Diagonal Circuit Identity Testing and Lower Bounds

Nitin Saxena[⋆]

Hausdorff Center for Mathematics
Endenicher Allee 60
D-53115 Bonn, Germany
ns@hcm.uni-bonn.de

Abstract. In this paper we give the first deterministic polynomial time algorithm for testing whether a *diagonal* depth-3 circuit $C(x_1, \ldots, x_n)$ (i.e. C is a sum of powers of linear functions) is zero. We also prove an exponential lower bound showing that such a circuit will compute determinant or permanent only if there are exponentially many linear functions. Our techniques generalize to the following new results:

1. Suppose we are given a depth-4 circuit (over any field $\mathbb{F}$) of the form:

$$C(x_1, \ldots, x_n) := \sum_{i=1}^{k} L_{i,1}^{e_{i,1}} \cdots L_{i,s}^{e_{i,s}}$$

   where, each $L_{i,j}$ is a sum of univariate polynomials in $\mathbb{F}[x_1, \ldots, x_n]$. We can test whether C is zero deterministically in $poly(size(C), max_i\{(1 + e_{i,1}) \cdots \ (1 + e_{i,s})\})$ field operations. In particular, this gives a deterministic polynomial time identity test for general depth-3 circuits C when the d :=degree(C) is logarithmic in the size(C).
2. We prove that if the above circuit $C(x_1, \ldots, x_n)$ computes the determinant (or permanent) of an $m \times m$ formal matrix with a "small" $s = o\left(\frac{m}{\log m}\right)$ then $k = 2^{\Omega(m)}$. Our lower bounds work for all fields $\mathbb{F}$. (Previous exponential lower bounds for depth-3 only work for nonzero characteristic.)
3. We also present an exponentially faster identity test for homogeneous diagonal circuits (deterministically in $poly(nk \log(d))$ field operations over finite fields).

Keywords: arithmetic circuit, identity testing, depth 3, depth 4, determinant, permanent, lower bounds.

1 Introduction

Identity Testing is the problem of checking whether a given arithmetic circuit $C(x_1, \ldots, x_n)$, computing a polynomial over a field $\mathbb{F}$, is the zero circuit. Ideally

[⋆] This work was done while the author was a postdoc in Centrum voor Wiskunde en Informatica, Amsterdam under BRICKS AFM1 and NWO VICI grants.

L. Aceto et al. (Eds.): ICALP 2008, Part I, LNCS 5125, pp. 60–71, 2008.

we would like to do identity testing deterministically in time polynomial in the size of the circuit C but no such algorithm is known. The simplest known general algorithm is randomized which was discovered independently by Schwartz [20] and Zippel [22]: it evaluates the given circuit at a random point and accepts if and only if the circuit evaluates to zero at that point. There are more involved randomized algorithms that use fewer random bits [2]. Besides being a natural algebraic problem, special cases of identity testing also appear in primality testing [3], testing equivalence of read-once branching programs [6], graph matching problems [15], interpolating sparse multivariate polynomials [7] and proving complexity theory results such as IP=PSPACE [21], NP=PCP($O(\log n)$, $O(1)$) [5]. Solving identity testing becomes all the more important by the work of Impagliazzo and Kabanets [11] who showed that – finding a deterministic algorithm for identity testing is, roughly, equivalent to proving circuit lower bounds for NEXP.

In this paper we consider arithmetic circuits of depth 4 and solve the identity testing problem for a natural restricted case. Our basic technique is to express the multiplication gate $(a_0+a_1x_1+\cdots+a_nx_n)^d$ in a *dual* form $\sum_j f_{j,1}(x_1)\cdots f_{j,n}(x_n)$. In full generality our dual form expresses a product-of-sum-of-univariates as a sum-of-product-of-univariates effectively (see Remark 1). Our technique of computing the dual is a new way to unfold a multiplication gate in an arithmetic circuit. The dual of a multiplication gate is obtained by using the tools of formal power series (of e^x), polynomial interpolation and working over local algebras. This dual computation is faster than a brute-force expansion and may have other applications. Finally, we also show that in the special case of homogeneous diagonal circuits we can actually do better and give a $poly(nk\log d)$ time identity test.

1.1 Known Results

There are deterministic algorithms known for identity testing only over restricted classes of arithmetic circuits. Raz and Shpilka [19] gave a deterministic polynomial time identity test for noncommutative arithmetic formulas. Dvir and Shpilka [8] attempted a characterization of zero depth-3 circuits and obtained a $poly(n, 2^{\log^{k-1} d})$ time identity test. Kayal and Saxena [14] used Chinese remaindering over local rings and gave a $poly(nd^k)$ time identity test for depth-3 circuits which is clearly a polynomial time identity test if k, the top fanin of the circuit, is bounded. In this work we allow the top fanin to be unbounded but impose the restriction that each multiplication gate has only "few" *distinct* functions as input. All these identity tests are non-black-box, i.e. they look *inside* the circuit instead of just evaluating it at points. Recently, there has been some attempts towards black-box identity testing for depth-3 circuits (see [12]). A black-box identity test even for depth-4 circuits would have important repercussions for the general identity testing problem [4].

In this paper we also prove exponential lower bounds for computing determinant or permanent by certain restricted depth-4 circuits. These restricted depth-4 circuits are the ones for which we give a deterministic polynomial time identity test. Grigoriev, Karpinski and Razborov [9,10] have also shown such lower bounds for general depth-3 circuits but assuming a nonzero characteristic. Our lower bounds are new in the sense that they hold over all fields.

1.2 Definitions and Statement of Results

We will use $poly(M,N)$ to refer to a real function in M and N whose value is upper bounded by $(MN)^{c_1}$ for all $M,N > c_2$ where $c_1, c_2 > 0$ are absolute constants. When using $poly(M,N)$ we will not specify the value of c_1, c_2 as our main interest in this paper is only in their existence. We will use $[n]$ to refer to the set $\{1,\dots,n\}$. We will denote the characteristic of a field $\mathbb{F}$ (i.e. smallest integer $t > 0$ such that $t = 0$ in $\mathbb{F}$ or zero if there is no such t) by $char(\mathbb{F})$. An algebra R over a field $\mathbb{F}$ is simply a ring containing $\mathbb{F}$. In this paper only finite dimensional commutative algebras appear, i.e. there is an integer $N > 0$ and basis elements $b_1,\dots,b_N \in R$ such that any element in R can be uniquely expressed as $\sum_{i=1}^{N} \alpha_i b_i$ with α_i's in $\mathbb{F}$. We call N the *dimension* of the algebra R over $\mathbb{F}$, denoted by $dim(R)$. It is a simple exercise to see that basic operations (e.g. multiplication of two elements) in R can be done using $poly(N)$ field operations (in $\mathbb{F}$).

Our main concern in this paper are depth-3 (or depth-4) circuits. For the purposes of identity testing (also lower bounds for determinant and permanent) the hardest case is when the circuit has an addition gate at the top. These circuits are called $\Sigma\Pi\Sigma$ (or $\Sigma\Pi\Sigma\Pi$). It is clear that the output of such a $\Sigma\Pi\Sigma$ circuit $C(x_1,\dots,x_n)$ would be: $\sum_{i=1}^{k} \ell_{i,1}\cdots\ell_{i,d_i}$, where the $\ell_{i,j} = (a_{i,j,0} + a_{i,j,1}x_1 + \cdots + a_{i,j,n}x_n)$ are linear functions over a field $\mathbb{F}$. We call k the top fanin of C, d_i the degree of the i-th multiplication gate and $d = max_i\{d_i\}$ the degree of C. The size of an arithmetic circuit is the number of addition, multiplication and input gates in its representation as a directed acyclic graph. Clearly, in the above setting $size(C)$ is dominated by knd. It is easy to see that by brute-force we can check whether a $\Sigma\Pi\Sigma$ circuit C is a zero circuit in time polynomial in $k \cdot \binom{n+d}{d}$ but this is generally exponential in $size(C)$.

In this paper we start with the case where each of the multiplication gates in C has only one distinct linear function as input. We call such a C a *diagonal* circuit and it looks like: $C(x_1,\dots,x_n) = \sum_{i=1}^{k} b_i \cdot \ell_i^{d_i}$, where the b_i's are in $\mathbb{F}$ and the ℓ_i's are linear functions. Our techniques extend upto depth-4 circuits of the form:

$$C(x_1,\dots,x_n) = \sum_{i=1}^{k} L_{i,1}^{e_{i,1}} \cdots L_{i,s}^{e_{i,s}} \qquad (1)$$

where the $L_{i,j}$'s are not linear functions but sums of univariate polynomials, i.e. for all $i \in [k], j \in [s]$:

$$L_{i,j}(x_1,\dots,x_n) = g_{i,j,1}(x_1) + \cdots + g_{i,j,n}(x_n)$$

where $g_{i,j,j'} \in \mathbb{F}[x_{j'}]$. Our first main theorem is:

Theorem 1. *Over any field $\mathbb{F}$, let C be a circuit given as in Equation (1). Then we can deterministically check whether C is a zero circuit in $poly(size(C), max_i\{ (1+e_{i,1})\cdots(1+e_{i,s})\})$ field operations.*

Thus, when s is constant or when s is logarithmic but $e_{i,j}$'s are constants we get a deterministic polynomial time identity test.

The lower bounds that we get, basically show that if a depth-3 circuit (or a restricted depth-4 circuit) computes determinant (or permanent) then either some of the multiplication gates have "lots" of distinct functions as inputs or the top fanin of the circuit is exponential. Our second main theorem is:

Theorem 2. *Over any field $\mathbb{F}$, if the circuit in Equation (1) expresses the determinant (or permanent) of a general $m \times m$ matrix with parameters $s = o\left(\frac{m}{\log m}\right)$, $n = m^2$ and $d = poly(m)$ then $k = 2^{\Omega(m)}$.*

Note that determinant (or permanent) of an $m \times m$ matrix is just a sum of $m!$ monomials. A monomial $y_1 \cdots y_m$ can be expressed as a sum of powers of $2^{O(m)}$ linear forms. Hence, determinant can be expressed by a sum of powers of at most $O(m!)$ linear forms and our lower bounds show that this is almost tight.

1.3 Our Techniques

We use non-black-box methods, i.e. we heavily use the structure of the given circuit. We use tools that previously have been used to understand noncommutative formulas, for example by Nisan, Wigderson [16,17], Raz and Shpilka [19]. We apply these old tools in a nontrivial way to understand depth-3 and depth-4 (commutative) circuits. For clarity let us note here the two old theorems in a generalized form.

A circuit $D(x_1, \ldots, x_n)$, over an algebra R over a field $\mathbb{F}$, is called noncommutative if each of its multiplication gate has ordered inputs and the variables $x_1, \ldots, x_n$ do not commute, i.e. for all $i \neq j$, $x_i \cdot x_j \neq x_j \cdot x_i$. The output $D(x_1, \ldots, x_n)$ is a formal expression in the ring $R\{x_1, \ldots, x_n\}$ of polynomials over noncommutative variables $x_1, \ldots, x_n$. Clearly, any commutative circuit $C(x_1, \ldots, x_n)$ can be turned into a noncommutative circuit $\tilde{C}(x_1, \ldots, x_n)$ by imposing an order on the inputs to its multiplication gates and assuming $x_i \cdot x_j \neq x_j \cdot x_i$ for all $i \neq j$. But now circuits C and $\tilde{C}$ are computing different polynomials and it may happen that C is a zero circuit but $\tilde{C}$ is a nonzero circuit. However, if $\tilde{C}$ is a zero circuit then C is surely a zero circuit as well. A circuit is called a *formula* if the fan-out of every gate in the circuit is at most one. Noncommutative formulas are easier to analyze compared to the commutative ones and the following identity test is relevant to us:

Theorem 3 (Theorem 2.5 of [19] generalized over algebras). *Let R be an algebra over a field $\mathbb{F}$. Given a noncommutative formula $C(x_1, \ldots, x_n) \in R\{x_1, \ldots, x_n\}$ we can verify deterministically in $poly(size(C), dim(R))$ field operations whether C is zero.*

The second result relevant to us is an extension of Theorem 5.1 of [19] that proves lower bounds for pure circuits using the partial derivative space (see the proof idea in Lemma 5.3 of [19]).

Theorem 4 (Theorem 5.1 of [19] generalized over algebras). *Let R be an algebra over a field $\mathbb{F}$, $r \in R \setminus \{0\}$, $r' \in R$ and let $det(x_{1,1}, \ldots, x_{n,n})$ denote the determinant of a formal $n \times n$ matrix $((x_{i,j}))$. If $det(x_{1,1}, \ldots, x_{n,n}) \cdot r - r'$ can be expressed as a circuit:*

$$C(x_{1,1}, \ldots, x_{n,n}) = \sum_{i=1}^{k} f_{i,1,1}(x_{1,1}) \cdots f_{i,n,n}(x_{n,n})$$

where the f_{i,j_1,j_2}'s are univariate polynomials over R, then $k \cdot dim(R) = 2^{\Omega(n)}$. A similar lower bound holds for the permanent as well.

Proof (Sketch). Since determinant is a multilinear polynomial we can ignore the nonlinear terms in the f_{i,j_1,j_2}'s. Now if we look at the suitably defined partial derivative space (as in [19]) of the circuit C then it has rank, over $\mathbb{F}$, at most $k \cdot dim(R)$ because there are k multiplication gates and the coefficients in f_{i,j_1,j_2}'s are themselves of dimension $dim(R)$ over $\mathbb{F}$. On the other hand it is known that the corresponding rank of determinant is $2^{\Omega(n)}$.

Our main contribution is a novel way to transform the multiplication gates of a circuit, hence the overall circuit, to a form on which we can apply Theorems 3 and 4. Our basic technique is to use the formal power series $e^x = 1 + x + \frac{x^2}{2!} + \cdots$ and polynomial interpolation to express the multiplication gate $(a_0 + a_1x_1 + \cdots + a_nx_n)^d$ in a *dual* form: $\sum_j f_{j,1}(x_1) \cdots f_{j,n}(x_n)$. Now this is a nice circuit as the variables $x_1, \ldots, x_n$ in it are "separated" and we can invoke the known tricks. For example, it can be viewed as a circuit in which the variables $x_1, \ldots, x_n$ do not commute, thus by Theorem 3 we get a deterministic polynomial time identity test for diagonal circuits. Also, by the lower bounds of Theorem 4 we get that a diagonal circuit can compute determinant or permanent only if it is of exponential size. These ideas generalize to circuits with $s > 1$ in Equation (1) but require more algebraic sophistication as then we work with the formal power series on larger local algebras instead of working on the base field $\mathbb{F}$.

1.4 Organization

The paper is organized as follows. In section 2 we present our results for the basic case of diagonal circuits over zero characteristic. In section 3 we show how to extend our results to restricted depth-4 circuits over zero characteristic. In section 4 we extend the previous results to nonzero characteristic. Finally, in section 5 we present an exponentially faster identity test for homogeneous diagonal circuits (deterministically in $poly(nk\log(d))$ field operations over finite fields). Some of the proofs have been omitted from the extended abstract due to space constraints.

2 Diagonal Depth-3 Circuits

The aim of this section is to define a dual expression for multiplication gates of the form $(a_0 + a_1x_1 + \cdots + a_nx_n)^d$ and use that form to give an identity test for diagonal circuits and to prove lower bounds. We will assume throughout this section that the base field $\mathbb{F}$ is of characteristic zero. We will use the fairly standard notation $[m]f(x_1, \ldots, x_n)$ to denote the coefficient of the monomial m in a polynomial (more generally, a power series) f. For example, $[xyz](x + y + z)^3 = 6$.

2.1 Dual of a Multiplication Gate

The following lemma formalizes and computes the dual of an affine power.

Lemma 1. *Let $a_0, a_1, \ldots, a_n$ be in a field $\mathbb{F}$ of zero characteristic. Then we can compute univariate polynomials $f_{i,j}$'s in poly(nd) field operations such that for $t = (nd + d + 1)$:*

$$(a_0 + a_1x_1 + \cdots + a_nx_n)^d \;=\; \sum_{i=1}^{t} f_{i,1}(x_1) \cdots f_{i,n}(x_n)$$

Proof. We will prove this using the formal power series: $\exp(x) = 1 + x + \frac{x^2}{2!} + \cdots$, where $\exp(x) = e^x$ and e is the base of natural logarithm. Define the degree d truncation of the series to be $E_d(x) = 1 + x + \cdots + \frac{x^d}{d!}$. Observe that:

$$\begin{aligned}(d!)^{-1} \cdot (a_0 + a_1x_1 + \cdots + a_nx_n)^d \;&=\; [z^d] \exp\left((a_0 + a_1x_1 + \cdots + a_nx_n) \cdot z\right)\\ &=\; [z^d] \exp(a_0z) \cdot \exp(a_1x_1z) \cdots \exp(a_nx_nz)\\ &=\; [z^d]\; E_d(a_0z) \cdot E_d(a_1x_1z) \cdots E_d(a_nx_nz)\end{aligned}$$

The product $E_d(a_0z) \cdot E_d(a_1x_1z) \cdots E_d(a_nx_nz)$ can be viewed as a univariate polynomial in z of degree $(nd+d)$. Hence, its coefficient of z^d can be computed by evaluating the polynomial at $(nd+d+1)$ distinct points $\alpha_1, \ldots, \alpha_{nd+d+1} \in \mathbb{F}$ (remember $\mathbb{F}$ is large enough) and by interpolation we can compute $\beta_1, \ldots, \beta_{nd+d+1} \in \mathbb{F}$ such that:

$$\begin{aligned}&[z^d]\; E_d(a_0z) \cdot E_d(a_1x_1z) \cdots E_d(a_nx_nz)\\ &= \sum_{i=1}^{nd+d+1} \beta_i \cdot E_d(a_0\alpha_i) \cdot E_d(a_1\alpha_ix_1) \cdots E_d(a_n\alpha_ix_n)\end{aligned}$$

This is the dual form of $(a_0+a_1x_1+\cdots+a_nx_n)^d$ as required. It is routine to verify that all the univariate polynomials $E_d(\cdot)$ in the above sum can be computed in *poly*(nd) field operations.

2.2 Identity Test and Lower Bounds

The dual form of multiplication gates obtained in Lemma 1 is easy to analyze. We give the ideas in the following theorems.

Theorem 5. *Over zero characteristic, identity testing for diagonal circuits can be done in deterministic polynomial time (poly(nkd) field operations).*

Proof. Suppose we are given a diagonal circuit C:

$$C(x_1, \ldots, x_n) = \sum_{i=1}^{k} b_i \cdot \ell_i^{d_i}$$

Then by Lemma 1 we can compute the dual form of each of the k multiplication gates such that:

$$C(x_1, \ldots, x_n) = \sum_{i=1}^{k} \sum_{j=1}^{nd_i+d_i+1} f_{i,j,1}(x_1) \cdots f_{i,j,n}(x_n) \tag{2}$$

where the univariate polynomials $f_{i,j,j'}$'s are of degree at most d_i.

Now observe that the variables in the circuit on the RHS of Equation (2) can be assumed to be noncommutative without affecting the output, i.e. circuit C. Thus, if we apply the identity testing algorithm of Theorem 3 to the circuit on the RHS of Equation (2) we will correctly know whether C is zero or not. Hence, C can be verified for zeroness deterministically in $poly(nkd)$ field operations.

Theorem 6. *Over zero characteristic, if a diagonal circuit expresses the determinant (or permanent) of a formal $m \times m$ matrix with $n = m^2$ variables and degree $d = poly(m)$ then the top fanin $k = 2^{\Omega(m)}$.*

Proof. Suppose a diagonal circuit C computes the determinant of a general $m \times m$ matrix. Then by Lemma 1 determinant is being computed by a circuit as given in Equation (2). Now the exponential lower bound of Theorem 4 applies and we get that $poly(ndk) = 2^{\Omega(m)}$ implying $k = 2^{\Omega(m)}$.

3 Extension to Restricted Depth-4 Circuits

In this section we extend the results of the last section to depth-4 circuits (with some success). The starting point is a dual expression for multiplication gates of the form $L_1^{e_1} \cdots L_s^{e_s}$ where the L_i's are sums of univariate polynomials in $\mathbb{F}[x_1, \ldots, x_n]$. The proof is along the same lines as presented before but now we will work in local algebras over $\mathbb{F}$. Finally, we use that form to give identity test and prove lower bounds. We will again assume throughout this section that the base field $\mathbb{F}$ is of characteristic zero.

3.1 Dual of a Multiplication Gate

We compute the dual form of a multiplication gate of the form:

$$M(x_1, \ldots, x_n) = (g_{1,1}(x_1) + \cdots + g_{1,n}(x_n))^{e_1} \cdots (g_{s,1}(x_1) + \cdots + g_{s,n}(x_n))^{e_s} \tag{3}$$

which means that we express M as an expression: $\sum_{i=1}^{t} f_{i,1}(x_1) \cdots\ f_{i,n}(x_n)$ where the $f_{i,j}$'s are univariate polynomials over an $\mathbb{F}$-algebra R (unlike the diagonal case where we worked over $\mathbb{F}$). This expression with variables $x_1, \ldots, x_n$ "separated" we call a *dual* of the multiplication gate. The following lemma shows that such a dual is computable but we pay a price in terms of the dimension of algebra R which is $(e_1 + 1) \cdots (e_s + 1)$.

Lemma 2. *Let $M(x_1,\ldots,x_n)$ be the multiplication gate of Equation (3) over a field $\mathbb{F}$ of zero characteristic and $e=(e_1+\cdots+e_s)$. Then we can compute univariate polynomials $f_{i,j}$'s over an algebra $R:=\mathbb{F}[z_1,\ldots,z_s]/(z_1^{e_1+1},\ldots,z_s^{e_s+1})$ in poly(size(M), dim(R)) field operations such that for $t=(ne+1)$:*

$$M(x_1,\ldots,x_n)\cdot z_1^{e_1}\cdots z_s^{e_s} \;=\; \sum_{i=1}^{t} f_{i,1}(x_1)\cdots f_{i,n}(x_n) \quad \textit{over } R$$

Remark 1. Note that we can informally describe the above equation as: a product-of-sums-of-univariates can be written as a sum-of-products-of-univariates. This justifies our continued usage of the phrase "dual form".

Proof. We will again prove this using the formal power series: $\exp(x)=1+x+\frac{x^2}{2!}+\cdots$, where $\exp(x)=e^x$ and e is the base of natural logarithm. Recall that the degree d truncation of the series is $E_d(x)=1+x+\cdots+\frac{x^d}{d!}$. Let $L_1,\ldots,L_s$ be the distinct factors of M (that are now not linear functions but sums of univariate polynomials). Observe that:

$$\begin{aligned}
(e_1!\cdots e_s!)^{-1}\cdot L_1^{e_1}\cdots L_s^{e_s} \;&=\; [z^e z_1^{e_1}\cdots z_s^{e_s}]\ \exp(L_1z_1z)\cdots\exp(L_sz_sz)\\
&=\; [z^e z_1^{e_1}\cdots z_s^{e_s}]\ \exp(L_1z_1z+\cdots+L_sz_sz)\\
&=\; [z^e z_1^{e_1}\cdots z_s^{e_s}]\ \exp\left((g_{1,1}z_1+\cdots+g_{s,1}z_s)z\right)\cdots\\
&\qquad\cdots\exp\left((g_{1,n}z_1+\cdots+g_{s,n}z_s)z\right)\\
&=\; [z^e z_1^{e_1}\cdots z_s^{e_s}]\ E_e\left((g_{1,1}z_1+\cdots+g_{s,1}z_s)z\right)\cdots\\
&\qquad\cdots E_e\left((g_{1,n}z_1+\cdots+g_{s,n}z_s)z\right) \qquad (4)
\end{aligned}$$

Note that the last product can be viewed as a univariate polynomial in z of degree ne. Hence, its coefficient of z^e can be computed by evaluating the polynomial at $(ne+1)$ distinct points $\alpha_1,\ldots,\alpha_{ne+1}\in\mathbb{F}$ (remember that $\mathbb{F}$ is large enough) and by interpolation we can compute $\beta_1,\ldots,\beta_{ne+1}\in\mathbb{F}$ such that:

$$\begin{aligned}
&[z^e z_1^{e_1}\cdots z_s^{e_s}]\ E_e\left((g_{1,1}z_1+\cdots+g_{s,1}z_s)z\right)\cdots E_e\left((g_{1,n}z_1+\cdots+g_{s,n}z_s)z\right)\\
&\quad=\;[z_1^{e_1}\cdots z_s^{e_s}]\ \sum_{i=1}^{ne+1}\beta_i\cdot E_e\left((g_{1,1}z_1+\cdots+g_{s,1}z_s)\alpha_i\right)\cdots\\
&\qquad\qquad\cdots E_e\left((g_{1,n}z_1+\cdots+g_{s,n}z_s)\alpha_i\right)
\end{aligned}$$

Notice that the monomials having nonzero coefficients in the above sum are of the form $z_1^{t_1}\cdots z_s^{t_s}$ such that $t_1+\cdots+t_s=e=e_1+\cdots+e_s$. Thus, if we look at the above sum modulo the ideal $(z_1^{e_1+1},\ldots,z_s^{e_s+1})$ then the surviving monomials $z_1^{t_1}\cdots z_s^{t_s}$ would be those that have $t_1\leqslant e_1,\ldots,t_s\leqslant e_s$ which together with $t_1+\cdots+t_s=e=e_1+\cdots+e_s$ uniquely determines the surviving monomial as $z_1^{e_1}\cdots z_s^{e_s}$. Consequently, we can summarize the above computations as, over R:

$$\begin{aligned}
&(e_1!\cdots e_s!)^{-1}\cdot L_1^{e_1}\cdots L_s^{e_s}\cdot z_1^{e_1}\cdots z_s^{e_s}\\
&\quad=\;\sum_{i=1}^{ne+1}\beta_i\cdot E_e\left((g_{1,1}z_1+\cdots+g_{s,1}z_s)\alpha_i\right)\cdots E_e\left((g_{1,n}z_1+\cdots+g_{s,n}z_s)\alpha_i\right)\ .
\end{aligned}$$

This is the dual form of M as required. Notice that there is a nonconstant factor $z_1^{e_1} \cdots z_s^{e_s}$ appearing on the LHS but since this factor is a nonzero element of the algebra R, the dual form will be good enough for our purposes. It is routine to verify that the univariate polynomials $E_e(\cdot)$ over R in this sum can be computed in $poly(size(M), dim(R))$ field operations and that the dimension of R is $(e_1 + 1) \cdots (e_s + 1)$.

3.2 Identity Test and Lower Bounds

We can now apply the dual form of Lemma 2 to k multiplication gates and work on a bigger algebra. We formalize this idea in the following theorems.

Theorem 7. *Given a circuit C over a field $\mathbb{F}$ of zero characteristic:*

$$C(x_1, \ldots, x_n) = \sum_{i=1}^{k} L_{i,1}^{e_{i,1}} \cdots L_{i,s}^{e_{i,s}}$$

where the $L_{i,j}$'s are sums of univariate polynomials and (wlog) for all $i, e_{i,1} \neq 0$. We can test whether C is a zero circuit deterministically in $poly(size(C), max_i\{ (1 + e_{i,1}) \cdots (1 + e_{i,s})\})$ field operations.

Proof. Let us apply the dual form of Lemma 2 to the i-th multiplication gate M_i, with $e_i := (e_{i,1} + \cdots + e_{i,s})$, and compute the univariate polynomials f_{i,j_1,j_2}'s, for all $1 \leqslant j_1 \leqslant t_i = (ne_i + 1)$ and $j_2 \in [n]$, over the algebra $R_i := \mathbb{F}[z_{i,1}, \ldots, z_{i,s}]/(z_{i,1}^{e_{i,1}+1}, \ldots, z_{i,s}^{e_{i,s}+1})$ in $poly(size(M_i), dim(R_i))$ field operations such that:

$$L_{i,1}^{e_{i,1}} \cdots L_{i,s}^{e_{i,s}} \cdot z_{i,1}^{e_{i,1}} \cdots z_{i,s}^{e_{i,s}} \;=\; \sum_{j_1=1}^{t_i} f_{i,j_1,1}(x_1) \cdots f_{i,j_1,n}(x_n) \quad \text{over } R_i \tag{5}$$

With the aim of getting a dual form of the circuit C let us define a commutative algebra R that contains the algebras corresponding to each multiplication gate, i.e. $R_1, \ldots, R_k$, as "orthogonal" subalgebras and in which the following $(k-1)$ relations hold: $z_{1,1}^{e_{1,1}} \cdots z_{1,s}^{e_{1,s}} = \cdots = z_{k,1}^{e_{k,1}} \cdots z_{k,s}^{e_{k,s}}$. Explicitly, the algebra R is: $\mathbb{F}[z_{i,j} \mid \forall i \in [k], \forall j \in [s]]/\mathcal{I}$, where the ideal $\mathcal{I}$ is generated by the following three sets of relations:

- $z_{i,j}^{e_{i,j}+1} = 0$, for all $i \in [k], j \in [s]$.
- $z_{i,j} \cdot z_{i',j'} = 0$, whenever $i \neq i'$.
- $z_{i,1}^{e_{i,1}} \cdots z_{i,s}^{e_{i,s}} = z_{i',1}^{e_{i',1}} \cdots z_{i',s}^{e_{i',s}}$, for all $i, i' \in [k]$.

Note that the first set of relations just make $R_1, \ldots, R_k$ as subalgebras of R while the other two sets impose relations on certain zero-divisors in R ($e_{i,1} \neq 0$ implies that $z_{i,1}^{e_{i,1}} \cdots z_{i,s}^{e_{i,s}}$ is a zero-divisor of R). The second set of relations are put in so that the dimension of R gets down to roughly sum of the dimensions of $R_1, \ldots, R_k$. Note that the dimension of R over the base field $\mathbb{F}$ is exactly $\sum_{i=1}^{k}(1 + e_{i,1}) \cdots (1 + e_{i,s}) - 2(k-1)$ which is nonzero.

Now by using the third set of relations in R and summing up Equation (5) for all the k multiplication gates, we get over the algebra R:

$$\begin{aligned} C(x_1,\ldots,x_n)\cdot z_{1,1}^{e_{1,1}}\cdots z_{1,s}^{e_{1,s}} &= \sum_{i=1}^{k} L_{i,1}^{e_{i,1}}\cdots L_{i,s}^{e_{i,s}}\cdot z_{i,1}^{e_{i,1}}\cdots z_{i,s}^{e_{i,s}} \\ &= \sum_{i=1}^{k}\sum_{j_1=1}^{t_i} f_{i,j_1,1}(x_1)\cdots f_{i,j_1,n}(x_n) \qquad (6) \end{aligned}$$

This last expression can be viewed as a noncommutative formula in variables $x_1,\ldots,x_n$ over the algebra R. Clearly, it is zero *iff* $C(x_1,\ldots,x_n)\cdot z_{1,1}^{e_{1,1}}\cdots z_{1,s}^{e_{1,s}}$ is zero over R *iff* C is zero over $\mathbb{F}$. Thus, it is sufficient to test the circuit on the RHS of Equation (6) for zeroness. This can be done by applying the identity testing algorithm of Theorem 3, now working over the algebra R. Hence, we can deterministically verify whether C is zero in $poly(size(C), dim(R))$ field operations as required.

Theorem 8. *Following the notation of the last theorem, if C expresses the determinant (or permanent) of a formal $m\times m$ matrix with parameters $s = o\left(\frac{m}{\log m}\right)$, $n=m^2$ and $(e_1+\cdots+e_k) := e = poly(m)$ then $k = 2^{\Omega(m)}$.*

Proof. Suppose the circuit C computes the determinant of a general $m\times m$ matrix. Recall that C has a dual form as given in Equation (6). Thus, we can apply Theorem 4 to deduce that $poly(nek, dim(R)) = 2^{\Omega(m)}$ implying:

$$poly\left(nek, max_i\{(1+e_{i,1})\cdots(1+e_{i,s})\}\right) = 2^{\Omega(m)}$$

As the $e_{i,j}$'s are at most $poly(m)$ the above implies $poly(nek, m^s) = 2^{\Omega(m)}$ which using the hypothesis further implies $k = 2^{\Omega(m)}$.

4 Extension to the Nonzero Characteristic Case

In the last section we defined the dual form of a multiplication gate $L_1^{e_1}\cdots L_s^{e_s}$, where the L_i's are sums of univariates over a field $\mathbb{F}$ of zero characteristic. In this section we note how to obtain the dual form when the characteristic of $\mathbb{F}$ is a prime $p > 1$. Note that over such a field the expressions used in the proof of Lemma 2 may not be well defined, for example if $p|d!$ then $\frac{1}{d!}$ is undefined in $\mathbb{F}$. We can show that such issues can be taken care of by working in a local algebra over a *Galois ring* of characteristic p^b instead of working over $\mathbb{F}$. This finishes the proofs of our main Theorems 1 and 2.

5 A Faster Identity Test for Diagonal Circuits

Identity testing for homogeneous diagonal circuits can be made exponentially faster in the degree d of the circuit. Unfortunately, we only know how to do this over a finite field $\mathbb{F}$ with an extra assumption that $d < char(\mathbb{F})$ (we do believe it should be possible to do this in general). The main idea to speed up the identity test is that if the degree d of a diagonal circuit C is large compared to fanin k then an argument using Vandermonde's matrix shows that C can be zero only if each multiplication gate is zero, which can be tested in time $poly(nk\log(d))$. Thus, wlog we can assume $d \leqslant k$ and the identity test given in this paper tests $\sum_{i=1}^{k} b_i \cdot \ell_i^d = 0$ deterministically in $poly(nk)$ field operations.

6 Conclusion

In this work we gave a deterministic polynomial time identity test for restricted depth-4 circuits. Our basic idea was to define a dual operation on the multiplication gates in a depth-3 circuit that converts a product gate into a sum of product of univariate polynomials over a local algebra. This dual is efficiently computable when the multiplication gate has "few" distinct linear functions as input. In the case of a general multiplication gate of a depth-3 circuit of degree d the dual computation takes exponential time: $poly(n2^d)$. This dual computation can be viewed as a new way to unfold a given depth-3 circuit better than the direct brute-force expansion. We leave it as an open question to improve this duality to solve the identity testing problem for general depth-3 circuits.

Kayal [13] has observed that Theorems 1 and 2 for depth-3 circuits can also be obtained (nontrivially) by using the space of partial derivatives first defined by Nisan and Wigderson [17]. The basic reason is that the space of partial derivatives of a diagonal circuit has "low" rank and this can be exploited to give an identity test and proving lower bounds. However, in the case of our restricted depth-4 circuits the space of partial derivatives typically has "high" rank. For example, the partial derivative space of $(x_1^2 + \cdots + x_n^2)^n$ is of rank more than 2^n. Thus, the dual form analyzes the restricted depth-4 circuits in ways stronger than the partial derivative space.

Acknowledgements

I would like to thank Harry Buhrman and Ronald de Wolf for various useful discussions during the course of this work. I thank Neeraj Kayal and Avi Wigderson for providing insightful observations on a preliminary version of this paper.

References

1. Adleman, L.M., Lenstra, H.W.: Finding irreducible polynomials over finite fields. In: 18th ACM Symposium on Theory of Computing, pp. 350–355. ACM Press, New York (1986)
2. Agrawal, M., Biswas, S.: Primality and identity testing via Chinese remaindering. Journal of the ACM 50(4), 429–443 (2003)
3. Agrawal, M., Kayal, N., Saxena, N.: Primes is in P. Annals of Mathematics 160(2), 781–793 (2004)

4. Agrawal, M., Vinay, V.: Arithmetic Circuits: A Chasm at Depth Four (preprint, 2008)
5. Arora, S., Safra, S.: Probabilistic Checking of Proofs: A New Characterization of NP. Journal of the ACM 45(1), 70–122 (1998)
6. Blum, M., Chandra, A.K., Wegman, M.N.: Equivalence of free Boolean graphs can be tested in polynomial time. Information Processing Letters 10, 80–82 (1980)
7. Clausen, M., Dress, A., Grabmeier, J., Karpinski, M.: On zero-testing and interpolation of k-sparse multivariate polynomials over finite fields. Theoretical Computer Science 84(2), 151–164 (1991)
8. Dvir, Z., Shpilka, A.: Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits. SIAM J. Comput. 36(5), 1404–1434 (2007)
9. Grigoriev, D., Karpinski, M.: An Exponential Lower Bound for Depth 3 Arithmetic Circuits. In: 30th ACM Symposium on Theory of Computing, pp. 577–582. ACM Press, New York (1998)
10. Grigoriev, D., Razborov, A.A.: Exponential Lower Bounds for Depth 3 Arithmetic Circuits in Algebras of Functions over Finite Fields. Appl. Algebra Eng. Commun. Comput. 10(6), 465–487 (2000)
11. Impagliazzo, R., Kabanets, V.: Derandomizing polynomial identity tests means proving circuit lower bounds. Computational Complexity 13(1/2), 1–46 (2004)
12. Karnin, Z., Shpilka, A.: Deterministic black box polynomial identity testing of depth-3 arithmetic circuits with bounded top fan-in. ECCC Report TR07-042 (2007)
13. Kayal, N.: Private Communication. Summer (2007)
14. Kayal, N., Saxena, N.: Polynomial Identity Testing for Depth 3 Circuits. Computational Complexity 16(2), 115–138 (2007)
15. Lovasz, L.: On determinants, matchings, and random algorithms. In: Fundamentals of Computing Theory, pp. 565–574. Akademia-Verlag (1979)
16. Nisan, N.: Lower bounds for non-commutative computation. In: 23rd ACM Symposium on Theory of Computing, pp. 410–418. ACM Press, New York (1991)
17. Nisan, N., Wigderson, A.: Lower bounds on arithmetic circuits via partial derivatives. Computational Complexity 6(3), 217–234 (1997)
18. Raz, R.: Multi-linear formulas for permanent and determinant are of super-polynomial size. In: 36th ACM Symposium on Theory of Computing, pp. 633–641. ACM Press, New York (2004)
19. Raz, R., Shpilka, A.: Deterministic polynomial identity testing in non-commutative models. Computational Complexity 14(1), 1–19 (2005)
20. Schwartz, J.T.: Fast Probabilistic Algorithms for Verification of Polynomial Identities. Journal of the ACM 27(4), 701–717 (1980)
21. Shamir, A.: IP=PSPACE. Journal of the ACM 39(4), 869–877 (1992)
22. Zippel, R.: Probabilistic Algorithms for Sparse Polynomials. In: International Symposium on Symbolic and Algebraic Computation, pp. 216–226. Springer, Heidelberg (1979)