Linear Gaps Between Degrees for the Polynomial Calculus Modulo Distinct Primes

Sam Buss[1,2]
Department of Mathematics
Univ. of Calif., San Diego
La Jolla, CA 92093-0112
sbuss@ucsd.edu

Dima Grigoriev
IMR Universite Rennes-1
Beaulieu 35042 Rennes, France
dima@maths.univ-rennes1.fr

Russell Impagliazzo[1,3]
Computer Science and Engineering
Univ. of Calif., San Diego
La Jolla, CA 92093-0114
russell@cs.ucsd.edu

Toniann Pitassi[1,4]
Computer Science
University of Arizona
Tucson, AZ 85721-0077
toni@cs.arizona.edu

Abstract

This paper gives nearly optimal lower bounds on the minimum degree of polynomial calculus refutations of Tseitin's graph tautologies and the mod p counting principles, $p \geq 2$. The lower bounds apply to the polynomial calculus over fields or rings. These are the first linear lower bounds for the polynomial calculus for k-CNF formulas. As a consequence, it follows that the Gröbner basis algorithm, used as a heuristic for k-SAT, requires exponential time in the worst-case. Moreover, our lower bounds distinguish linearly between proofs over fields of characteristic q and r, $q \neq r$, and more generally distinguish linearly the rings Z_q and Z_r where q and r do not have the identical prime factors.

[1]Supported in part by international grant INT-9600919/ME-103 from the NSF (USA) and the MŠMT (Czech republic)

[2]Supported in part by NSF grant DMS-9803515

[3]Supported in part by NSF grant CCR-9734911, Sloan Research Fellowship BR-3311, and US-Israel BSF grant 97-00188.

[4]Supported in part by NSF grant CCR-9457783 and US-Israel BSF grant 95-00238.

1 Introduction

The problem of recognizing when a proposition formula is a tautology is dual to the satisfiability problem and is therefore central to computer science. A principal method of establishing that a formula is a tautology is to find a proof of it in a formal system such as resolution or (extended) Frege systems. In fact, many algorithms for establishing propositional validity are essentially a search for a proof in a particular formal system. In recent years, several algebraic proof systems, including the Nullstellensatz system and the polynomial calculus (also called the 'Gröbner' system) have been proposed: these systems are motivated in part by the desire to identify powerful proof systems which support efficient search algorithms and in part by the desire to extend lower bounds on proposition proof complexity to stronger proof systems.

The Nullstellensatz proof system is a propositional proof system based on Hilbert's Nullstellensatz and was introduced in [2]. The polynomial calculus (PC) is a stronger propositional proof system introduced first by [8]. (See [15] and [6] for subsequent, more general treatments of algebraic proof systems.) In the polynomial calculus, one begins with an initial set of polynomials and the goal is to prove that they cannot be simultaneously equal to zero over a field F. A polynomial calculus (PC) derivation of P_l from a set of polynomials Q is a sequence of polynomials $P_1, \ldots, P_l$ such that each polynomial is either an initial polynomial from Q, or follows from one of the following two rules: (i) If P_i and P_j are previous polynomials, then $cP_i + dP_j$ can be derived, where $c, d \in F$; (ii) if P_i is a previous polynomial and x is a variable, then xP_i can be derived. The *degree* of a PC derivation is the maximum degree of the P_i's, and its *size* is the sum over all i of the number of (non-zero) monomials in P_i. We identify polynomials P_i with the equations $P_i = 0$ and a PC refutation of Q (a proof that the equations $Q = 0$ are not solvable over F) is simply a PC derivation of 1 from Q (i.e., of $1 = 0$). For Boolean tautologies, we usually consider sets Q that contain the polynomials, $x_i^2 - x_i$ for each variable x_i, corresponding to the axiom that each variable is either 0 or 1. However, in this paper, we will be using non-Boolean systems as a tool for reasoning about Boolean systems, so it will be useful to also consider proofs without these axioms. For non-Boolean systems it is often desirable to use an algebraicly closed field F so as to ensure the completeness of the polynomial calculus.

The PC *degree* and *size* of the system of inconsistent polynomials Q is the minimum degree, resp. size, of any PC refutation of the system. While size might be considered the more natural measure, it is tightly correlated with degree. For example, [8] show that a constant degree lower bound implies a polynomial size lower bound, whereas [12] show that a linear degree lower bound implies an exponential size lower bound. In fact, the latter result is also true for size in the PCR system which can simulate both resolution and polynomial calculus refutations ([1]). So our results also imply the (known) lower bounds for resolution proofs of the Tseitin tautologies ([18]). In fact, the relationship is somewhat stronger, as we mention in the concluding comments.

The definition of the polynomial calculus depends implicitly on the choice of a field F such that all polynomials are over the field F. A number of authors also consider the polynomial calculus over rings ([6, 5]). The only difference in the definition of the PC system is that a PC refutation over a ring is a derivation of r (i.e., of $r = 0$) for some non-zero r in the ring. Our main results apply to both fields and rings.

The mod p counting principle can be formulated as a set MOD_p^n of constant-degree polynomials expressing the negation of the counting principle, and the present paper gives tight lower bounds on the degree of polynomial calculus refutations of MOD_p^n over fields of characteristic $q \neq p$. (Our bounds are $\Theta(n)$, but this is really sub-linear, since the MOD_p^n principle has a number of variables which is a polynomial in n.) However, we obtain our bounds via a reduction from variants of the Tseitin graph tautologies for constant degree expanders. (The variants coincide with the normal Tseitin graph tautologies when $p = 2$, and we are proving a lower bound for a field of odd characteristic. The variation generalizes these tautologies to allow us to prove lower bounds for MOD_p^n for $p \neq 2$ over any field of characteristic other than p. Because the $p = 2$ case is somewhat simpler, we describe it separately.) For these tautologies, which can be expressed as the negation of k-CNF's for a constant k depending on the degree of the graph, we obtain a truly linear lower bound. As a consequence, it follows that the Gröbner basis algorithm is worst-case exponential time when used as a heuristic for k-SAT in the manner suggested by [8].

Some lower bounds on the degree of Nullstellensatz proofs of the mod p counting principles have been given in prior work: [2] gave non-constant lower bounds and [6] gave lower bounds of the form n^ϵ. For the polynomial calculus, the best lower bound on the degree of PC refutations of MOD_p^n was Krajíček's $\Omega(\log\log n)$ lower bound based on a general lower bound for symmetrically specified polynomials [13].

Polynomial calculus lower bounds have been obtained for other families of tautologies. Razborov [16] established $\sqrt{n}$ lower bounds on the degree of polynomial calculus proofs of the pigeon-hole principle, as well as a linear

$\Omega(n)$ lower bound on a variant of this principle.[6]

Recently, Grigoriev [9] succeeded in giving very simple linear lower bounds on the degree of Nullstellensatz refutations of the Tseitin mod 2 graph tautologies. The present work is motivated by this paper, and in particular by the idea of working in the Fourier basis which greatly simplifies the argument. More precisely, by performing a simple linear transformation on the variables, we transform the Boolean clauses of the graph tautologies into an equivalent (in terms of refutation degree) system of non-Boolean binomials. We then give a normal form for polynomial calculus derivations involving only binomials, which allows us to prove the degree lower bound for the transformed system. Finally, we reduce the graph tautologies to MOD_p^n, thus proving the same lower bound for the modular counting principles. Our main contribution may be that, once the tautologies are reduced to binomials, our arguments are much simpler than for previous lower bounds.

The present paper establishes linear (in n) lower bounds to the degree of polynomial calculus over a field of characteristic $q \nmid p$, for any PC refutation of the MOD_p^n polynomials. In section 8 we generalize this linear lower bound to the polynomial calculus over rings Z_q provided p and q are relatively prime. As it is well-known to be easy to give constant degree polynomial calculus (and even Nullstellensatz) refutations of the MOD_p^n polynomials over F_p, our results imply that the MOD_p^n polynomials have a linear gap between proof complexity for the polynomial calculus over F_p and over F_q. This is the first separation of more than $\log\log n$.

When considering the polynomial calculus as a tool for analyzing heuristics for k-SAT based on polynomial reasoning, such as the Gröbner basis algorithm [8], it is interesting to consider size bounds for polynomial calculus proofs for tautologies expressible as k-CNF's. By the results of [12], an exponential lower bound on size for a Boolean tautology follows from a degree bound that is linear in the number of variables. Unfortunately, the MOD_p^n polynomials are not expressible themselves as k-CNF's, and have a number of variables that is polynomial in n. However, our lower bounds for the MOD_p^n polynomials are obtained via a reduction from variants of the Tseitin Graph Tautologies for constant degree expanders. These tautologies are expressible as k-CNF's and have only a linear number of variables in n. Thus, we immediately obtain an exponential lower bound for the size of any polynomial calculus proof of these tautologies, and hence show that the

[6] We would like to take this opportunity to correct an embarrassing mistake that appeared in the conference version of this paper. Razbarov's result is the first linear lower bound for the polynomial calculus. However, we believe our result is the first such result for a tautology in k-CNF form or a system of bounded degree polynomials.

Gröbner basis algorithm requires worst-case exponential time when used as a heuristic for k-SAT.

It follows from a result of Krajíček [14] that our linear lower bounds on the degree of PC refutations imply exponential lower bounds of $AC^0[q]$-Frege proofs of the mod p principles when Mod-q gates are present only at the top (root) of formulas.

A final motivation for our work, and perhaps the most significant, is that our proof technique is much simpler than previous lower bounds for the polynomial calculus. This core simplicity gets lost for the most general statement of our results. Therefore, we first prove our lower bound for the special case of the original Tseitin graph tautologies over a field of odd characteristic, and only later state the result in full generality.

2 Tseitin tautologies: Boolean and polynomial versions

Tseitin's (mod 2) graph tautologies are based on the following idea. Let G_n be a connected undirected graph on n vertices, where each node in the graph has an associated charge of either 0 or 1, and where the total sum of the charges is odd. Then it is impossible to choose a subset of the edges E' from E so that for every vertex $v \in V$, the number of E'-edges incident to v is equal mod 2 to the charge of v. This impossibility follows from a simple parity argument, since summing the degrees of all vertices in the subgraph counts each edge twice, and so is even, whereas it should also be the sum of all the charges, which is odd.

For an r-regular graph G_n with n odd, and charges all 1, we can express this principle as the inconsistency of the following system of polynomials over a finite field F of characteristic different from 2: There will be $rn/2$ underlying variables, one for each edge of G_n. We will denote the variable corresponding to the edge $e = \{i, j\}$ from i to j by $y_e = y_{\{i,j\}}$. For each variable y_e, we have the equation $y_e^2 - 1 = 0$; this forces the variables to take on values of either 1 or -1, with $y_e = -1$ corresponding to the presence of e in the subgraph E'. Secondly, corresponding to each vertex i in G_n, we will have the equation $1 + y_{\{i,j_1\}} y_{\{i,j_\}} \cdots y_{\{i,j_r\}} = 0$, where $j_1, \ldots, j_r$ are the neighbors of i in G_n. This corresponds to saying that the degree of i in the subgraph E' is odd. This set of equations, representing the Tseitin mod 2 graph formula, will be denoted by $TS_n(2)$.

For any prime p, we can generalize the above principle to obtain a mod p version as follows. Again, we fix an underlying r-regular, undirected graph G_n, and then let G'_n be the corresponding directed graph where each undirected edge is replaced by two directed edges. Each vertex in G'_n will have an associated label, or charge in $[0, p-1]$ such that the sum of the vertex charges is congruent to 1 mod p. The mod p principle states that it is impossible to assign values in $[0, p-1]$ to each of the directed edges so that: (i) for any pair of complementary edges $\langle i, j\rangle$ and $\langle j, i\rangle$, $v(\langle i, j\rangle) + v(\langle j, i\rangle) \equiv 0 \pmod p$, and (ii) for every vertex i, the sum of the edge values coming out of vertex i is congruent to the charge of that vertex mod p. Again, this is impossible since if we sum the edges in pairs, we obtain 0 mod p, but summing them by vertices gives the total charge of 1 mod p.

Let F be a finite field with characteristic $q \neq p$ that contains a primitive p-th root of unity ω. Assume all charges of vertices are 1, and that $n \equiv 1 \pmod p$. We can express the mod p Tseitin principle for G'_n as the unsatisfiability of the following system of polynomials over F: We have rn underlying variables y_e, one for every directed edge e. For each variable y_e we have the equation $y_e^p - 1 = 0$; this forces variables to take on values in $1, \omega, \omega^2, \ldots, \omega^{p-1}$. (The power of ω corresponds to the value assigned to e.) Secondly, for each vertex i in G'_n, we will have the equation $y_{\langle i,j_1\rangle} y_{\langle i,j_2\rangle}, \ldots, y_{\langle i,j_r\rangle} - \omega = 0$, where $j_1, \ldots, j_r$ are the neighbors of i. Third, for each edge $e = \langle i, j\rangle$ we have the equation $y_{\langle i,j\rangle} y_{\langle j,i\rangle} - 1 = 0$. This set of equations, representing the Tseitin mod p formula, will be denoted by $TS_n(p)$.

Alternatively, we can express the above tautologies as Boolean formulas. In the graph case, we have one Boolean variable x_e for each edge, determining whether $e \in E'$, and we express the constraint that the degree of a node i is odd as a CNF formula, a system of 2^{r-1} clauses of size r. For the general case, we would have one Boolean variable $x_{e,i}$ for each directed edge e and residue mod p, i, interpreted as an indicator variable for the event that edge e has charge i. We would have constraints that say each edge has one charge, that an edge's charge is the negation of that of the reversed edge, and that the sum of charges of edges incident to a given node sum to 1 mod p. Each of these constraints involve at most $\max\{r, p\}$ variables and the tautology can be expressed as a k-CNF for some constant k. We call these the Boolean version of the tautologies. The Boolean version is also a system of polynomials in the usual way, writing negation as $1 - x_{e,i}$, "or" as the product of the negations, and including the axioms $x_{e,i}^2 - x_{e,i}$. We let $BTS_{n,p}$ denote this system of polynomials.

3 The mod p principle and low degree reductions

A related principle is the mod p counting principle. Intuitively, it states that it is not possible to partition a set of size n into groups of size p, if n is congruent to 1 mod p. We will express this by polynomial equations as follows. The underlying variables are x_e, where e ranges over all p element subsets of $[1, n]$. The degree 2 equations expressing the negation of the principle are: (1) $x_e^2 - x_e = 0$ for each e; (2) $x_e x_f = 0$, for each e, f such that $e \cap f \neq \emptyset$ and $e \neq f$; (3) $1 - \sum_{e, i \in e} x_e = 0$, for each $i \in [1, n]$. Let the above set of equations be denoted by MOD_p^n.

We want to show that a low degree PC refutation of the mod p counting principle implies a low degree PC refutation of the Tseitin mod p graph equations. To do this, we define the following general notion of a low degree reduction.

Definition. *Let $P(\overline{x})$, $Q(\overline{y})$ be two sets of polynomials over a field F. Then P is (d_1, d_2)-reducible to Q if: (1) For every y_i, there is a degree d_1 definition of y_i in terms of the x's. That is, for every i, there exists a degree d_1 polynomial r_i where y_i will be viewed as being defined by $r_i(x_1, \ldots, x_n)$; (2) there exists a degree d_2 PC derivation of the polynomials $Q(\overline{r}(x_1, \ldots, x_n))$ from the polynomials $P(\overline{x})$.*

Lemma 1 *Suppose that $P(x)$ is (d_1, d_2)-reducible to $Q(y)$. Then if there is a degree d_3 PC refutation of $Q(y)$, then there is a degree $max(d_2, d_3 d_1)$ PC refutation of $P(x)$.*

With this definition, it is easy to see that the polynomial and Boolean versions of the Tseitin tautologies are equivalent as far as their degrees go.

Lemma 2 *For all n and p, and for any field F of characteristic q, where $q \nmid p$, and F includes the primitive p-th root of unity, $BTS_{n,p}$ is $(1, pr)$-reducible to $TS_{n,p}$.*

Proof of Lemma 2. For the graph case, we define $y_e = 1 - 2x_e$. In the general case, we define $y_e = \sum_i x_{e,i} \omega^i$. Proofs of degree pr of the corresponding constraints follow from the completeness of polynomial calculus, and the fact that each constraint involves at most pr variables. □

We can also prove an analog for the usual reduction from k-SAT to 3-SAT:

Lemma 3 *Given any unsatisfiable k-CNF formula Φ on n variables with m clauses, we can construct an unsatisfiable 3-CNF formula Ψ with $n + km$ variables, so that Φ is (k, k) reducible to Ψ.*

Proof of Lemma 3. We start with the variables of Ψ. For each clause of Φ, $x_1 \vee x_2 \vee .. \vee x_k$, Ψ will also have $k-2$ new variables, $z_1, ..z_{k-2}$ and clauses that represent $z_1 = x_1 \vee x_2$, $z_i = z_{i-1} \vee x_{i+1}$ for $i = 2, ..k-2$ and $z_{k-2} \vee x_k$. (This is the usual reduction from k-SAT to $3-SAT$.) The reduction simply defines z_i as $x_1 \vee x_2 \vee ... \vee x_{i+1}$. That the clauses of Ψ have degree k proofs from the clauses the corresponding clauses of Φ follows from the completeness of polynomial calculus, and the fact that each constraint involves at most k original variables. $\Box$

Lemma 4 *For all n and p, and for any field F of characteristic q, where $q \nmid p$, and F includes the primitive p-th root of unity, $TS_n(p)$ is (d_1, d_2)-reducible to MOD_p^m over F, where $m = n + nrp$, $d_1 = 2pr$ and $d_2 = 2pr$.*

Proof of Lemma 4. Let G'_n be a directed Tseitin graph on n vertices, where $n \equiv 1 \pmod{p}$. That is, the underlying G'_n is an r-regular graph; each vertex of G'_n has a charge of 1, and the edges of G'_n are labeled with values from $[0, p-1]$. Thus, the total number of directed edges of G'_n is rn. From G we will define a universe U of size m, and a corresponding p-partition of this universe, where $m = n + nrp$. In U, there will be one element corresponding to each vertex of G'_n, and there will also be p elements corresponding to each directed edge of G'_n. We will denote the element of U corresponding to vertex i in G'_n by (i), and the vector of p elements of U corresponding to the edge $\langle i, j \rangle$ in G'_n will be denoted by $(i, j, *) = \langle (i, j, 1), (i, j, 2), \ldots, (i, j, p) \rangle$.

Definition. *The elements in U associated with node i will be (i), plus all elements $(i, k, *)$. (That is, the rp elements corresponding to outgoing edges from i plus the element corresponding to node i.) The elements in U associated with the pair of nodes i, j will be the rp elements corresponding to the directed edge $\langle i, j \rangle$ plus the rp elements corresponding to the directed edge $\langle j, i \rangle$.*

The partition of U is defined as follows. We will consider node i in G'_n, and the r labeled edges, $(i, j_1), (i, j_2), \ldots, (i, j_r)$, leading out of i, where $j_1 < j_2 < \cdots < j_r$. Suppose that the values of these edges are: a_1, a_2,...,a_r. Then for each ℓ, $1 \le \ell \le r$, we take the first a_ℓ elements in U from $(i, j_\ell, *)$, and group them with the first $(p - a_\ell)$ elements in U from $(j_\ell, i, *)$. (This gives us r p-partitions so far.) Note that the number of remaining, ungrouped elements associated with node i is $(p - a_1) + (p - a_2) + \cdots + (p - a_r) + 1$, which is congruent to 0 mod p as long as $(a_1 + \cdots + a_r) \bmod p = 1$.

We then group these remaining, ungrouped elements associated with i, p at a time, in accordance with the following ordering. Ungrouped elements from $(i, j_1, *)$ are first, followed by ungrouped elements from $(i, j_2, *)$, and so on until we get to the ungrouped elements from $(i, j_r, *)$, and lastly the element (i).

It should be intuitively clear that if the values $y_{i,j}$ satisfy $TS_n(p)$, that is, if they are set so that the mod p sum coming out of each vertex in G'_n is congruent to 1 mod p, and $y_{i,j}y_{j,i} = 1$ and $y^p_{i,j} = 1$, then the corresponding partition of U is a proper p partition. We want to prove this now formally, with small-degree PC refutations. There are two steps to this reduction. First, for each variable x_e underlying MOD^m_p, we want to define a degree at most rp polynomial, call it $r_e(\overline{y})$, in the $y_{i,j}$ variables that corresponds to the above reduction. Secondly, we want to show that there is a small degree PC derivation of $MOD^m_p(\overline{r}_e)$ from $TS_n(p)$.

Step 1: Defining r_e. We will first describe the defining polynomial r_e for x_e. Recall that e is a particular p-set from U. In the above reduction, the valid p-partitions are of two types: (i) where the elements of e are a subset of the elements associated with a pair of nodes i, j in G'_n; (ii) where the elements of e are a subset of the elements associated with a node i. Thus, if the underlying p elements from e are not one of these two types, then x_e is just set to 0.

Now consider case (i); that is, the elements of e are a subset of the elements associated with the pair of nodes i, j. Suppose that e is the set $\{(i, j, 1), (i, j, 2), \ldots, (i, j, a_1), (j, i, 1), \ldots, (j, i, p - a_1)\}$. That is, e consists of an initial segment of size a_1 of the p elements associated with directed edge $\langle i, j\rangle$ and an initial segment of size $p - a_1$ of the p elements associated with $\langle j, i\rangle$. (If e is not of this form, then again x_e is just 0.) Then x_e should be 1 if $y_{i,j} = \omega^{a_1}$, $y_{j,i} = \omega^{p-a_1}$ and should be 0 otherwise. This is defined by the following polynomial:

$$\prod_{a \neq a_1} (\omega^{a_1} - \omega^a)^{-1}(y_{i,j} - \omega^a) \times \prod_{b \neq p-a_1} (\omega^{p-a_1} - \omega^b)^{-1}(y_{j,i} - \omega^b) \qquad (1)$$

More generally, suppose that we want to define a 0-1 valued variable x so that $x = 1$ if $y_1 = \omega^{p_1}$ and $y_2 = \omega^{p_2}$ and ... and $y_k = \omega^{p_k}$, and otherwise $x = 0$. Then this is accomplished by the following degree kp polynomial:

$$\prod_i \prod_{p \neq p_i} (\omega^{p_i} - \omega^p)^{-1}(y_i - \omega^p) \qquad (2)$$

Case (ii) is handled similarly but is somewhat more complicated. Now the elements of e are a subset of elements associated with i, and moreover we can assume without loss of generality that they must be end-segments of $(i, j_{l1}, *)$, $(i, j_{l2}, *), \ldots, (i, j_{ll}, *)$ plus possibly either (i) or a consecutive segment of $(i, j_{ll+1}, *)$. (Otherwise, x_e is just set to zero.) Then x_e should be 1 if and only if there exists values $a_1, \ldots, a_r$ assigned to the outgoing edges $(i, j_1), \ldots, (i, j_r)$ such that the partition described in the reduction above groups the elements of e together. This is a big OR (translated as a sum) (of size at most p^r) over the good values of $a_1, \ldots, a_r$ that group e together. Thus, it is expressible by a polynomial in the variables $y_{i,j_1}, y_{i,j_2}, \ldots, y_{i,j_r}$ of degree at most pr.

Step 2: Deriving $MOD_p^m(r_e)$ from $TS_n(p)$. We will now describe how to give small degree PC derivations of the equations $MOD_p^m(r_e)$ from $TS_n(p)$. Recall that the equations in $MOD_p^m(r_e)$ are as follows.

1. $r_e^2 - r_e = 0$ for all p-sets e
2. $r_e r_f = 0$ for all e, f such that $e \cap f \neq 0$, $e \neq f$
3. $\sum_{e, u \in e} r_e - 1 = 0$, for all $u \in [m]$.

We want to show that for every equation E that we need to derive as described above, that E is a tautological consequence of a small, constant number of equations from $TS_n(p)$. Then, since each equation of $TS_n(p)$ involves only a constant number of variables, by completeness of PC it will follow that there is a small-degree derivation of each equation E.

Definition. *Let $f_1 = 0, \ldots, f_k = 0, g = 0$ be polynomial equations over a field F with underlying variables $x_1, \ldots, x_n$. Then g is a tautological consequence of $f_1, \ldots, f_k$ if for every assignment α to the underlying variables, if all of the equations $f_1, \ldots, f_k$ are satisfied by α, then $g = 0$ is also satisfied by α.*

By generalizing slightly the completeness result in [6], (Theorem 5.2 part 2), it can be shown that if g is a tautological consequence of $f_1, \ldots, f_k$, all with underlying variables $x_1, \ldots, x_n$, and if $f_1, \ldots, f_k$ includes the equations $x_i^p = 1$ for all variables x, then there is a degree pn derivation of g from $f_1, \ldots, f_k$.

In light of the above, it is just a matter of verifying that each of the above equations E is a tautological consequence of a small number of equations from $TS_n(p)$ involving a small number of variables. In particular, equations of type (1) require degree pr and equations of type (2) and (3) each require degree at most $2pr$.

This completes the proof of Lemma 4. $\Box$.

4 Intuition and an upper bound

In order to first give some intuition behind the lower bound for the Tseitin tautologies, it is helpful to think about the natural PC refutation for these equations. To be concrete, we consider the mod 2 case; the others are similar.

Initially, the equations say that the number of edges out of a single vertex v is odd. These equations have degree r. Then in degree at most $2r$, one can combine two of these equations to say that the number of edges out of a set of vertices of size 2 is even. Continuing in this way, if $S \subset V$, then one can derive an equation saying that the number of edges out of S, $E(S)$, has the same parity as the size of S. This equation is most naturally expressed as $m - 1 = 0$ if $|S|$ is even, and $m + 1 = 0$ if $|S|$ is odd, where m is the product of the variables corresponding to edges $E(S)$, that cross between S and its complement. Thus, the degree of this polynomial is equal to the size of $E(S)$. Proceeding this way, we eventually obtain two equations, one saying that the number of edges out of a set S_1 is odd, and the other one saying that the number of edges out of a set S_2 is even, where S_1 and S_2 are disjoint, and $S_1 \cup S_2 = V$. This will lead to a derivation of 1, since we have now derived $m + 1$ and $m - 1$ for some monomial m. If G_n is highly expanding, the degree of this refutation will be large since at some point we must pass through a relatively large set of edges. Thus, any binomial expressing that the number of edges out of this set must have the same parity as the size of the set, will have a large number of variables and hence large degree.

We want to show that the above almost completely characterizes what can be done with the initial equations. Suppose we have derived $m - 1 = 0$, where m is the set of edges $E(S)$, such that $|E(S)| = d$, and $|S|$ is even. (Or similarly, we have derived $m + 1 = 0$ when m is the set of edges of $E(S)$ but now $|S|$ is odd.) However, now it is possible to rewrite this equation in a slightly different form so that it has smaller degree. In particular, we can divide up the edges of m into two halves, m_1 and m_2 and rewrite the equation $m - 1 = 0$ instead as $m_1 - m_2 = 0$. This is derived from $m - 1$ in degree d by multiplying $m-1$ by edges of m_2, one at a time, thus transferring the edges of m_2 over to the second term, one at a time. This new equation, $m_1 - m_2 = 0$ has degree $d' = \lceil d/2 \rceil$, and in general is not derivable by a degree d' PC refutation. The (degree d) equations that interest us are this larger set of equations, which express the fact that the edges coming out of a set S are even (or odd) by a pair of monomials.

There are two key steps to making this intuition a proof. First, we must show that, although the PC proof can contain arbitrary polynomials, the important lines are equalities as above, or *binomials* if viewed as a difference.

This is made formal in a very general way in section 5. Secondly, that the set of degree d equations described above, although not all provable with degree d proofs, is more natural and thus easier to understand, and they span all of the degree d derivable PC polynomials. In contrast, an explicit construction of the exact set of degree d derivable PC polynomials (as done by Razborov [16] for pigeonhole principle) seems much more difficult, although it is possible for symmetric systems ([13]).

5 Binomial systems and bounds for PC

In the previous section, we reduced the problem of proving lower bounds for the mod counting principles to that of proving lower bounds for the Tseitin graph tautologies. The reason this is progress is that the Tseitin graph tautologies are expressed as a system of polynomials of a very simple form: each polynomial is a *binomial*, the difference of two *terms* (i.e., the weighted sum of two monomials with coefficients over the field.) (This fact was earlier used by Grigoriev [9] in giving lower bounds for Nullstellensatz.) A binomial $a_1m_1 - a_2m_2$ can be viewed as the equation between two terms, $a_1m_1 = a_2m_2$. Intuitively, an algebraic proof for a binomial system should be expressible as a sequence of such equations.

We next recall a general characterization of things provable in PC from [8], and then will show that this characterization can be refined for binomial systems.

Definition. *A degree d* pseudo-ideal *I is a vector space of degree at most d polynomials so that if $p \in I$ and p has degree $\leq d-1$, then $xp \in I$ for every variable x.*

Theorem 5 *[8] Let P be a system of polynomials, and let $I_d(P)$ be the set of all polynomials q that have a degree d PC proof from P. Then $I_d(P)$ is a d-pseudo-ideal, and for any d-pseudo-ideal I containing P, $I_d(P) \subseteq I$.*

So pseudo-ideals capture provability in polynomial calculus. If equational reasoning is complete for polynomial calculus for binomial systems, it should follow that the pseudo-ideals for such systems are determined by which terms are "provably equal" from the system. In other words, pseudo-ideals should be determined by an equivalence relation on degree d terms with certain closure properties. This is formalized below. (By 'ring' we always mean 'commutative ring'.)

Definition. *Let R be a ring and R^* a multiplicative subgroup of R, and let $x_1, \ldots, x_n$ be variables. (i.e., R^* consists only of invertible elements and is* *closed under products and inverses). An R^*-term is a term whose coefficient is from R^*. An R^*-binomial is the difference of two R^*-terms. A d-*Laurent relation[7] *over R^*-terms is an equivalence relation $\equiv_d$ on R^*-terms of degree at most d with the following properties: Let t_1, t_2, be R^*-terms of degree at most d and let $r \in R^*$.*

(a) *$t_1 \equiv_d t_2$ iff $rt_1 \equiv_d rt_2$; and*

(b) *If t_1 and t_2 are degree at most $d-1$, and $t_1 \equiv_d t_2$ then $x_i t_1 \equiv_d x_i t_2$ for any variable x_i.*

If $\equiv_d$ is a d-Laurent relation, we define a corresponding set of binomials $B_{\equiv_d} = \{t_1 - t_2 | t_1 \equiv_d t_2\}$ and a set of polynomials $S_{\equiv_d} = SPAN_R(B_{\equiv_d})$, the set of linear combinations of binomials in $B_{\equiv_d}$.

R will usually be a field, but in section 8 we will need the more general version. Intuitively, $\equiv_d$ represents the set of pairs of terms that can be proved equal using equational-type reasoning, where we are allowed to multiply both sides of a known equation by a constant or variable, as long as we don't exceed degree d.

We now show that lower bounds on polynomial calculus proofs can be established by exhibiting a non-trivial d-Laurent relation.

Theorem 6 *Let Q be a set of R^* binomials. If $\equiv_d$ is a d-Laurent relation with $Q \subseteq B_{\equiv_d}$ and $1 \not\equiv_d a$ for any $a \in R^*, a \neq 1$, then Q has no degree d polynomial calculus refutation over R.*

The proof of this theorem follows from a sequence of lemmas that take up the rest of this section. Lemma 7 is the main technical lemma, and the other lemmas describe how to use it to prove the theorem.

Lemma 7 *Assume $\equiv_d$ is d-Laurent.[8] Suppose $f \in S_{\equiv_d}$. Then f can be rewritten as a linear combination $f = \sum_{j=1}^{T} a_j(t_j - t'_j)$ of binomials from $B_{\equiv_d}$ such that no monomial completely cancels out, i.e., every monomial t_j, t'_j in the linear combination appears in f with non-zero coefficient.*

Proof. Let $f = \sum_j a_j(t_j - t'_j)$, where each pair of monomials in the above sum is a polynomial from $B_{\equiv_d}$. We prove the lemma by induction on the number of distinct monomials in the above sum. At each step, if cancellation of a monomial occurs, we will rewrite f by an equivalent sum of elements of $B_{\equiv_d}$ such that the number of monomials in the new sum is strictly smaller.

Assume m appears in the sum, without loss of generality in exactly the first T' differences, but has zero coefficient in f. Because each element

[7] We name them in honor of Laurent, who initiated the study of equational reasoning for term algebras.

[8] Actually, we do not need property (b) in the definition of d-Laurent relation for this lemma.

$a_j \in R^*$, and so has an inverse, by factoring out the coefficient of m in each term, we can rewrite any elements that m appears in: $c_k(a_k m - a'_k m'_k) = c_k a_k(m - a'_k a_k^{-1} m'_k) = d_k(m - t_k)$ for some R^* term t_k. Also, by the closure properties of $\equiv_d$ for multiplication by constants from R^*, $m \equiv_d t_k$. Now, since m has coefficient 0 in f, $\sum_k d_k = 0$.

We claim that the sum of binomials containing m, $\sum_{k=1}^{T'} d_k(m - t_k)$, can be rewritten as $\sum_{k=2}^{T'} d_k(t_1 - t_k)$. This is because $\sum_{k=2}^{T'} d_k(t_1 - t_k) = (\sum_{k=2}^{T'} d_k)(t_1) - \sum_{k=2}^{T'} d_k t_k = -d_1(t_1) - \sum_{k=2}^{T'} d_k t_k = -\sum_{k=1}^{T'} d_k t_k = (\sum_{k=1}^{T'} d_k)m - \sum_{k=1}^{T'} d_k t_k = \sum_{k=1}^{T'} d_k(m - t_k)$.

Since $\equiv_d$ is transitive, $t_1 \equiv t_k$ for all k. So this substitution rewrites f as a weighted sum of members of $B_{\equiv_d}$. The new sum is without m and without any monomial not in the previous sum, so contains one fewer monomial. □

Lemma 8 *If $\equiv_d$ is d-Laurent, and there is a $c \in R, c \neq 0$ with $c \in S_{\equiv_d}$, then there is an $a \in R^*, a \neq 1$ with $1 \equiv_d a$.*

Proof. If $c \in S_{\equiv_d}$, by Lemma 7, c can be written as a sum of equivalent terms which only have monomials that appear in c, i.e, are constants. Thus, at least two distinct constants $a \equiv_d a'$, and then $1 \equiv_d a'a^{-1}$. □

Lemma 9 *If $\equiv_d$ is d-Laurent, then $S_{\equiv_d}$ is a degree d pseudo-ideal.*

Proof. By definition, $S_{\equiv_d}$ is a vector space of polynomials of degree at most d, so we just need to show closure under multiplication by a variable, provided the total degree is at most d. Assume $f \in S_{\equiv_d}$ has degree at most $d-1$. By Lemma 7, we can write $f = \sum_{i=1}^{T} c_i(t_i - t'_i)$, where $t_i \equiv_d t'_i$ and each t_i, t'_i comes from a monomial with non-zero coefficient in f. In particular, each t_i, t'_i has degree at most $d-1$. Therefore, $xt_i \equiv xt'_i$ by the second closure property in the definition of d-Laurent relation. So $xf = \sum_{i=1}^{T} c_i(xt_i - xt'_i) \in S_{\equiv_d}$. □

Proof (of Theorem 6). Let $\equiv_d$ be a d-Laurent relation with $Q \subseteq B_{\equiv_d}$, and that $1 \not\equiv_d a$ for any $1 \neq a \in R^*$. Assume Q has a polynomial calculus refutation of degree d over R, i.e., proves some $c \neq 0$, $c \in R$. Then $c \in S_{\equiv_d}$, since the latter is a pseudo-ideal containing Q. But then $1 \equiv_d a$ for some $a \neq 1, a \in R^*$. This contradiction proves the theorem. □.

The notion of d-Laurent relations is similar to the definition of the *d-Laurent* proof system, which is an algebraic proof system introduced in [10] and shown therein to be closely related to the restriction of the polynomial calculus to binomials in that lower bounds on the degree of Laurent proofs imply lower bounds on the degree of polynomial calculus proofs. [10] also introduces the related algebraic *Thue* systems and proves their equivalence with Nullstellensatz proofs.

6 PC lower bound for mod 2

We first prove linear lower bounds for the Tseitin principle $TS_n(2)$ for polynomial calculus over fields of characteristic $q > 2$, provided the underlying graph is an expander graph.

Definition. *Let $G = (V, E)$ be an undirected graph, and let $\epsilon > 0$. The graph G has* expansion ϵ *if for any subset S of vertices with $|S| \leq |V|/2$, $|N(S)| \geq (1+\epsilon)|S|$, where $N(S)$ is the set of nodes adjacent to nodes in S.*

Theorem 10 *Let F be a field and let G_n be a graph on n nodes with expansion ϵ. For all $d < \epsilon n/8$, there is no degree d PC refutation of $TS_n(2)$ over F.*

Note that there is no restriction on the characteristic q of the field F or the degree of the graph. When q is an odd prime or zero, then the $TS_n(2)$ polynomials are unsatisfiable and therefore have a PC refutation over F, of degree which is necessarily linear by the theorem. When $q = 2$, then the $TS_n(2)$ polynomials are easily seen to be satisfiable (trivially, since $1 = -1$), and there is no PC-refutation of $TS_n(2)$ at all. Similarly, while the theorem is true for all graphs, it is only interesting for low degree graphs, since the degree of G_n is the degree of the polynomials in $TS_n(2)$.

It is an easy corollary of Theorem 10 and Lemmas 1 and 4 that over a field of characteristic $q \neq 2$, PC-refutations of the MOD_2^n polynomials require size linear in n: this is established as Corollary 20 below for general p in place of 2.

Preparatory to proving Theorem 10, we establish some definitions and lemmas. In what follows, we will reduce all polynomials by $y_{i,j}^2 = 1$ for all variables, thus obtaining only multilinear polynomials.

Definition. *For a monomial $m = \prod_i y_i^{f_i}$, define the* multilinearization $\overline{m}$ *of m to be $\prod_i y_i^{f_i \bmod 2}$. For a multilinear monomial m we define E_m to be the set of edges e such that y_e is a factor of m.*

Definition. *For two sets A, B, $A +_2 B$ denotes the disjoint union of A and B.*

Definition. *Let $S \subseteq V$, where V is the set of vertices in G_n. Then $E(S)$ is defined to be the set of edges with exactly one endpoint in S and one endpoint outside of S.*

Proposition 11 *Let G_n be an expander graph with expansion ϵ. If $S \subseteq V$, $|S| \leq n/2$, then $|E(S)| \geq \epsilon|S|$.*

Proof Since $|S| \leq n/2$, $|N(S)| \geq (1+\epsilon)|S|$ by the definition of expansion. Then $|N(S) - S| \geq \epsilon|S|$, and each node in $N(S) - S$ is the endpoint of at least one edge in $E(S)$. $\Box$.

We shall prove Theorem 10 as a corollary to Theorem 6: for this, we let $R = F$ and $R^* = \{-1, 1\}$. The R^*-terms are thus just the terms m and $-m$ where m is a monomial.

Definition. *We define an equivalence relation $\equiv_d$ on the R^*-terms of degree at most d multilinear monomial, as follows. Let $b_1, b_2 \in \{0,1\}$, $(-1)^{b_1} m_1 \equiv_d (-1)^{b_2} m_2$ if there exists a set $S \subset V$ such that*

1. $E_{\overline{m_1 m_2}} = E(S)$.
2. $|S| < n/2$; *and*
3. $|S| \equiv b_2 - b_1 \pmod 2$.

We will show that there is no degree $d < \epsilon n/8$ PC refutation of $TS_n(2)$ by showing that that $\equiv_d$ is a d-Laurent relation.

Lemma 12 *If $d < \epsilon n/8$, then the relation $\equiv_d$ is an equivalence relation.*

Proof It is easy to see from the definitions that $(-1)^b m \equiv_d (-1)^b m$ and that $(-1)^{b_1} m_1 \equiv_d (-1)^{b_2} m_2$ iff $(-1)^{b_2} m_2 \equiv_d (-1)^{b_1} m_1$. We need to show that if $(-1)^{b_1} m_1 \equiv_d (-1)^{b_2} m_2$ and $(-1)^{b_2} m_2 \equiv_d (-1)^{b_3} m_3$, then $(-1)^{b_1} m_1 \equiv_d (-1)^{b_3} m_3$. Let S_1 be the set of vertices such that $E(S_1) = E_{\overline{m_1 m_2}}$, $|S_1| \equiv b_2 - b_1 \pmod 2$, $|S_1| < n/2$, and similarly let S_2 be the set of vertices such that $E(S_2) = E_{\overline{m_2 m_3}}$, $|S_2| \equiv b_3 - b_2 \pmod 2$, $|S_2| < n/2$. We want to show that $S' = S_1 +_2 S_2$ is a set of vertices such that $E(S') = E_{\overline{m_1 m_3}}$, $|S'| = b_3 - b_1$, and $|S'| < n/2$. Intuitively, this is saying that if S_1 has parity $b_2 - b_1$ which equals the parity of $|E(S_1)|$, and S_2 has parity $b_3 - b_2$, which equals the parity of $|E(S_2)|$, then $S_1 +_2 S_2$ has parity $b_3 - b_1$, which equals the parity of $|E(S_1 +_2 S_2)|$. And furthermore, $|S_1 +_2 S_2|$ is not too large.

Clearly, $|S'| \bmod 2 = |S_1| \bmod 2 + |S_2| \bmod 2 = b_2 - b_1 + b_3 - b_2 = b_3 - b_1$. Also we have: $E(S_1 +_2 S_2) = E(S_1) +_2 E(S_2) = E_{\overline{m_1 m_2 m_2 m_3}} = E_{\overline{m_1 m_3}}$.

It is left to show that $|S'| < n/2$. Since $|m_1|, |m_2| \leq d$, it follows that $|E(S_1)| \leq |m_1 m_2| \leq 2d$. Since G_n is an expander graph, Proposition 11 implies that $|E(S_1)| \geq \epsilon|S_1|$, and thus it follows that $|S_1| \leq 2d/\epsilon < n/4$. Similarly, $|S_2| \leq n/4$. Thus, $|S'| \leq |S_1| + |S_2| < n/2$. (In fact, since $|E(S')| \leq |m_1 m_3| \leq 2d$, Proposition 11 further implies that $|S'| \leq n/4$.) $\Box$

Lemma 13 *For $d \leq \epsilon n/8$, $\equiv_d$ is a d-Laurent relation.*

Proof. Let $d \leq \epsilon n/8$. We just established that $\equiv_d$ is an equivalence relation. Condition (a) of the definition of d-Laurent is trivially satisfied from the definition of $\equiv_d$. Also, since $\overline{(xM_1)(xm_2)} = \overline{m_1 m_2}$ for any variable x, condition (b) of the definition of d-Laurent is also satisfied. $\Box$.

Lemma 14 *Every polynomial of $TS_n(2)$ is a binomial from $B_{\equiv_d}$.*

Proof. There are two kinds of polynomials in $TS_n(2)$. For the equations $y_e^2 - 1$, we must show that $y_e^2 \equiv_d 1$. This is easily done by taking $S = \emptyset$ and noting that since $\overline{y_e^2} = 1$, the three conditions of the definition of $\equiv_d$ are trivially satisfied. For the equations of the form $1 + y_{\{i,j_1\}} y_{\{i,j_2\}} \cdots y_{\{i,j_r\}} = 0$, we must show that $1 \equiv_d (-1) y_{\{i,j_1\}} y_{\{i,j_2\}} \cdots y_{\{i,j_r\}}$. This is easily seen to hold with $S = \{i\}$. $\square$

Proof of Theorem 10. This is a consequence of Theorem 6. First, Lemma 13 shows $\equiv_d$ is d-Laurent. Second, Lemma 14 shows $TS_n(2) \subset B_{\equiv_d}$. It remains to show that $1 \not\equiv_d (-1)$. To prove this suppose $1 = (-1)^0 \equiv_d (-1) = (-1)^1$ holds with some set S satisfying the conditions of the definition $\equiv_d$. On the one hand, we must have $|S| = 1 - 0 = 1 mod 2$, and also we must have $E(S) = \emptyset$. But on the other hand, $|S| < n/2$, so Lemma 11 implies $E(S)$ is non-empty — a contradiction. Therefore, the hypotheses of Theorem 6 hold, and there is no PC refutation of $TS_n(2)$ over F of degree d. $\square$

7 PC lower bound for the general case

This section extends our linear lower bounds to the degrees of PC refutations of $TS_n(p)$ over a field F of characteristic q.

Theorem 15 *Let F be a field of characteristic q containing a primitive p-th root of unity, and let G_n be an r-regular graph with expansion ϵ. Then, for all $d < \epsilon n/8$, there is no degree d PC refutation of $TS_n(p)$ over F.*

As a corollary to this theorem and Lemmas 1 and 4, we shall prove (as Corollary 20) that when $q \nmid p$, any PC refutation of the MOD_p^n polynomials over F requires linear degree.

In order to express the $TS_n(p)$ polynomials, F must contain a p-th primitive root of unity, ω. We let $R = F$ and R^* be the powers of the root of unity, i.e., $R^* = \{1, \omega, \omega^2, \ldots, \omega^{p-1}\}$. For the rest of this section, it is sufficient to assume only that R is a ring (rather than a field). See section 8 for more explanation of what it means for a ring to have a p-th root of unity.

Definition. *We consider multisets with multiplicities in Z_p, i.e., a function from elements of a fixed universe to Z_p. An element from the universe that does not occur in the multiset is considered to have multiplicity 0. The size of a multiset is the number of elements of non-zero multiplicity. We identify a set with the multiset given by its characteristic function. Let A and B be two multisets. Then $A +_p B$ denotes the multiset, where if x occurs in A with* *multiplicity a, and in B with multiplicity b, then x occurs in $A +_p B$ with multiplicity $(a+b) \bmod p$. (Note that when A and B are ordinary sets and $p=2$, then $A +_2 B$ is simply the disjoint union of A and B.)*

Definition. *For each undirected edge $\{i,j\} \in E$, fix a directed edge on the same set of vertices arbitrarily; call the set of directed edges $\hat{E}$. For any monomial m in the variables y_e, $m = \Pi_{(i,j)|\{i,j\}\in E} y_{(i,j)}^{a_{(i,j)}}$, let the* normalization *of m, $\overline{m}$, be given by $\overline{m} = \Pi_{(i,j)\in\hat{E}} y_{(i,j)}^{a_{(i,j)}-a_{(j,i)} \bmod p}$. (Intuitively, we derive $\overline{m}$ from m using $y_{(i,j)}^p = 1$ and $y_{(i,j)}y_{(j,i)} = 1$.) The* inverse *m^{-1} is $\Pi_{(i,j)|\{i,j\}\in E} y_{(i,j)}^{p-a_{(i,j)}}$.*

Let S be a multiset over the vertices V in G, where vertex i occurs in S with multiplicity s_i. $E(S)$ will denote a multiset of edges from $\hat{E}$ where edge (i,j) occurs with multiplicity $s_i - s_j$.

We identify a normalized monomial with the corresponding multiset of edges from $\hat{E}$.

Proposition 16 *Let G_n be an expander graph with expansion ϵ. If $|S| \le n/2$, then $|E(S)| \ge \epsilon|S|$.*

Proof. Consider any edge $e = (i,j) \in \hat{E}$ where i has non-zero multiplicity in S and j has 0 multiplicity in S, or vice versa. Then e has non-zero multiplicity in $E(S)$. Thus the proposition follows from Proposition 11, since every undirected edge between the sets of nodes with zero and non-zero multiplicities corresponds to such an edge in $\hat{E}$. (In fact, when members of S have different non-zero multiplicities, it only makes the size of $E(S)$ increase.) $\Box$

Definition. *We define the binary relation $\equiv_d$ on the R^*-terms $\omega^b m$ where m is a degree at most d monomial and $0 \le b < p$. $\omega^{b_1} m_1 \equiv_d \omega^{b_2} m_2$ if there exists a multiset S of vertices such that (i) $\overline{m_1 m_2^{-1}}$ corresponds to $E(S)$; (ii) $|S| < n/2$; and (iii) $\sum_i s_i \equiv b_2 - b_1$ (mod p).*

The next three lemmas are proved exactly analogously to Lemmas 12-14.

Lemma 17 *For $d \le \epsilon n/8$, the relation $\equiv_d$ is an equivalence relation.*

Lemma 18 *For $d \le \epsilon n/8$, $\equiv_d$ is a d-Laurent relation.*

Lemma 19 *Every polynomial of $TS_n(2)$ is a binomial from $B_{\equiv_d}$.*

Proof of Theorem 15. Exactly as argued in the proof of Theorem 10, we have that $1 \not\equiv_d a$ for any $a \in R^*$ distinct from 1, i.e., $1 \not\equiv_d \omega^i$ for all $0 < i < p-1$. Thus Theorem 15 follows from Theorem 6 using Lemmas 18 and 19. $\Box$

Corollary 20 *Let $q \geq 2$ be a prime such that $q \nmid p$ and let F be a field of characteristic q. Any PC-refutation of the MOD_p^n polynomials requires degree $> \delta n$, for some constant $\delta > 0$.*

Proof Choose constants ϵ and r so that there are r-regular graphs G_n of expansion ϵ for all n. Let $d_1 = d_2 = 2pr$. Suppose MOD_p^m has a degree d_3 PC refutation, where $m = n + nrp$. By Lemmas 1 and 4 $TS_n(p)$ has a degree $d_3 d_1$ PC refutation, so by Theorem 15, $d_3 d_1 > \epsilon n/(8pr)$. Thus, $d_3 > \epsilon m/(16p^2r^2(1+rp))$. Since ϵ, r, p, d are constants, this proves the Corollary. □

Corollary 21 *Let $q \geq 2$ be a prime such that $q \nmid p$ and let F be a field of characteristic q. Any PC-refutation of the $BTS_{n,p}$ polynomials requires degree $> \delta n$, for some constant $\delta > 0$.*

Proof Let F' be a finite extension field of F containing a p-th root of unity. By Lemma 2, $BTS_{n,p}$ is $(1, rp)$ reducible to $TS_{n,p}$, and by Theorem 15, polynomial calculus refutations of $TS_{n,p}$ require degree $\Omega(n)$ over F'. Thus, the same is true for proofs of $BTS_{n,p}$ over F'. Since a polynomial calculus proof over F is a special case of a polynomial calculus proof over F' (provided the original polynomials have co-efficients in F), the same is true over F. □

Corollary 22 *Any implementation of the Gröbner basis algorithm over any finite field that uses an explicit representation of polynomials as the sum of their non-zero monomials requires exponential worst-case time to solve 3-SAT, with the usual representation as a system of polynomials.*

Proof The size of a polynomial calculus proof is the number of non-zero monomials in all of its lines. As noted in [8], any implementation as above produces a polynomial calculus proof whose total size is bounded above by the time of the algorithm. In [12], it is shown that any polynomial calculus refutation of a constant degree system of polynomials requires size $2^{\Omega(d^2/n)}$, where d is the minimal degree of such a refutation, and n is the number of variables. In particular, assume the algorithm uses some field of characteristic q, and choose a different prime p. For the k-CNF $BTS_{n,p}$, we have $d = \Omega(n)$ as a lower bound on the degree. Since $BTS_{n,p}$ has $O(n)$ clauses, by Lemma 3, we can find a 3-CNF translation that also requires degree $\Omega(n)$ and has $O(n)$ variables. Hence, any proof requires size $2^{\Omega(n)}$, and so the running time of the Gröbner basis algorithm on the corresponding 3-CNF is exponential. □

8 Polynomial calculus over rings

We now consider the polynomial calculus over rings instead of over fields. For this, we consider a fixed ring R and the polynomials have coefficients from R. Since the definition of the polynomial calculus did not use any field-specific properties, e.g., since the definition did not depend on the existence of inverses, it is completely natural to consider the polynomial calculus over rings. As before, we define a PC derivation to be a sequence of polynomials $\langle P_i \rangle_i$ with the same rules of addition and multiplication. However, we modify the definition of a PC refutation of Q to be a PC derivation that ends with a constant polynomial m where $m \in R$ is non-zero (and its derivation thus corresponds to a derivation of the contradiction $m = 0$).

It is known that the polynomial calculus over rings is complete with respect to Boolean reasoning, i.e., if the initial polynomials include $x^2 - x = 0$ for each variable x then any unsatisfiable set of polynomials has a PC refutation. However, the polynomial calculus over rings is not complete for general derivations, see the examples in [5]. In this respect the polynomial calculus over a field is stronger than the polynomial calculus over a ring. On the other hand, if the ring R is $\mathbb{Z}_m$ where $m = p_1 \cdot p_2$ for distinct primes p_1, p_2, then it is well-known that there are constant-degree polynomial calculus proofs of $MOD^n_{p_1}$ and $MOD^n_{p_2}$. But Theorem 15 implies that there is no single field for which the polynomial calculus has constant degree proofs of both these principles.

The situation is a little analogous to an important open problem in circuit complexity. Namely, Smolensky [17] showed that polynomial size constant-depth circuits with mod-q gates cannot compute the mod-p function for distinct primes p, q. However, it is open whether this is true for composite values of q where $p \nmid q$.

We prove below that if p and q are relatively prime, then over the ring Z_q, any PC refutation of MOD^n_p requires degree $\geq \delta n$ for some constant δ. The general outline of the proof is similar to the approach used for the proof of Theorem 15.

In the next section, we do some preliminary work introducing rings with roots of unity. Following that, we discuss the reduction of the Tseitin principle to the mod p counting principle and then discuss the lower bound for Tseitin principle.

8.1 Rings with roots of unity

We are mostly interested in lower bounds on the degree of polynomial calculus refutations over rings $R = \mathbb{Z}_q$; however, our method of proof depends strongly on the use of p-th roots of unity, and on the existence of inverses of certain terms involving the p-th root of unity. In this section, we prove that there exist rings containing $\mathbb{Z}_q$ with the desired p-th roots of unity.

Theorem 23 *Let $p, q > 1$ be relatively prime. Then there is a finite ring $R \supset Z_q$ which contains a p-th root of unity ω such that*
(a) *p is the least positive integer i such that $\omega^i = 1$,*
(b) *For all $0 \leq j < k < p$, $(\omega^k - \omega^j)$ has an inverse in R.*

Proof. First we shall give a simple proof for the case where q is a product of *distinct* primes $q = r_1 \cdot r_2 \cdots r_m$. For any prime r, let GF_r denote the field of order r. Let $F_i = GF_{r_i}[\sqrt[p]{1}]$ be the extension of GF_{r_i} obtained by adjoining a p-th root of unity. We use ω_i to denote a p-th root of unity in F_i. Define R to be the ring with domain $\prod_i F_i$ and component-wise addition and multiplication. An element of R is an m-tuple $\langle a_1, \ldots, a_m \rangle$ with $a_i \in F_i$. By the Chinese remainder theorem, a copy of Z_q is embedded in R by $n \mapsto \langle n \bmod r_1, \ldots, n \bmod r_m \rangle$. The element $\langle a_1, \ldots, a_m \rangle$ has an inverse in R iff each $a_i \neq 0$. Letting $\omega = \langle \omega_1, \ldots, \omega_m \rangle$, it is easy to see that ω is a p-th root of unity in R and satisfies property (a). Likewise, $\omega_i^k - \omega_i^j \in F_i$ is non-zero for all i and thus $(\omega^k - \omega^j)^{-1}$ exists in R.

Now consider the general case, where q is not a product of distinct primes. (We don't use any special properties of Z_q beyond the fact that p^{-1} exists in Z_q, which follows from the fact that p and q are relatively prime.) We shall let R be the ring of polynomials from $\mathbb{Q}[\sqrt[p]{1}]$ with integer coefficients modulo q; R can be explicitly constructed as follows: Consider a primitive p-th root of unity, ν, over the field of rationals. As a root of unity, ν is a root of the polynomial $x^{p-1} + x^{p-2} + \cdots + x^2 + x + 1$. Likewise, for any $\ell < p$ dividing p, $\nu^{p/\ell}$ is a primitive ℓ-th root of unity, so ν is a root of $x^{(\ell-1)p/\ell} + x^{(\ell-2)p/\ell} + \cdots + x^{p/\ell} + 1$. It follows that there is a non-constant polynomial $Q(x)$ which is the greatest common divisor of each of these polynomials which has ν as a root. Furthermore, by Gauss's lemma (c.f. [11]), we may choose the polynomial $Q(x)$ with leading coefficient 1 and integer coefficients. We define R to be the extension ring $Z_q[\omega]/(Q(\omega))$. Formally, this means we define an equivalence relation on the set $Z_q[\omega]$ of univariate polynomials over Z_q by

$$f \sim g \quad \leftrightarrow \quad \exists h \in Z_q[\omega],\ f(\omega) - g(\omega) = h(\omega) \cdot Q(\omega).$$

Clearly this is an equivalence relation, and addition and multiplication respect $\sim$. The ring $R = Z_q[\omega]/(Q(\omega)) =_{df} Z_q[\omega]/\sim$ has domain the set of $\sim$-equivalence classes (but we generally abuse notation by writing $f \in R$ instead of $[f] \in R$, etc.) Clearly R is a ring. In R, each polynomial $x^{(\ell-1)p/\ell} + x^{(\ell-2)}p/\ell + \cdots + x^{p/\ell} + 1$ is equal to zero, since it is a multiple of Q. Therefore $\omega^p = 1$ in R (i.e., $w^p \sim 1$) since

$$(\omega - 1) \cdot (\omega^{p-1} + \omega^{p-2} + \cdots + \omega + 1) \;=\; \omega^p - 1.$$

Also note that no constant of Z_q becomes equal to zero in R: this is immediate from the fact that Q is a non-constant, monic polynomial over Z_q.

It remains to prove that if $k \neq \ell$, $0 \leq k, \ell < p$, then $(\omega^k - \omega^\ell)$ has a (multiplicative) inverse in R. Since $(\omega^k - \omega^l) = \omega^\ell(\omega^{k-\ell} - 1)$ and ω^ℓ has inverse in R, it will suffice to prove that $(\omega^k - 1)$ has an inverse in R for all $1 \leq k < p$.

Define $i_0 = 0$ and $i_{n+1} = i_n + k \bmod p$. Let ℓ be the least value such that $i_\ell = 0$; of course ℓ divides p. Therefore the values $i_0, \ldots, i_{\ell-1}$ are distinct and enumerate all the values in $\{0, p/\ell, 2p/\ell, \ldots, (\ell-1)p/\ell\}$. For $0 \leq j < \ell$, let $v(j)$ be the value such that $i_{v(j)} = j$, $0 \leq v(j) < \ell$. Define

$$f(\omega) \;=\; \sum_{j=0}^{\ell-1} v(j)\omega^j \;=\; \sum_{n=0}^{\ell-1} n \cdot \omega^{i_n}.$$

Claim: $(\omega^k - 1)f(\omega) = \ell$ holds in R.

Since ℓ has an inverse in Z_q, the claim immediately implies that $(\omega^k - 1)$ has in inverse in R, namely, $\ell^{-1}f(\omega)$.

In R we have

$$\begin{aligned}
(\omega^k - 1) \cdot f(\omega) &= \sum_{n=0}^{\ell-1} n\omega^{i_n+k} - \sum_{n=0}^{\ell-1} n\omega^{i_n} \\
&= \sum_{n=1}^{\ell} (n-1)\omega^{i_n} - \sum_{n=1}^{\ell-1} n\omega^{i_n} \\
&= (\ell - 1) \cdot \omega^{i_\ell} - \sum_{n=1}^{\ell-1} \omega^{i_n} \\
&= (\ell - 1) \cdot 1 + 1 - \sum_{n=0}^{\ell-1} \omega^{i_n}
\end{aligned}$$

$$
\begin{aligned}
&= \ell - \sum_{n=0}^{\ell-1} \omega^{in} = \ell - \sum_{n=0}^{\ell-1} \omega^{np/\ell} \\
&= \ell - 0 = \ell.
\end{aligned}
$$

That completes the proof of the claim and of Theorem 23. □

For the next two sections, we shall consider p and q to be fixed and let R be as in Theorem 23.

8.2 Low degree reductions

Lemma 1 clearly still applies to the polynomial calculus over rings, but Lemma 4 needs to be reproved for rings. Let q, p, R be as in the previous theorem.

Lemma 24 *Over the ring R, $TS_n(p)$ is (d_1, d_2) reducible to MOD_p^m, where $m = n + nrp$, $d_1 = 2pr$ and $d_2 = 2pr$.*

Proof. The reduction is exactly the same as the reduction used for the proof of Lemma 4. Examination of the definition of r_e in Step 1 of that proof reveals that the only place where inverses were used was in the polynomials (1) and (2) and these were inverses of elements of the form $\omega^{a_1} - \omega^a$ which do exist in R. So it remains to re-do Step 2 of the proof of Lemma 4.

Recall that we must find small degree PC derivations of $MOD_p^m(r_e)$ equations:

1. $r_e^2 - r_e = 0$ for all p-sets e
2. $r_e r_f = 0$ for all e, f such that $e \cap f \neq 0$, $e \neq f$
3. $\sum_{e, u \in e} r_e - 1 = 0$, for all $u \in [m]$.

As discussed before, each single equation is a tautological consequence of a constant number of equations of $TS_n(p)$. We now need to extend the completeness theorem of [6], Theorem 5.2, to apply to the polynomial calculus over R.

Lemma 25 *Let $z_1, \ldots, z_k$ be variables, and $f(\vec{z})$ be a polynomial. Suppose that in the ring R, $f(z_1, \ldots, z_k) = 0$ for all values of $z_1, \ldots, z_k \in \{1, \omega, \omega^2, \ldots, \omega^{p-1}\}$. Then there is PC derivation of $f(\vec{z})$ from the polynomials $z_i^p - 1$, of degree $\leq pk \cdot \deg(f)$.*

Proof. We give the proof for the case $k = 1$ and leave it the reader to formulate the proof by induction for the case $k > 1$. (All the essential difficulties arise already in the case $k = 1$.) Let P_a be the polynomial

$$(z_1 - \omega^0)(z_1 - \omega^1) \cdots (z_1 - \omega^{a-1})(z_1 - \omega^{a+1}) \cdots (z_1 - \omega^{p-1}).$$

Note that $P_a \cdot (z_1 - \omega^a)$ is the polynomial $z_1^p - 1$ (this is immediate from the fact that they are the same polynomial in each field F_{q_i}).

Claim: Let $c = f(\omega^a) \in R$. The polynomial $P_a \cdot (f(z_1) - c)$ is PC derivable from $z_i^p - 1$ in degree $(p-1) \cdot \deg(f)$.

The claim is proved by induction on the size of the polynomial f. The base case where f is a constant is trivial. The second base case where $f(z_1)$ is just z_1 is immediate from the observation above that $P_a \cdot (z_1 - \omega^a) = z_1^p - 1$. The induction steps of addition and multiplication are handled by the following two constructions:

$$\frac{P_a \cdot (f - c) \qquad P_a \cdot (g - d)}{P_a \cdot ((f + g) - (c + d))}$$

and

$$\frac{\dfrac{P_a \cdot (f - c)}{P_a \cdot (fg - cg)} \qquad \dfrac{P_a \cdot (g - d)}{P_a \cdot (cg - cd)}}{P_a \cdot (fg - cd)}$$

and this proves the claim.

Now let P_a^ℓ be the polynomial $\prod_{\substack{i \geq \ell \\ i \neq a}} (z_1 - \omega^i)$. We only use this polynomial when $a \geq \ell$.

Claim: Let $\ell \geq 0$ and let $c = f(\omega^a) \in R$. The polynomial $P_a^\ell \cdot f(z_1)$ is PC derivable from $z_i^p - 1$ in degree $(p-1)\deg(f)$.

The second claim is proved by induction on ℓ. The base case, where $\ell = 0$ is already established by the first claim, since $P_a^0 = P_a$. For the induction step, let $a \geq \ell + 1$. The induction hypothesis tells us that $P_\ell^\ell \cdot f(z_1)$ and $P_a^\ell \cdot f(z_1)$ are both PC derivable. Subtracting these gives

$$(\omega^\ell - \omega^a) P_a^{\ell+1} \cdot f(z_1).$$

Since $(\omega^\ell - \omega^a)$ is invertible in R, we may multiply by $(\omega^\ell - \omega^a)^{-1}$ to derive $P_a^{\ell+1} \cdot f(z_1)$, and the claim is proved.

The base case $k = 1$ of Lemma 24 is immediate from the second claim, with $\ell = p$. The argument for the induction step is similar and is left to the reader. □

8.3 PC lower bound for rings

We now prove the main theorems giving lower bounds the degrees of of PC derivations over $\mathbb{Z}_q$. Fix p, q,R as above.

Theorem 26 *Let G_n be an r-regular graph with expansion ϵ. Then, for all $d < \epsilon n/4$, there is no degree d PC refutation of $TS_n(p)$ over R.*

This plus Lemmas 1 and 24 immediately imply:

Corollary 27 *Let $p, q \geq 2$ be relatively prime. Any PC-refutation over Z_q of the MOD_p^n polynomials requires degree $> \delta n$, for some constant $\delta > 0$.*

The constant δ depends on p and q. To prove Theorem 26, we need merely note that the proof of Theorem 15 still applies: We take $R^* = \{1, \omega, \omega^2, \ldots, \omega^{p-1}\}$ and then, as already noted near the beginning of section 7, the proof of Theorem 15 establishes Theorem 26.

9 Concluding remarks

(1) Our proof of the lower bounds for polynomial calculus proofs of the mod p principles proceeded by first transforming the mod p principle from additive form into the equivalent 'multiplicative' Tseitin principles $TS_n(p)$. In the additive form, variables take on values 0 through $p-1$, whereas in the multiplicative form variables range over powers of the p-th root of unity ω. This transformation into multiplicative form is seemingly necessary, since unsatisfiable sets of degree d binomial polynomials over variables that take on 0/1 values (i.e., where the initial polynomials include $x^2 - x$) have degree $d+1$ polynomial calculus refutations, assuming we are working over a field. This can be proved by noting that $x_1 x_2 \cdots x_k - y_1 y_2 \cdots y_n = 0$ is equivalent to the set of Horn clauses $x_1 \wedge \cdots \wedge x_k \rightarrow y_j$ $(1 \leq j \leq n)$ and $y_1 \wedge \cdots \wedge y_n \rightarrow x_i$ $(1 \leq i \leq k)$. Similarly, $x_1 \cdots x_k - 1 = 0$ is equivalent to the set of unit Horn clauses $\rightarrow x_i$. and the monomial $x_1 \cdots x_k = 0$ is equivalent to the Horn clause $x_1 \wedge \cdots \wedge x_k \rightarrow \perp$. Conversely, a Horn clause $x_1 \wedge \cdots x_k \rightarrow y$ is equivalent to the binomial identity $x_1 \cdots x_k = x_1 \cdots x_k y$; and other Horn clauses can be translated analogously.

Then, if a set of binomial polynomials including $x^2 - x$ for all variables x is unsatisfiable, the SLD-resolution refutation of the equivalent Horn clauses may be used to derive a polynomial calculus refutation with degree no larger than one plus the degree of the initial polynomials. This PC refutation proceeds as follows: for any equation $x_1 x_2 \ldots x_k = a$ with a a non-zero scalar, derive $x_1 = 1$ by multiplying by $(x_1 - 1)$ and combining with $x_1^2 - x_1 = 0$. (If $a \notin \{0, 1\}$, then already a contradiction can be obtained.)

Then x_1 may be "erased" from all terms. This is iterated until an equation $a = 0$ is derived for some non-zero scalar a. If, on the other hand, no such equation is obtained, the polynomials can be satisfied by the truth assignment which gives all processed variables the value 1, and gives the remaining variables the value 0.

(2) Our proofs of the lower bounds for polynomial calculus proofs are closely related to the lower bounds of [9] for Nullstellensatz refutations of binomial systems. Thus a natural question is whether the Nullstellensatz proof system and the polynomial calculus proof systems are equivalent with respect to binomials systems, i.e., whether any degree d polynomial calculus refutation of binomials can be transformed into a degree d Nullstellensatz refutation. This, however, is not the case: [7] have shown that the induction principle is a binomial system which has constant degree polynomial calculus refutations, but requires $\log n$-degree Nullstellensatz proofs, and [10] has obtained the same results for binomial systems expressing a graph principle related to the Tseitin principle. The former separation uses 0/1 valued variables and the latter uses ± 1 valued variables. In addition, the induction principle can be translated into multiplicative form, and this is a binomial system over ± 1 valued variables which has constant degree PC refutations and requires logarithmic degree Nullstellensatz refutations.

(3) The proof of our main result can alternatively be proven by a direct reduction to resolution [4]. In particular, if one begins with binomial equations, plus extra equations $x_i^2 = 1$, then it can be shown that if there is a PC refutation of the equations of degree d, then there is a resolution refutation of clause-width $O(d)$ of the corresponding unsatisfiable formula obtained by converting the binomial equations in the natural way. Thus, our linear lower bounds for the Tseitin graph tautologies can be obtained as a corollary to the corresponding result for resolution of Urquhart [18]. The idea behind the reduction is to first observe that any line in an optimal degree PC refutation over GF_p is a binomial, and thus can be transformed into a linear equation (mod p) where the number of variables in the linear equation is at most twice the PC degree. (See [3].) And secondly, show that any linear equation (mod p) involving d variables can be expressed as a CNF formula of clause width d, and such that each width d CNF formula can be derived from the previous one by a small-width resolution refutation.

Acknowledgement. We thank A. Wadsworth for discussions about ring extensions. We also thank the reviewers for their substantial comments.

References

[1] M. Alekhnovich, E. Ben-Sasson, A. Razborov, and A. Widgerson, *Space complexity in propositional calculus*, in Proceedings of the 32nd Annual ACM Symposium on Theory of Computation, 2000, pp. ???–???

[2] P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi, and P. Pudlák, *Lower bounds on Hilbert's Nullstellensatz and propositional proofs*, Proceedings of the London Mathematical Society, 73 (1996), pp. 1–26.

[3] E. Ben-Sasson and R. Impagliazzo, *Random CNF's are hard for the polynomial calculus*. Typeset manuscript, 1999.

[4] J. Buresh-Oppenheim, M. Clegg, R. Impagliazzo, and T. Pitassi, *Homogenization and the polynomial calculus*. Typeset manuscript, 1999.

[5] S. R. Buss, *Lower bounds on Nullstellensatz proofs via designs*, in Proof Complexity and Feasible Arithmetics, P. Beame and S. Buss, eds., American Mathematical Society, 1998, pp. 59–71.

[6] S. R. Buss, R. Impagliazzo, J. Krajíček, P. Pudlák, A. A. Razborov, and J. Sgall, *Proof complexity in algebraic systems and bounded depth Frege systems with modular counting*, Computational Complexity, 6 (1996/1997), pp. 256–298.

[7] S. R. Buss and T. Pitassi, *Good degree lower bounds on nullstellensatz refutations of the induction principle*, in Proceedings of the Eleventh Annual Conference on Structure in Complexity Theory, IEEE Computer Society, 1996, pp. 233–242. Revised version to appear in *J. Computer and System Sciences*.

[8] M. Clegg, J. Edmonds, and R. Impagliazzo, *Using the Groebner basis algorithm to find proofs of unsatisfiability*, in Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing, 1996, pp. 174–183.

[9] D. Grigoriev, *Nullstellensatz lower bounds for Tseitin tautologies*, in Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science, IEEE Computer Society Press, 1998, pp. 648–652.

[10] ——, *Lower bounds for polynomial calculus proofs for binomial ideals*. Typeset manuscript, 1999.

[11] I. N. Herstein, *Topics in Algebra*, Xerox College Publishing, 2nd ed., 1975.

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

[13] J. Krajíček, *On the degree of ideal membership proofs from uniform families of polynomials over a finite field.* To appear in *Illinois J. of Mathematics.*

[14] ——, *Lower bounds for a proof system with an exponential speed-up over constant depth Frege systems and over polynomial calculus*, in Mathematical Foundations of Computer Science (MFCS), Lecture Notes in Computer Science #1295, Berlin, 1997, Spring-Verlag.

[15] T. Pitassi, *Algebraic propositional proof systems*, in Descriptive Complexity and Finite Models, N. Immerman and P. Kolaitis, eds., DIMACS Series in Discrete Mathematics and Theoretical Computer Science #31, American Mathematics Society, 1996, pp. 215–244.

[16] A. A. Razborov, *Lower bounds for the polynomial calculus*, Computation Complexity, 7 (1998), pp. 291–324.

[17] R. Smolensky, *Algebraic methods in the theory of lower bounds for Boolean circuit complexity*, in Proceedings of the Nineteenth Annual ACM Symposium on the Theory of Computing, ACM Press, 1987, pp. 77–82.

[18] A. Urquhart, *Hard examples for resolution*, J. Assoc. Comput. Mach., 34 (1987), pp. 209–219.