ be $|Vars(f)|$.*

*For $\mathcal{F}$ a set of functions, the* boundary *of $\mathcal{F}$, denoted $\partial\mathcal{F}$, is the set of variables $x$ such that there is* exactly one *function $f \in \mathcal{F}$ that depends on $x$.*

Definition 3.2 (Expansion). *For $\mathcal{F}$ an unsatisfiable set of functions, let $s$ be the minimal size of an unsatisfiable subset of $\mathcal{F}$. The* expansion *of $\mathcal{F}$ is*

$$e(\mathcal{F}) \triangleq \min\left\{ |\partial\mathcal{F}'| : \mathcal{F}' \subset \mathcal{F}, \quad \frac{s}{3} \leq |\mathcal{F}'| \leq \frac{2s}{3} \right\}.$$

An alternative, graph oriented, definition of expansion is as follows. Let $G$ be the bipartite graph with left vertex set $\mathcal{F}$, right vertex set $\mathrm{Vars}(\mathcal{F})$ and an edge connecting $f$ to $x$ if $f$ depends on $x$. Then $e(\mathcal{F})$ is the minimal size of the set of *unique-neighbors* of a "medium size" subset of the left hand vertices.

The following theorem is the main result of this section. The proof closely follows the width lower bound method of Ben-Sasson & Wigderson (2001).

THEOREM 3.3. *Let $\mathcal{L}$ be an unsatisfiable set of linear equations modulo 2 of width at most $k$ and let $F$ be a field of characteristic $p \neq 2$. Then,*

$$d\big(\mathcal{P}_F(\mathcal{L})\big) \geq \max\left\{k, \frac{1}{2}e(\mathcal{L})\right\}.$$

PROOF. We assume that $k < \frac{1}{2}e(\mathcal{L})$. By Theorem 2.7 we only have to prove that $w_G(\mathcal{L}) \geq e(\mathcal{L})$. Let $s$ be the minimal size of an unsatisfiable subset of $\mathcal{L}$. We say that a set of equations $\mathcal{L}' \subseteq \mathcal{L}$ *implies* an equation $\ell$ and denote this by $\mathcal{L}' \models \ell$ if and only if every assignment that satisfies $\mathcal{L}'$ also satisfies $\ell$. Define the following measure on linear equations:

$$\mu(\ell) \triangleq \min\{|\mathcal{L}'| : \mathcal{L}' \subseteq \mathcal{L},\ \mathcal{L}' \models \ell\}.$$

The following claims can be verified by inspection:

1. For $\ell \in \mathcal{L}$, $\mu(\ell) \leq 1$.
2. $\mu(1 = 0) = s$.
3. $\mu(\ell_1 + \ell_2) \leq \mu(\ell_1) + \mu(\ell_2)$.
   This is because if $\mathcal{L}_1 \models \ell_1,\ \mathcal{L}_2 \models \ell_2$, then clearly $\mathcal{L}_1 \cup \mathcal{L}_2 \models \ell_1 + \ell_2$.

Hence, in every Gaussian refutation there must be a line $\ell$ for which $\frac{s}{3} \leq \mu(\ell) \leq \frac{2s}{3}$. Fix such a line $\ell$, and let $\mathcal{L}'$ be a minimal subset of size $\mu(\ell)$ implying $\ell$. We claim that $\ell$ depends on every variable in $\partial\mathcal{L}'$. Assume this is not the case, i.e. assume $x_i \in \partial\mathcal{L}' \setminus Vars(\ell)$, and let $\ell'$ be the unique equation in $\mathcal{L}'$ which depends on $x_i$. By the minimality of $\mathcal{L}'$ there is some assignment $\alpha$ that satisfies all of $\mathcal{L}' \setminus \{\ell'\}$ but falsifies $\ell'$ and $\ell$. One can flip the assignment of $\alpha$ on $x_i$ so as to satisfy $\ell'$, without affecting $\ell$ or the rest of $\mathcal{L}'$, thus $\mathcal{L}' \not\models \ell$. Contradiction. □

With the aid of Theorem 3.3 we can derive the lower bounds of Buss *et al.* (2001) on the refutation degree of Tseitin contradictions and related formulas. As an example we study the case of Tseitin contradictions and prove Theorem 1.4.

PROOF OF THEOREM 1.4. First we notice as in Tseitin (1968); Urquhart (1987) that if $G$ is connected then $\mathcal{T}(G,b)$ is minimal unsatisfiable, i.e., removing any single constraint from it makes the residual system satisfiable. Next, we argue that the expansion of $\mathcal{T}(G,b)$ as per Definition 3.2 coincides with the definition of the term in Theorem 1.4. Indeed, $S \subset V$ has $t$ edges going from $S$ to $V \setminus S$ if and only if the set of constraints

$$\sum_{e \ni v} x_e \equiv b(v) \pmod 2, \quad v \in S$$

has a boundary of size $t$, because for each crossing edge $e=(v,u), v \in S, u \in V \setminus S$ the variable $x_e$ appears in exactly one constraint, namely, the constraint over vertex $v$. Now that we have shown that the two definitions, in Definition 3.2 and in Theorem 1.4, of the term "expansion" coincide, we apply Theorem 3.3 and complete our proof. □

## 4. Random $k$-CNF's

In this section we prove Theorem 1.2 and show that PC is not much better than Resolution in refuting random $k$-CNF's. Our proof goes by reducing the refutation-degree of a random CNF to the refutation-degree of a closely related random set of linear equations.

**4.1. Proof of Theorem 1.2.** The standard formulation of a clause $C$ as a polynomial is $P_C := \prod_{i \in \text{positive}}(1-x_i) \cdot \prod_{i \in \text{negative}} x_i$, where positive is the set of positive literals appearing in $C$ (the set negative is similarly defined). A CNF formula $\mathcal{C} = \{C_i\}_{i=1}^m$, formulated in polynomial calculus, is the following set of polynomials $\{P_{C_i}\}_{i=1}^m \cup \{x_i^2 - x_i\}_{i=1}^n$. From now on, we shall think of all CNF formulas as formulated in this manner.

To connect random CNFs to random linear equations we use a different method for producing the distribution $\mathcal{F}_k^{n,\Delta}$ over random $k$-CNFs from Definition 1.1.

DEFINITION 4.1 (Random linear equations). *Let $\mathcal{L} \sim \mathcal{L}_k^{n,\Delta}$ denote that $\mathcal{L}$ is a random set of linear equations mod 2, each having $k$ variables, chosen at random by picking $\Delta n$ equation i.i.d from the set of all $2\binom{n}{k}$ equations, with repetitions.*

Given $\mathcal{L}_k^{n,\Delta}$, one can produce $\mathcal{F}_k^{n,\Delta}$ from Definition 1.1 by replacing each $\ell \in \mathcal{L}$ with one of its $2^{k-1}$ defining clauses (as defined before Definition 1.3) picked at random, with equal probability. Let us denote by $\mathcal{C}_\mathcal{L}$ any CNF formula

derived from $\mathcal{L}$ by this process. Notice that $\mathcal{L} \models \mathcal{C}_{\mathcal{L}}$, but the converse is not necessarily true. In the other direction, for any clause $C$ there is a unique linear equation mod 2 $\ell_C$ such that $\ell_C$ depends on all variables of $C$ and $\ell_C \models C$. For $\mathcal{C}$ a CNF formula, let $\mathcal{L}_{\mathcal{C}} = \{\ell_C : C \in \mathcal{C}\}$ be the (unique) *linear closure* of $\mathcal{C}$. This gives a way to produce $\mathcal{L}_k^{n,\Delta}$ from $\mathcal{F}_k^{n,\Delta}$.

The following theorem is the main observation of this section. It says that taking the linear closure of a unsatisfiable CNF which has high expansion, cannot help decrease the refutation-degree substantially.

THEOREM 4.2. *For $\mathcal{C}$ a $k$-CNF, and $\mathcal{P}_F(\mathcal{L}_{\mathcal{C}})$ the formulation of the linear closure of $\mathcal{C}$ as a set of polynomials in the field $F$, of characteristic greater than 2, we have $d(\mathcal{C}) \geq \max\{d(\mathcal{P}_F(\mathcal{L}_{\mathcal{C}})), k+1\}$.*

The following lemma appears originally in Chvátal & Szemerédi (1988) for CNF formulas. It is proved using what is by now a standard union bound and applies to sets of linear equations just as well (see, e.g. (Ben-Sasson & Wigderson 2001, Section 6.3)).

LEMMA 4.3. *For integer $k \geq 3$ and $\Delta > 0$ there exists a constant $\kappa = \kappa(k, \Delta) > 0$ such that*

$$\Pr_{\mathcal{L} \sim \mathcal{L}_{\mathrm{k}}^{\mathrm{n},\Delta}}[\mathrm{e}(\mathcal{L}) \geq \kappa \mathrm{n}] = 1 - \mathrm{o}(1)\,.$$

PROOF OF THEOREM 1.2. Lemma 4.3 claims that with high probability a random set of linear equations $\mathcal{L}$ has expansion that is linear in the number of underlying variables. Theorem 3.3 implies the refutation-degree of a random set of linear equations is with high probability linear in $n$, as long as the underlying field has characteristic $> 2$. Theorem 4.2 completes the proof by showing that with high probability a random CNF requires linear refutation-degree. □

**4.2. Proof of Theorem 4.2.** The following definition and lemma from Buss *et al.* (2001) will assist us in our reduction of random CNFs to random systems of linear equations.

DEFINITION 4.4 (Buss *et al.* 2001). *Let $\mathcal{P}(\overline{x})$, $\mathcal{Q}(\overline{y})$ be two sets of polynomials over a field $F$. Then $\mathcal{P}(\overline{x})$ is $(d_1, d_2)$-reducible to $\mathcal{Q}(\overline{y})$ if there exists a degree $d_1$ polynomial $r_i(\overline{x})$ for every $y_i$ such that $\mathcal{Q}(\overline{r}(\overline{x}))$ has a PC derivation of degree $d_2$ from $\mathcal{P}(\overline{x})$.*

LEMMA 4.5 (Buss *et al.* 2001). *Suppose that $\mathcal{P}(x)$ is $(d_1, d_2)$-reducible to $\mathcal{Q}(y)$. Then if there is a degree $d_3$ PC refutation of $\mathcal{Q}(y)$, then there is a degree $max(d_2, d_3 d_1)$ PC refutation of $\mathcal{P}(x)$.*

Our proof of Theorem 4.2 also requires the following lemma.

LEMMA 4.6. *Let $\ell$ be a linear equation mod 2, over the variables $x_1, \ldots, x_k$, and let $\mathcal{P} = P_F(\ell) \cup \{y_i^2 - 1\}_{i=1}^k$ be its formulation as a set of polynomials over a field $F$ of characteristic $\neq 2$. Let $C$ be a clause over $x_1, \ldots, x_k$ such that $\ell \models C$. $\mathcal{P}$ is $(1, k+1)$-reducible to $P_C$.*

PROOF. Set $x_i = (1 + y_i)/2$. From this definition and $y_i^2 - 1$ one can deduce $x_i^2 - x_i$. PC is implicationally complete, thus, there is a derivation of $C$ from $\mathcal{P}$, using only variables that appear either in $\mathcal{P}$ or in $C$. This derivation can be assumed to have degree at most $k + 1$, because every time a monomial with degree $k + 1$ appears in the derivation, some variable $x$ has degree 2, and the monomial can be reduced to degree $k$ by use of the previously derived polynomials $x_i^2 - x_i$. □

We conclude this section with the proof of Theorem 4.2.

PROOF OF THEOREM 4.2. Let $\bar{\mathcal{C}}_\ell$ be the $k$-CNF definition of the linear equation $\ell$, and let $\bar{\mathcal{C}}_\mathcal{L}$ be the conjunction of the $\mathcal{C}_\ell$'s, for all $\ell \in \mathcal{L}$ (formulated as a set of polynomials). Clearly, $\mathcal{C} \subseteq \bar{\mathcal{C}}_\mathcal{L}$, hence any refutation of $\mathcal{C}$ is also a refutation of $\bar{\mathcal{C}}_\mathcal{L}$. The previous Lemma 4.6 implies that $\mathcal{P}_F(\mathcal{L}_\mathcal{C})$ is $(1, k+1)$-reducible to $\bar{\mathcal{C}}_\mathcal{L}$, and the theorem follows. □

## 5. Upper bounds on refutations of linear equations

In this section we prove nontrivial upper bounds on the complexity of refuting unsatisfiable systems of linear equations modulo 2 in Resolution and the Polynomial Calculus. All upper bounds stated below follow directly from the following theorem, the proof of which constitutes the bulk of this section.

THEOREM 5.1 (Upper bounds on Gaussian width). *For $\mathcal{L}$ an unsatisfiable system of linear equations modulo 2 of width $k = O(1)$ over $n$ variables,*

$$w_G(\mathcal{L}) \leq n/2 + o(n)\,.$$

*Moreover, there exists a refutation of size at most $n$ that achieves this width.*

Let $S_R(\mathcal{F}), w_R(\mathcal{F})$ respectively denote the minimal *size*, *width*, of a Resolution refutation of the CNF $\mathcal{F}$. The previous result and its proof have the following implications:

COROLLARY 5.2. *For $\mathcal{F}(\mathcal{L})$ the CNF formulation of a system of width $O(1)$ linear equations mod 2,*

$$w_R\big(\mathcal{F}(\mathcal{L})\big) \leq n/2 + o(n)\,.$$

COROLLARY 5.3. *For $\mathcal{F}(\mathcal{L})$ the CNF formulation of a system of width $O(1)$ linear equations mod 2,*

$$S_R\big(\mathcal{F}(\mathcal{L})\big) \leq 2^{n/2+o(n)}\,.$$

COROLLARY 5.4. *For $\mathcal{P}_p$ a system of linear equations mod 2, in the field $F_p$, $p > 2$,*

$$d(\mathcal{P}_p) \leq n/4 + o(n)\,.$$

The corollaries mentioned above show that for tautologies expressing systems of linear equations modulo 2, proof complexity can be significantly less than that of exhaustive search refutation. These results contrast with the work of Pudlák & Impagliazzo (2000) which showed that tree-like resolution refutations of such systems can require size $2^{(1-\epsilon)n}$ for arbitrary $\epsilon > 0$. Hence, resolution based methods with more memory could be faster than DLL algorithms for larger clause sizes.

PROOF OF THEOREM 5.1. Let the equations be $\sum_{i\in S_j} x_i = b_j$ for $j = 1, \ldots, m$, with$|S_j| \leq k$. Assume without loss of generality that the equations are minimally dependent, i.e., that $\oplus_j S_j = \emptyset$ and $\sum_j b_j = 1 \pmod 2$. Let $\pi$ be a random permutation of $1, \ldots, m$, and consider the refutation obtained by always adding the equations one by one in the order $\pi(1), \ldots, \pi(m)$. We claim that with high probability the width of this refutation, which by construction has size $m \leq n$, is at most $n/2 + o(n)$. Let $1 \leq j \leq m$ ; we will show that with high probability, the width of the $j$'th line, $W_j = |\oplus_{j'\leq j} S_{\pi(j')}| \leq n/2+o(n)$. Let $q = j/n$. Consider the random experiment of picking equations with probability $q$ independently and summing the picked equations to get $\sum_{i\in S'} x_i = b'$ for some set $S'$ and bit $b'$. Let $C$ be the number of chosen equations. Then if we subsequently subtract $C - j$ chosen equations if $C > j$ or add $j - C$ random unchosen equations otherwise, the result is distributed as the $j$'th equation. The difference in widths of the two processes is bounded by $|C - j|k$. Since the expectation of $C$ is $j$ and it is normally distributed, $\Pr[|C-j| \geq \alpha n^{1/2}] \leq e^{-\alpha^2/2}$. Thus, we concentrate on the width of $S'$.

Let $t_i$ be the number of equations $x_i$ appears in; since the sum of equations gives $0 = 1$, $t_i$ is even. Let $Bad = \{i|t_i \geq kn^{1/4}\}$. Since $m < n$, we know that

$|Bad| \leq n^{3/4}$. Look at the random indicator variable $R_i$ which is 1 if $x_i \in S'$. Let $R = \sum_{i \notin Bad} R_i$ be the number of "good" elements in $S'$.

$$\begin{aligned}\Pr(R_i = 0) - \Pr(R_i = 1) &= \sum_{l=0}^{l=t_i} (-1)^l \binom{t_i}{l} q^l (1-q)^{t_i - l} \\ &= (1-2q)^{t_i} > 0\,.\end{aligned}$$

This implies $E(R_i) \leq 1/2$ for every $i$. Additionally,

$$\begin{aligned}V(R) &= E(R^2) - \big(E(R)^2\big) \\ &= \sum_{i,i' \notin Bad} \big(E(R_i R_{i'}) - E(R_i)E(R_{i'})\big) \\ &\leq \sum_{i,i' \notin Bad \text{ and } \exists c \text{ such that } i,i' \in S_c} \big(E(R_i R_{i'})E(R_i)E(R_{i'})\big) \\ &\leq kn^{5/4}\,.\end{aligned}$$

So,

$$\begin{aligned}\Pr(R \geq n/2 + kn^{3/4}) &\leq \Pr\left(|R - E(R)| \geq kn^{3/4}\right) \\ &\leq kn^{5/4}/(k^2 n^{3/2}) \leq n^{-1/4}\,.\end{aligned}$$

Hence, with probability $1 - 2n^{-1/4}$,

$$\begin{aligned}W_j &\leq R + |Bad| + |C - j|k \\ &\leq n/2 + kn^{3/4} + n^{3/4} + n^{1/2}\ln n \\ &= n/2 + O(kn^{3/4})\,.\end{aligned}$$

We only obtained a probability $1 - O(n^{-1/4})$ bound on the above probability. However, this suffices to show that with probability 3/4, each $j$ which is a multiple of $Dn^{3/4}$ for a sufficiently large constant $D$ has $W_j \leq n/2 + O(kn^{3/4})$. However, if this is true, then since $W_j \leq W_{j'} + k|j - j'|$, the same is true for any $j$, using $j'$ the nearest multiple of $Dn^{3/4}$.

Note that the process of adding the lines one by one ensures that the set of variables for each new equation involves at most $k$ variables that do not appear in the old one. This is used to prove Corollary 5.3, since then the derivation of the $j$'th line in the Gaussian refutation is logically valid and depends on at most $W_j + k$ variables. It thus follows from the implicational completeness of resolution that the line can be simulated by a resolution refutation of size $2^{W_j + k}$. □

## Acknowledgements

We thank Avi Wigderson and the late Mikhail Alekhnovich for many helpful discussions. We are grateful to the anonymous referees for comments that improved the presentation of the paper. An initial version of this paper was presented in the 40th Annual Symposium on Foundations of Computer Science (FOCS) in 1999. Research of first co-author supported by grants from the Israeli Science Foundation and the US-Israel Binational Science Foundation. Research of second co-author supported by NSF Award CCR-9734911, Sloan Research Fellowship BR-3311, grant #93025 of the joint US-Czechoslovak Science and Technology Program, and USA-Israel BSF Grant 97-00188.

## References

M. Alekhnovich & A. Razborov (2003). Lower bounds for polynomial calculus: Non-binomial case. *Proceedings of the Steklov Institute of Mathematics* **242**, 18–35.

P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi & P. Pudlák (1994). Lower bounds on Hilbert's Nullstellensatz and propositional proofs. In *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science, FOCS'94 (Santa Fe, New Mexico, November 20–22, 1994)*, 794–806. IEEE Computer Society Press, Los Alamitos-Washington-Brussels-Tokyo.

E. Ben-Sasson & A. Wigderson (2001). Short proofs are narrow–resolution made simple. *Journal of the ACM* **48**(2), 149–169. ISSN 0004-5411. URL `http://www.acm.org/pubs/citations/journals/jacm/2001-48-2/p149-ben-sasson/`.

S. Buss, D. Grigoriev, R. Impagliazzo & T. Pitassi (2001). Linear gaps between degrees for the polynomial calculus modulo distinct primes. *J. Comput. Syst. Sci.* **62**(2), 267–289.

V. Chvátal & E. Szemerédi (1988). Many hard examples for resolution. *Journal of the ACM* **35**(4), 759–768.

M. Clegg, J. Edmonds & R. Impagliazzo (1996). Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, 174–183. Philadelphia, Pennsylvania.

A. Haken (1985). The intractability of resolution. *Theoretical Computer Science* **39**(2–3), 297–308.

R. Impagliazzo, P. Pudlák & J. Sgall (1999a). Lower bounds for the polynomial calculus and the Gröbner basis algorithm. *Computational Complexity* **8**(2), 127–144.

R. Impagliazzo, P. Pudlák & J. Sgall (1999b). Lower bounds for the polynomial calculus and the Gröbner basis algorithm. *Computational Complexity* **8**(2), 127–144.

T. Pitassi (1996). Algebraic propositional proof systems. In *Descriptive Complexity and Finite Models, Proceedings of a DIMACS Workshop, January 14–17, 1996, Princeton University*, N. Immerman & P. G. Kolaitis, editors, volume 31 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 215–244. American Mathematical Society. ISBN 0-8218-0517-7.

P. Pudlák & R. Impagliazzo (2000). A lower bound for DLL algorithms for k-SAT (preliminary version). In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, 128–136. ACM Press, N.Y.

A. A. Razborov (1998). Lower bounds for the polynomial calculus. *Computational Complexity* **7**(4), 291–324. URL `http://link.springer.de/link/service/journals/00037/bibs/8007004/80070291.htm`.

G. S. Tseitin (1968). On the complexity of derivation in the propositional calculus. *Zapiski nauchnykh seminarov LOMI* **8**, 234–259. English translation of this volume: Consultants Bureau, N.Y., 1970, pp. 115–125.

A. Urquhart (1987). Hard examples for resolution. *Journal of the ACM* **34**(1), 209–219.

Manuscript received 20 October 2008

Eli Ben-Sasson
Department of Computer Science
Technion – Israel Institute of Technology
Haifa, 32000, Israel
`eli@cs.technion.ac.il`

Russell Impagliazzo
School of Mathematics
Institute for Advanced Study
Einstein Drive
Princeton, NJ 08540, USA
and
Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093-0114, USA
`russell@cs.ucsd.edu`