



# Arithmetic circuits: the chasm at depth four gets wider

Pascal Koiran

## ► To cite this version:

| Pascal Koiran. Arithmetic circuits: the chasm at depth four gets wider. 2010. ensl-00494642v1

**HAL Id: ensl-00494642**

**<https://ens-lyon.hal.science/ensl-00494642v1>**

Preprint submitted on 23 Jun 2010 (v1), last revised 22 Mar 2012 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Arithmetic Circuits: The Chasm at Depth Four Gets Wider

Pascal Koiran

LIP\*, École Normale Supérieure de Lyon, Université de Lyon  
Department of Computer Science, University of Toronto<sup>†</sup>  
`Pascal.Koiran@ens-lyon.fr`

June 23, 2010

## Abstract

In their paper on the “chasm at depth four”, Agrawal and Vinay have shown that polynomials in  $m$  variables of degree  $O(m)$  which admit arithmetic circuits of size  $2^{o(m)}$  also admit arithmetic circuits of depth four and size  $2^{o(m)}$ . This theorem shows that for problems such as arithmetic circuit lower bounds or black-box derandomization of identity testing, the case of depth four circuits is in a certain sense the general case.

In this paper we show that smaller depth four circuits can be obtained if we start from polynomial size arithmetic circuits. For instance, we show that if the permanent of  $n \times n$  matrices has circuits of size polynomial in  $n$ , then it also has depth 4 circuits of size  $n^{O(\sqrt{n} \log n)}$ . Our depth four circuits use integer constants of polynomial size. These results have potential applications to lower bounds and deterministic identity testing, in particular for sums of products of sparse univariate polynomials.

---

\*UMR 5668 ENS Lyon, CNRS, UCBL, INRIA.

<sup>†</sup>A part of this work was done during a visit to the Fields Institute.

# 1 Introduction

Agarwal and Vinay have shown that polynomials of degree  $d = O(m)$  in  $m$  variables which admit nontrivial arithmetic circuits also admit nontrivial arithmetic circuits of depth four [1]. Here, “nontrivial” means of size  $2^{o(d+d \log \frac{m}{d})}$ . The resulting depth 4 circuits are  $\sum \prod \sum \prod$  arithmetic formulas: the output gate (at depth 4) and the gates at depth 2 are addition gates, and the other gates are multiplication gates. This theorem shows that for problems such as arithmetic circuit lower bounds or black-box derandomization of identity testing, the case of depth four circuits is in a certain sense the general case.

But what if we start from arithmetic circuits of size smaller than  $2^{o(m)}$  (for instance, of size polynomial in  $m$ ) ? It is reasonable to expect that the size of the corresponding depth four circuits will be reduced accordingly, but such a result cannot be found in [1]. One of the main results of this paper is a depth reduction theorem for VP families (i.e., families  $(f_n)$  of polynomials of degree and arithmetic circuit complexity polynomially bounded in  $n$ ). We show in Theorem 5 that any VP family  $(f_n)$  has depth 4 arithmetic formulas of size  $n^{O(\sqrt{d_n} \log d_n)}$ , where  $d_n$  is the degree of  $f_n$ . For instance, this result shows that if the permanent of  $n \times n$  matrices has circuits of size polynomial in  $n$ , then it also has depth 4 formulas of size  $n^{O(\sqrt{n} \log n)}$ . This is potentially useful for a lower bound proof: to show that the permanent does not have polynomial size circuits, we “only” have to show that it does not have depth 4 formulas of size  $n^{O(\sqrt{n} \log n)}$ . This is still certainly far away from the known lower bounds for constant depth arithmetic circuits: currently we have superpolynomial lower bound for the permanent for circuits of depth 3 only, and only in finite fields [5, 6]. In the restricted setting of multilinear arithmetic circuits, superpolynomial lower bounds can be obtained for circuits of arbitrary constant depth [14].

As a second contribution we perform an analysis of the size of the integer constants used by the depth 4 circuit simulating a given polynomial size circuit (a similar analysis for the construction in [1] has not been carried out yet to the author’s knowledge). Roughly speaking, we show that reduction to depth 4 does not require the introduction of large constants. In particular, we give in Theorem 6 an analogue of Theorem 5 for  $VP^0$  (this is a constant-free version of VP). This result is used in [9], where we show that black-box derandomization of identity testing for sums of products of sparse univariate polynomials with sparse coefficients would imply a lower bound for the permanent.

## 1.1 Main Ideas and Comparison with Previous Work

The main depth reduction result in [1] is as follows.

**Theorem 1** *Let  $P(x_1, \dots, x_m)$  be a polynomial of degree  $d = O(m)$  over a field  $F$ . If there exists an arithmetic circuit of size  $2^{o(d + d \log \frac{m}{d})}$  for  $P$  then there exists a depth 4 arithmetic circuit of size  $2^{o(d + d \log \frac{m}{d})}$ .*

Theorem 2.4 in [1] also provides some bounds on the fan-in of the gates in the resulting depth 4 circuits.

For multilinear polynomials, their result (Corollary 2.5 in [1]) reads as follows:

**Corollary 1** *A multilinear polynomial in  $m$  variables which has an arithmetic circuit of size  $2^{o(m)}$  also has a depth 4 arithmetic circuit of size  $2^{o(m)}$ .*

We give the (simple) proof, which is omitted from [1]. For  $d = m$  the result is clear since the exponent  $d + d \log \frac{m}{d}$  in Theorem 1 is equal to  $m$ . Consider now the case of a polynomial  $P(X_1, \dots, X_m)$  of degree  $d < m$ , having a circuit of size  $2^{o(m)}$ . Let  $Q = P + \prod_{i=1}^m X_i$ . Since the number of variables of  $Q$  is equal to its degree, we are back to the first case:  $Q$  has a depth four circuit of size  $2^{o(m)}$ . We can obtain a circuit of size  $2^{o(m)}$  for  $P$  by subtracting the product  $\prod_{i=1}^m X_i$  (this requires only  $m$  additional arithmetic operations). Note that this corollary (and its proof) hold more generally for any polynomial of degree  $d \leq m$ .

By specializing the multilinear polynomial to the permanent, Agrawal and Vinay then state in Corollary 2.6 that if every depth 4 arithmetic circuit for the permanent requires exponential size, the same is true for arithmetic circuits of unbounded depth. It is not made precise in [1] what “exponential size” exactly means. In this context (arithmetic complexity of the permanent) the most standard interpretation is probably that an exponential size circuit for the  $n \times n$  permanent is of size  $2^{\Omega(n)}$  (note that the number of variables is  $m = n^2$ ). With this interpretation, it is not clear why Corollary 2.6 of [1] would follow from Theorem 1 or Corollary 1.

Since the permanent of a  $n \times n$  matrix has degree  $d = n$  and  $m = n^2$  variables, we can deduce the following from Theorem 1: If there exists an arithmetic circuit of size  $2^{o(n \log n)}$  for the  $n \times n$  permanent then there exists also a depth 4 arithmetic circuit of size  $2^{o(n \log n)}$ . This statement is not very useful since we already know (by Ryser’s formula [15]) that the permanent has depth 3 arithmetic formulas of size  $O(n2^n)$ . Note that applying Corollary 1 directly to the permanent would give an even worse bound (namely, we would obtain depth 4 formulas of size  $2^{o(n^2)}$ ). As explained earlier, we

can show that if the permanent has polynomial size circuits it must also have depth 4 formulas of size  $n^{O(\sqrt{n} \log n)}$ .

Before describing their general depth reduction algorithm, Agrawal and Vinay begin with the special case of matrix powering. For this problem there is a very simple and elegant reduction to depth four. Then they treat the general case with an apparently different approach: their construction builds on the depth reduction algorithm of Allender, Jiao, Mahajan and Vinay [2]. In this paper we show that the matrix powering idea is powerful enough to handle arbitrary polynomial-size arithmetic circuits. Weakly skew circuits are one of the main tools that we use to reduce the evaluation of arbitrary arithmetic circuits to matrix powering. This class of circuits is known to capture the complexity of a number of problems from linear algebra such as e.g. matrix powering, iterated matrix multiplication or computation of the determinant [16, 12].

## 1.2 Organization of the paper

In Section 2 we present the two main computation models that we will use: arithmetic circuits and arithmetic branching programs. We define some of the corresponding complexity classes, and give some basic properties. In Section 3, building on a construction of Malod and Portier [12] we give an efficient simulation of arithmetic circuits by arithmetic branching programs. Compared to [12], we take extra care to construct branching programs of small depth because the square root of the depth appears in the exponent of the size estimate for the final depth 4 circuit. In section 4 we reduce branching programs to depth 4 circuits using the matrix powering idea from [1]. Then we state our main technical result in Theorem 3. We show in particular that an arithmetic circuit of size  $t$  and formal degree  $d$  has a depth 4 circuit of size  $t^{O(\sqrt{d} \log d)}$ . Finally we draw some consequences for depth reduction of VP families in Section 5, and for depth reduction of  $\text{VP}^0$  families in Section 6.

## 2 Arithmetic Circuits and Branching Programs

We recall that an arithmetic circuit contains addition and multiplication gates. In addition to these arithmetic gates there are input gates, labelled by variables or constants from some field  $K$ . An output gate is of fan-out zero. We often assume that there is a single output gate. In this case an arithmetic circuit therefore represents a polynomial with coefficients in  $K$ .

We often assume that the arithmetic gates have arity 2, but in constant-depth circuits we naturally allow addition and multiplication gates of unbounded fan-in (we often also some explicit upper bounds on the fan-in, see for instance Theorem 3). In some of our intermediate constructions (e.g. Proposition 2) we also work with weighted addition gates.

**Definition 1** *A  $n$ -ary weighted addition gate computes a linear combination  $a_1x_1 + \dots + a_nx_n$  of its inputs  $x_1, \dots, x_n$ . Here  $a_i$  is the weight associated to the  $i$ -th input of the gate. The total weight of the gate is  $\sum_{i=1}^n |a_i|$ .*

For instance, a subtraction gate is a binary weighted addition gate with weights  $(1, -1)$ . We sometimes refer to binary unweighted addition gates as “ordinary addition gates”. The size of a circuit is its total number of gates (including input gates).

**Definition 2** *Fix a field  $K$ . A sequence  $(f_n)$  of polynomials with coefficients in  $K$  belongs to  $\mathsf{VP}$  if there exists a polynomial  $p(n)$  and a sequence  $(C_n)$  of arithmetic circuits such that  $\deg(f_n) \leq p(n)$ ,  $C_n$  computes  $f_n$  and is of size at most  $p(n)$ .*

The size constraint implies in particular that  $f_n$  depends on polynomially many variables. The above definition is fairly robust. For instance we obtain the same class with circuits using gates of fan-in 2 or of unbounded fan-in, weighted or unweighted addition gates.

An arithmetic formula is a circuit where all gates are of fan-out one, except of course the output gate. In the constant depth setting, arithmetic formulas and arithmetic circuits are polynomially related ([14], Claim 2.2).

The complexity of several problems from linear algebra such as iterated matrix multiplication or computing the determinant is captured by a restricted class of arithmetic circuits called *weakly skew circuits* [16, 12] (there is also an essentially equivalent notion of *skew circuit* [16, 7, 8]). Let  $C$  be an arithmetic circuit where all multiplication gates are binary. A multiplication gate  $\alpha$  in  $C$  is said to be disjoint if at least one of its two subcircuits is disjoint from the remainder of  $C$ , except of course for the edge from the subcircuit to  $\alpha$  (removing this edge would therefore disconnect  $C$ ). The circuit is weakly skew if its multiplication gates are all disjoint. This definition is usually given only for circuits where all addition gates are binary unweighted, but we will use our slightly more general definition instead (see Propositions 2 and 3).

A circuit where the only constants are from the set  $\{0, -1, 1\}$  is said to be constant-free. A constant-free circuit represents a polynomial in  $\mathbb{Z}[X_1, \dots, X_n]$ , where  $X_1, \dots, X_n$  are the variables labelling the input gates.

The constant-free model was systematically studied by Malod [11]. In particular, he defined a class  $\text{VP}^0$  of polynomial families that are “easy to compute” by constant-free arithmetic circuits. First we need to recall the notion of formal degree:

- (i) The formal degree of an input gate is equal to 1.
- (ii) The formal degree of an addition gate is the maximum of the formal degrees of its incoming gates, and the formal degree of a multiplication gate is the sum of these formal degrees.

Finally, the formal degree of a circuit is equal to the formal degree of its output gate. This is obviously an upper bound on the degree of the polynomial computed by the circuit. Note that this definition can be applied to circuits with weighted addition gates of arbitrary fan-in. For instance, the polynomial  $x - 2y$  can be computed by a circuit containing one ordinary addition gate, one multiplication gate and three inputs labeled by  $x$ ,  $y$  and the constant  $-2$ . This circuit has formal degree two. The same polynomial can be computed by another circuit containing a binary weighted addition gate (of total weight  $1 + |-2| = 3$ ) with inputs  $x$  and  $y$ . The second circuit has formal degree 1.

**Definition 3** *A sequence  $(f_n)$  of polynomials belongs to  $\text{VP}^0$  if there exists a polynomial  $p(n)$  and a sequence  $(C_n)$  of constant-free arithmetic circuits (with unweighted addition gates) such that  $C_n$  computes  $f_n$  and is of size and formal degree at most  $p(n)$ .*

The constraint on the formal degree forbids the computation of polynomials of high degree such as e.g.  $X^{2^n}$ ; it also forbids the computation of large constants such as  $2^{2^n}$ . The class  $\text{VP}^0$  is therefore a strict subset of  $\text{VP}$  (over the field of rational numbers, or more generally any field of characteristic 0). As for  $\text{VP}$  we obtain the same class with gates of fan-in 2 or of unbounded fan-in, but of course we cannot allow addition gates with arbitrary weights. We can however allow subtraction gates:

**Proposition 1** *Let  $C$  be a constant-free circuit of size  $t$  and formal degree  $d$ , where the arithmetic gates are multiplication, unweighted addition or subtraction gates (all of fan-in 2).*

*There is an equivalent constant-free circuit  $C'$  of formal degree  $d+1$  and size at most  $6t+3$ , where the arithmetic gates are binary multiplications or ordinary additions.*

*Proof.* We need to get rid of subtraction gates. A first idea would be to write each subtraction  $x - y$  as  $x + (-1) \times y$ , but the cumulative effect of the multiplications  $(-1) \times y$  could lead to an increase in the formal degree by more than 1. Instead we will represent each gate  $\alpha$  in  $C$  by a pair of gates  $(\alpha_1, \alpha_2)$  in  $C'$ . The output of  $\alpha$  will be equal to the difference of the outputs of  $\alpha_1$  and  $\alpha_2$ . An input  $x$  in  $C$  can be represented by the pair  $(x, 0)$ . To simulate the arithmetic operations in  $C$  we use the following rules:  $(\alpha_1 - \alpha_2) + (\beta_1 - \beta_2) = (\alpha_1 + \beta_1) - (\alpha_2 + \beta_2)$ ;  $(\alpha_1 - \alpha_2) - (\beta_1 - \beta_2) = (\alpha_1 + \beta_2) - (\alpha_2 + \beta_1)$ ;  $(\alpha_1 - \alpha_2) \times (\beta_1 - \beta_2) = (\alpha_1 \times \beta_1 + \alpha_2 \times \beta_2) - (\alpha_2 \times \beta_1 + \alpha_1 \times \beta_2)$ . A straightforward induction shows that the gates in a pair  $(\alpha_1, \alpha_2)$  will have same formal degree as the gate  $\alpha$  that they represent. Finally, to complete the construction of  $C'$  we come back to our first idea: if  $(\alpha_1, \alpha_2)$  is the pair representing the output gate of  $C$ , we write the difference  $\alpha_1 - \alpha_2$  as  $\alpha_1 + (-1) \times \alpha_2$ . This increases the formal degree by 1. Each arithmetic operation in  $C$  is simulated by at most 6 operations in  $C'$ , and we need 3 additional gates to perform the final subtraction.  $\square$

This modest increase in the formal degree cannot be avoided: without subtraction gates there is no better way to compute the polynomial  $f(x) = -x$  than by the formula  $f(x) = -1 \times x$ , which is of formal degree 2.

Finally we define the notion of arithmetic branching program. This is an edge-weighted directed acyclic graph with two distinguished vertices  $s$  and  $t$ . The output of the branching program is by definition equal to the sum of the weights of all paths from  $s$  to  $t$ , where the weight of a path is the product of the weights of its edges. In this paper we assume that the edge weights are constants from some field  $K$  or variables. Like an arithmetic circuit, a branching program therefore represents a polynomial with coefficients in  $K$ . The depth of a branching program is the length (in number of edges) of the longest path from  $s$  to  $t$ . The term *arithmetic* (or *algebraic*) *branching program* goes back at least to [13, 3] but these objects were used implicitly much earlier, for instance in [17].

### 3 From Circuits to Branching Programs

We first recall Lemma 4 from [12].

**Lemma 1** *Let  $C$  be a circuit of size  $t$  and formal degree  $d$ , containing only binary unweighted arithmetic gates. There exists a weakly skew circuit  $C'$  of formal degree  $d$  and size at most  $t^{\log 2d}$  which computes the same polynomial.*



The fact that  $C'$  has same formal degree as  $C$  is not explicitly stated in [12], but it can be checked that their construction does satisfy this additional property. We would like to apply this construction not to  $C$  itself, but to a “normal form” of  $C$  containing weighted addition gates.

We begin with an easy lemma.

**Lemma 2** *Let  $C$  be a circuit made only of input gates and (ordinary) addition or subtraction gates. Each gate of  $C$  is equivalent to a weighted addition gate of total weight at most  $2^s$ , where  $s$  is the number of arithmetic gates in  $C$ .*

*Proof.* By induction on  $s$ . The result is true for  $s \leq 1$  since an input gate can be viewed as a unary weighted addition gate of weight 1, and an ordinary addition or subtraction gate as a binary weighted addition gate of weight 2. For  $s > 1$ , consider an addition or subtraction gate which is an output of  $C$ . By induction hypothesis each of the two inputs of the gate computes a function of the form  $\sum_{i=1}^n a_i x_i$  where  $x_1, \dots, x_n$  are the inputs of  $C$  and  $\sum_i |a_i| \leq 2^{s-1}$ . Therefore the output gate computes a function of the same form with total weight at most  $2^s$ .  $\square$

**Lemma 3** *Let  $C$  be a circuit containing  $s$  (weighted) addition gates and  $m$  multiplication gates. There is an equivalent circuit  $C_+$  such that:*

- (i)  $C_+$  contains at most  $s$  addition gates and  $m$  multiplication gates.
- (ii) Any input to an addition gate is an input of  $C_+$  or a multiplication gate (in other words, the output of an addition gate can be fed only to multiplication gates).
- (iii) If all the addition gates of  $C$  are ordinary additions or subtractions, the total weight of every addition gate of  $C_+$  is at most  $2^s$ .

*Proof.* We will keep the same multiplication gates in  $C_+$  as in  $C$ . Consider a multiplication gate in  $C$  having at least one addition gate as an input. We can view such an addition gate as the output of a maximal subcircuit which does not contain any internal multiplication gate (the inputs to the subcircuit are therefore inputs of  $C$  or multiplication gates). The output of this subcircuit is a linear function of its inputs. We can therefore replace the subcircuit by a single (weighted) addition gate. Moreover, in the case where all the addition gates of  $C$  are ordinary additions or subtractions, this weighted addition gate can be taken of weight at most  $2^s$  by Lemma 2.

We perform this replacement simultaneously for all addition gates of  $C$  feeding into a multiplication gate. If the output of  $C$  is a multiplication gate, we are done. If the output is an addition gate, we likewise replace its maximal subcircuit by a weighted addition gate.  $\square$

The same transformation as in Lemma 1 can be applied to  $C_+$  instead of  $C$ . The resulting weakly skew circuit contains weighted addition gates.

**Proposition 2** *Let  $C$  be a circuit of size  $t$  and formal degree  $d$  where all multiplication gates are binary. There exists a weakly skew circuit  $C'$  of degree  $d$  and size at most  $t^{\log 2^d}$  which computes the same polynomial. In  $C'$ , any input to an addition gate is an input of the circuit or a multiplication gate. Moreover, if all the addition gates of  $C$  are ordinary additions or subtractions, the total weight of every addition gate of  $C'$  is at most  $2^s$ .*

We omit the proof since this is really the same construction as in Lemma 4 of [12]. One can check that this construction indeed preserves properties (ii) and (iii) from Lemma 3.

**Proposition 3** *Let  $C$  be a weakly skew circuit of size  $m$  and formal degree  $d$ , with weights of addition gates coming from some set  $W$ . Assume moreover that any input to an addition gate is an input of  $C$  or a multiplication gate. There exists an equivalent arithmetic branching program  $G$  of size at most  $m + 1$  and depth at most  $3d - 1$ . The edges of  $G$  are labeled by inputs of  $C$  or constants from  $W$ .*

*Proof.* The construction is similar to that of ([12], Lemma 5). The main new point is to check the depth bound. Recall from Section 2 that for every multiplication gate  $\alpha$  in  $C$  we have an *independent subcircuit* which is connected to the remainder of  $C$  only by the arrow from the subcircuit to  $\alpha$ . As in [12] we say that a gate is reusable if it does not belong to any independent subcircuit. Also as in [12], we will prove a version of Proposition 3 for circuits with multiple outputs.

We will show by induction that for any reusable gate  $\alpha$  of  $C$  there is a vertex  $t_\alpha$  in  $G$  such that the weight of  $(s, t_\alpha)$  is the polynomial computed by  $\alpha$ . As to the depth, we will show that if  $\alpha$  is an addition gate computing a polynomial of formal degree  $d_\alpha$ , the depth of  $t_\alpha$  in  $G$  (the length of the longest path from  $s$  to  $t_\alpha$ ) is at most  $3d_\alpha - 1$ ; if  $\alpha$  is a multiplication or input gate, its depth is at most  $3d_\alpha - 2$ .

The beginning of the induction is clear: a weakly skew circuit  $C$  of size  $m = 1$  is reduced to a single gate  $\alpha$  labeled by some input  $x$ . The

corresponding graph  $G$  has two nodes  $s$  and  $t$ , with an edge from  $s$  to  $t$  labeled by  $x$ . We take of course  $t_\alpha = t$ . We have  $d_\alpha = 1$ , and this gate is indeed at depth  $3d_\alpha - 2 = 1$ .

Consider now a weakly skew circuit  $C$  of size  $m \geq 2$ , and let  $\alpha$  be one of its output gates. Removing  $\alpha$  from  $C$ , we obtain a circuit  $C'$  of size  $m - 1$ . By induction hypothesis, there is a corresponding graph  $G'$  of size at most  $m$  with a distinguished vertex  $s$ .

If  $\alpha$  is an input gate labeled by  $x$ , we obtain  $G$  by adding a vertex  $t_\alpha$  to  $G'$ , and an edge from  $s$  to  $t_\alpha$  labeled by  $x$ .

Assume now that  $\alpha$  is a (weighted) addition gate, with  $k$  (distinct) inputs  $\alpha_1, \dots, \alpha_k$ . These  $k$  gates must be reusable, so by induction hypothesis we have vertices  $t_{\alpha_i}$  in  $C'$  so that the weight of  $(s, t_{\alpha_i})$  is equal to the polynomial computed by  $\alpha_i$ . Moreover, since  $\alpha$  is an addition gate the  $\alpha_i$  are multiplication or input gates, and are therefore at depth at most  $3d_{\alpha_i} - 2 \leq 3d_\alpha - 2$ . We obtain  $G$  by adding a new vertex  $t_\alpha$  to  $G'$ , and  $k$  new edges from the  $t_{\alpha_i}$  to  $t_\alpha$  (labeled by the same weights as the incoming edges of the addition gate  $\alpha$ ). The weight of  $(s, t_\alpha)$  in  $G$  is clearly equal to the polynomial computed by  $\alpha$ , and  $t_\alpha$  is at depth at most  $(3d_\alpha - 2) + 1 = 3d_\alpha - 1$ .

Assume finally that  $\alpha$  is a multiplication gate with inputs  $\beta$  and  $\gamma$ . Let  $C_\beta$  and  $C_\gamma$  be the corresponding subcircuits. Since  $C$  is weakly skew, one of the two subcircuits (say,  $C_\gamma$ ) is independent from the rest of  $C$ . Hence  $m = m_\beta + m_\gamma + 1$  where  $m_\beta$  and  $m_\gamma$  are the sizes of  $C_\beta$  and  $C_\gamma$ . We can apply separately the induction hypothesis to  $C_\beta$  and  $C_\gamma$ . This yields two graphs  $G_\beta$  and  $G_\gamma$  of respective sizes at most  $m_\beta + 1$  and  $m_\gamma + 1$ , with sources  $s_\beta$  and  $s_\gamma$ . In these graphs there are vertices  $t_\beta$  and  $t_\gamma$  such that the weight of  $(s_\beta, t_\beta)$  in  $C_\beta$  is equal to the polynomial computed by gate  $\beta$ , and the weight of  $(s_\gamma, t_\gamma)$  in  $C_\gamma$  is equal to the polynomial computed by gate  $\gamma$ . We construct  $G$  from these two graphs by identifying  $t_\beta$  and  $s_\gamma$ . The source of  $G$  is  $s = s_\beta$ . This graph is of size at most  $(m_\beta + 1) + (m_\gamma + 1) - 1 = m \leq m + 1$ . In  $G$ , the vertex associated to gate  $\alpha$  will be  $t_\alpha = t_\gamma$ . The weight of  $(s, t_\gamma)$  in  $G$  is indeed equal to the polynomial computed by gate  $\alpha$ . For vertices  $v$  in  $G_\gamma$  the weight of  $(s, v)$  in  $G$  is *not* equal to the weight of  $(s_\gamma, v)$  in  $G_\gamma$ , but as pointed out in [12] this does not matter since these vertices correspond to non-reusable gates of  $C$ .

Let  $d$ ,  $d_\beta$  and  $d_\gamma$  be the formal degrees of the circuits  $C$ ,  $C_\beta$  and  $C_\gamma$ . By induction hypothesis,  $t_\gamma$  is at depth at most  $3d_\gamma - 1$  in  $G_\gamma$ , and  $t_\beta$  is at depth at most  $3d_\beta - 1$  in  $G_\beta$ . In  $G$ ,  $t_\gamma$  is therefore at depth at most  $(3d_\beta - 1) + (3d_\gamma - 1) = 3d - 2$ .  $\square$

Combining Propositions 2 and 3 yields the following result.

**Theorem 2** *Let  $C$  be a circuit of size  $t$  and formal degree  $d$  where all multiplication gates are binary. There is an equivalent arithmetic branching program  $G$  of size at most  $t^{\log 2d} + 1$  and depth at most  $3d - 1$ . The edges of  $G$  are labeled by inputs of  $C$  or by constants. Moreover, if all the addition gates of  $C$  are ordinary additions or subtractions then these constants are integers of absolute value at most  $2^t$ .*

## 4 From Branching Programs to Depth-4 Circuits

In this section we complete the reduction to circuits of depth 4.

**Lemma 4** *Let  $G$  be an arithmetic branching program of size  $m$  and depth  $\delta$ , with edges labeled by elements from some set  $S$ . There is an  $m \times m$  matrix  $M$  such that the polynomial computed by  $G$  is equal to the entry at row 1 and column  $m$  of the matrix power  $M^p$ , for any  $p \geq \delta$ . Moreover, the entries of  $M$  are in the set  $S \cup \{0, 1\}$ .*

*Proof.* Fix a topological ordering of the nodes of  $G$ , with the source  $s$  labeled 1 and the target  $t$  labeled  $m$ . We define  $M$  as the adjacency matrix of the graph  $G'$  obtained from  $G$  by adding a loop of weight 1 on vertex  $t$ . In other words,  $M_{mm} = 1$  and in all other cases  $M_{ij}$  is the (possibly null) weight from node  $i$  to node  $j$  of  $G$ . Note that  $M$  is upper-diagonal, with all diagonal entries equal to 0 except  $M_{mm}$ . It follows from the classical relation between matrix powering and paths in graphs that  $(M^p)_{1m}$  is equal to the sum of weights of all  $st$ -paths of length exactly  $p$  in  $G'$ . This is also the sum of weights of all  $st$ -paths of length at most  $p$  in  $G$ , and for  $p \geq \delta$  this is the output of the arithmetic branching program.  $\square$

Note that for  $p \geq \delta$  all entries of  $M^p$  except  $(M^p)_{1m}$  are equal to zero.

In the last step in our series of reduction, we explain (following basically the same strategy as in [1]) how to perform the matrix powering operation in the above lemma with depth four formulas, and also depth four circuits.

**Proposition 4** *Let  $G$  be an arithmetic branching program of size  $m$  and depth  $\delta$ . There is an equivalent depth four circuit  $\Gamma$  with  $m^2 + 1$  unweighted addition gates and  $m^{\lceil \sqrt{\delta} \rceil + 1} + m^{\lceil \sqrt{\delta} \rceil - 1}$  multiplication gates. There is also an equivalent depth four formula  $\Gamma_f$  with  $m^{\lceil \sqrt{\delta} \rceil - 1} + 1$  unweighted addition gates and  $m^{\lceil \sqrt{\delta} \rceil - 1} + m^{2\lceil \sqrt{\delta} \rceil - 2}$  multiplication gates.*

*The inputs of  $\Gamma$  and  $\Gamma_f$  are from the set as the edge labels of  $G$ , and their multiplication gates are of fan-in  $\lceil \sqrt{\delta} \rceil$ .*

*Proof.* We need to compute  $M^p$ , where  $p \geq \delta$  and  $M$  is as in Lemma 4. Let  $p$  be the smallest square bigger or equal to  $\delta$ . From  $M$  we will compute  $N = M^{\sqrt{p}}$  by a depth 2 circuit  $\Gamma_2$ , and then from  $N$  we will compute  $M^p = N^{\sqrt{p}}$  using the same circuit. With a depth 2 circuit one cannot play clever tricks: we can only expand a polynomial as a sum of monomials. In this case we express each entry of  $N$  as a sum of  $m^{\sqrt{p}-1}$  products of length  $\sqrt{p}$ , by brute-force expansion of the product  $M^{\sqrt{p}}$ . This yields a circuit  $\Gamma_2$  with  $m^2$  addition gates (one for each entry of  $N$ ) and  $m^{\sqrt{p}+1}$  multiplication gates. We can double those estimates to upper bound the size of  $\Gamma$ . To arrive at the slightly better estimate in the statement of Proposition 4, note that the second copy of  $\Gamma_2$  only needs to compute a single entry of  $N^{\sqrt{p}}$ .

In order to obtain an arithmetic formula, we recompute from scratch each entry of  $N$  whenever it is used by the second copy of  $\Gamma_2$ . The arithmetic formula therefore computes a sum of  $m^{\sqrt{p}-1}$  products, where each product is a sum of  $m^{\sqrt{p}-1}$  products of entries of  $M$ . We therefore have one addition and  $m^{\sqrt{p}-1}$  products gates in the top two levels,  $m^{\sqrt{p}-1}$  addition and  $m^{2(\sqrt{p}-1)}$  multiplication gates in the two bottom levels.  $\square$

Note the significant saving in the number of addition gates if we use depth four circuits instead of depth four formulas. We can now prove our main depth reduction result.

**Theorem 3** *Let  $C$  be an arithmetic circuit of size  $t$  and formal degree  $d$  where all multiplication gates are binary. There is an equivalent depth four circuit  $\Gamma$  with at most  $(t^{\log 2d} + 1)^2 + 1$  unweighted addition gates and at most  $2(t^{\log 2d} + 1)^{\sqrt{3d}+2}$  multiplication gates.*

*There is an equivalent arithmetic formula  $\Gamma_f$  of depth four with at most  $(t^{\log 2d} + 1)^{\sqrt{3d}} + 1$  unweighted addition gates and at most  $2(t^{\log 2d} + 1)^{2\sqrt{3d}}$  multiplication gates. The inputs of  $\Gamma$  and  $\Gamma_f$  are inputs of  $C$  or constants; their multiplication gates are of fan-in at most  $\sqrt{3d} + 1$ .*

*If  $C$  is constant-free, and if all the addition gates of  $C$  are ordinary additions or subtractions, then these constants are integers of absolute value at most  $2^t$ .*

*Proof.* Combine Theorem 2 and Proposition 4.  $\square$

If we start from arithmetic formulas instead of arithmetic circuits, we can obtain depth four formulas and circuits of significantly smaller size. Also the proof is much simpler since we do not need Theorem 2. Instead, we use the classical transformation of arithmetic formulas into arithmetic branching

programs ([17], see also Lemma 2.15 of [4] and the improved bound in [10]).

**Lemma 5** *Let  $F$  be an arithmetic formula which contains  $e$  arithmetic operations. There is an equivalent algebraic branching program  $G$  of size at most  $e + 2 = (t + 3)/2$  and depth at most  $e - 1 = (t + 1)/2$ . Here  $t = 2e + 1$  is the size of  $F$  viewed as an arithmetic circuit. The edges of  $G$  are labeled by the inputs of  $F$ .*

**Theorem 4** *Let  $F$  be an arithmetic formula which contains  $e$  arithmetic operations. There is an equivalent depth four circuit  $\Gamma$  with at most  $(e + 2)^2 + 1$  unweighted addition gates and at most  $2(e + 2)^{\sqrt{e}+2}$  multiplication gates.*

*There is an equivalent arithmetic formula  $\Gamma_f$  of depth four with at most  $(e + 2)^{\sqrt{e}} + 1$  unweighted addition gates and at most  $2(e + 2)^{2\sqrt{e}}$  multiplication gates. The inputs of  $\Gamma$  and  $\Gamma_f$  are inputs of  $C$ , and their multiplication gates are of fan-in at most  $\sqrt{e} + 1$ .*

*Proof.* Combine Lemma 5 and Proposition 4.  $\square$

The savings in the number of addition gates in depth four circuits compared to depth four formulas are especially significant in the above theorem: our circuits contain only quadratically many addition gates. This is a relevant parameter since the number of addition gates (minus 1) is equal to the number of *distinct* sparse polynomials in a sum of products of sparse polynomials [9].

## 5 Depth Reduction for VP

In accordance with Definition 1, a unary weighted addition gate outputs  $\alpha \cdot x$ , where  $\alpha$  is the weight of the gate and  $x$  its input. Recall also from the definition of formal degree in Section 2 that the formal degree of such a gate is equal to that of its input.

The following result is essentially Lemma 2 from [11], written in a different language. We give the proof because we will build on it in the next section.

**Proposition 5** *Any VP family  $(f_n)$  can be computed by a polynomial-size family  $(C_n)$  of circuits of formal degree  $\deg(f_n)$ . The addition gates of  $C_n$  are unary weighted or binary unweighted (i.e., “ordinary”).*

*Proof.* Since  $(f_n)$  is in VP, this family can be computed by a family  $(C'_n)$  of arithmetic circuits of polynomial size where all the arithmetic gates are binary unweighted. To construct  $C_n$  from  $C'_n$  we use a small variation on the standard homogenization trick. In order to homogenize  $C'_n$  one would normally represent each gate  $\gamma$  computing a polynomial  $f_\gamma$  by a sequence  $\gamma_i$  of  $d_n + 1$  gates, where  $i$  ranges from 0 to  $d_n$  and  $\gamma_i$  computes the homogenous component of  $f_\gamma$  of degree  $i$ . The homogenous components of degree higher than  $d_n$  can be discarded since they cannot contribute to the final output. This construction preserves polynomial circuit size, and each gate now computes a polynomial of degree at most  $d_n$ . But formal degree can be higher due to multiplication by constants (i.e., homogenous components of degree 0).

To circumvent this difficulty, we get rid of the gates  $\gamma_0$  representing homogenous components of degree 0. We will therefore construct a circuit  $C''_n$  which computes the sum of all homogenous components of  $f_n$  of degree at least 1. Our final circuit  $C_n$  will then add the output of  $C''_n$  to the constant term of  $f_n$ , at the cost of one additional arithmetic operation.

We will use unweighted addition gates inside  $C''_n$ . Indeed, let  $\gamma$  be a multiplication gate of  $C_n$  with inputs  $\alpha$  and  $\beta$ . To obtain  $f_{\gamma,i}$ , the homogenous component of degree  $i$ , one normally writes  $f_{\gamma,i} = \sum_{j=0}^i f_{\alpha,j} f_{\beta,i-j}$ . This expression involves  $f_{\alpha,0}$  and  $f_{\beta,0}$ , which as we have said are not represented by any gate of  $C''_n$ . Therefore, to compute e.g.  $f_{\alpha,0} f_{\beta,i}$ , instead of a multiplication gate we use a unary addition gate with input  $f_{\beta,i}$  and weight  $f_{\alpha,0}$ . A straightforward induction shows that a gate  $\gamma_i$  in  $C''_n$  will have formal degree  $i$ . As a result,  $C''_n$  and  $C_n$  will be of formal degree  $d_n$ .  $\square$

**Theorem 5** *Let  $(f_n)$  be a VP family of polynomials of degree  $d_n = \deg(f_n)$ . This family can be computed by a family  $(\Gamma_n)$  of depth four circuits with  $n^{O(\log d_n)}$  addition gates and  $n^{O(\sqrt{d_n} \log d_n)}$  multiplication gates. The family  $(f_n)$  can also be computed by a family  $(F_n)$  of depth four arithmetic formulas of size  $n^{O(\sqrt{d_n} \log d_n)}$ . The inputs to  $\Gamma_n$  and  $F_n$  are variables of  $f_n$  or constants; their multiplication gates are of fan-in at most  $\sqrt{3d_n} + 1$ .*

*Proof.* This is an application of Theorem 3:  $t$  is polynomial in  $n$ , and by Proposition 5 we can take  $d = d_n$ .  $\square$

## 6 Depth Reduction for $VP^0$

We first show that a circuit of small size and degree where all inputs are in  $\{-1, 0, 1\}$  cannot compute a large integer.

**Lemma 6** *Let  $C$  be a constant-free and variable-free circuit of size  $t$  and formal degree  $d$  where all arithmetic gates are binary unweighted. The output of  $C$  is an integer of absolute value at most  $2^{td}$ .*

*Proof.* By induction on  $t$ . For  $t = 1$  the circuit contains a single input gate, which must carry an integer in  $\{-1, 0, 1\}$ . The result is therefore true for  $t = 1$ . Consider now a circuit  $C$  of size  $t \geq 2$ , and let  $d_1$  and  $d_2$  be the formal degrees of the two inputs to the output gate. By induction hypothesis these two gates carry integers of absolute value at most  $2^{(t-1)d_1}$  and  $2^{(t-1)d_2}$ . If the output gate is an addition we have  $d_1, d_2 \leq d$  and  $C$  therefore computes an integer of absolute value at most  $2^{(t-1)d} + 2^{(t-1)d} \leq 2^{td}$ . If the output gate is a multiplication, we have  $d = d_1 + d_2$  and  $C$  computes an integer of absolute value at most  $2^{(t-1)d_1} \times 2^{(t-1)d_2} \leq 2^{td}$ .  $\square$

**Proposition 6** *Any  $\text{VP}^0$  family  $(f_n)$  can be computed by a family  $(C_n)$  of constant-free circuits of polynomial size and formal degree  $\deg(f_n)$ . The arithmetic gates of  $C_n$  are binary multiplication, ordinary addition or subtraction gates.*

*Proof.* Since  $(f_n)$  is in  $\text{VP}^0$ , this family can be computed by a family  $(C'_n)$  of constant-free circuits of polynomial size and polynomial formal degree. All the arithmetic gates of  $C'_n$  can be assumed to be binary unweighted. To construct  $C_n$  from  $C'_n$  we proceed along the same lines as in Proposition 5. In particular, we will again construct a circuit  $C''_n$  which computes the sum of all homogenous components of  $f_n$  of degree at least 1. Our final circuit  $C_n$  then adds the output of  $C''_n$  to the constant term of  $f_n$  (call it  $c_n$ ). By Lemma 6,  $c_n$  has polynomial bit size (it is equal to the output of  $C'_n$  when all variables are set to 0). We can therefore compute  $|c_n|$  from scratch using a sequence of multiplications by 2 and additions of bits. We use an addition to perform a multiplication by 2, so this construction does not require any multiplication gate. Finally, depending on the sign of  $c_n$  we add or subtract  $|c_n|$  to the output of  $C''_n$ . The resulting circuit  $C_n$  will have same formal degree as  $C''_n$ .

We also need to use a similar trick inside  $C''_n$ . Indeed, let  $\gamma$  be a multiplication gate of  $C_n$  with inputs  $\alpha$  and  $\beta$ . To obtain  $f_{\gamma,i}$ , the homogenous component of degree  $i$ , one normally writes  $f_{\gamma,i} = \sum_{j=0}^i f_{\alpha,j} f_{\beta,i-j}$ . This expression involves  $f_{\alpha,0}$  and  $f_{\beta,0}$ , which as explained in the proof of Proposition 5 are not represented by any gate of  $C''_n$ . Therefore, to compute e.g.  $f_{\alpha,0} f_{\beta,i}$  we start from  $f_{\beta,i}$  and compute the product using a sequence of multiplications by 2 and additions of  $f_{\beta,i}$ . As explained above, thanks to



Lemma 6 this can be done with a polynomial number of addition gates, at most one subtraction and no multiplication gate. A straightforward induction shows that a gate  $\gamma_i$  in  $C_n''$  will have formal degree  $i$ . As a result,  $C_n''$  and  $C_n$  will be of formal degree  $d_n$ .  $\square$

By Proposition 1, one can get rid of the subtraction gates in Proposition 6 at the cost of a linear increase in circuit size and an increase in the formal degree by just 1 (using Lemma 3 from [11] instead of Proposition 1 would give a worse degree bound).

**Theorem 6** *Let  $(f_n)$  be a  $\text{VP}^0$  family of polynomials of degree  $d_n = \deg(f_n)$ . This family can be computed by a family  $(\Gamma_n)$  of depth four circuits with  $n^{O(\log d_n)}$  addition gates and  $n^{O(\sqrt{d_n} \log d_n)}$  multiplication gates. The family  $(f_n)$  can also be computed by a family  $(F_n)$  of depth four arithmetic formulas of size  $n^{O(\sqrt{d_n} \log d_n)}$ . The inputs to  $\Gamma_n$  and  $F_n$  are variables of  $f_n$  or relative integers of polynomial bit size; their multiplication gates are of fan-in at most  $\sqrt{3d_n} + 1$ .*

*Proof.* This is an application of Theorem 3:  $t$  is polynomial in  $n$ , and by Proposition 6 we can take  $d = d_n$ .  $\square$

## Acknowledgments

I thank Bruno Grenet and Natacha Portier for useful discussions on this work.

## References

- [1] M. Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *Proc. 49th IEEE Symposium on Foundations of Computer Science*, pages 67–75, 2008.
- [2] E. Allender, J. Jiao, M. Mahajan, and V. Vinay. Non-commutative arithmetic circuits: depth reduction and size lower bounds. *Theoretical Computer Science*, 209:47–86, 1998.
- [3] A. Beimel and A. Gál. On arithmetic branching programs. *Journal of Computer and System Sciences*, 59(2):195–220, 1999.

- [4] P. Bürgisser. *Completeness and Reduction in Algebraic Complexity Theory*. Number 7 in Algorithms and Computation in Mathematics. Springer, 2000.
- [5] D. Grigoriev and M. Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Proc. 30th ACM Symposium on Theory of Computing*, pages 577–582, 1998.
- [6] D. Grigoriev and A. Razborov. Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 6(10):465–487, 2000.
- [7] M. Jansen. Lower bounds for syntactically multilinear algebraic branching programs. In *Proc. MFCS 2008*, volume 5162 of *Lecture Notes in Computer Science*, pages 407–418. Springer-Verlag, 2008.
- [8] E. Kaltofen and P. Koiran. Expressing a fraction of two determinants as a determinant. In *Proc. 2008 International Symposium on Symbolic and Algebraic Computation (ISSAC)*, pages 141–146. ACM Press, 2008.
- [9] P. Koiran. Shallow circuits with high-powered inputs. <http://arxiv.org/abs/1004.4960>, 2010.
- [10] H. Liu and K.W. Regan. Improved construction for universality of determinant and permanent. *Information Processing Letters*, 100(6):233–237, 2006.
- [11] G. Malod. *Polynômes et coefficients*. PhD thesis, Université Claude Bernard - Lyon 1, 2003.
- [12] G. Malod and N. Portier. Characterizing Valiant’s complexity classes. *Journal of Complexity*, 24:16–38, 2008. Conference version in MFCS 2006.
- [13] N. Nisan. Lower bounds for non-commutative computation. In *Proc. 23rd ACM Symposium on Theory of Computing*, pages 410–418, 1991.
- [14] R. Raz and A. Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. *Computational Complexity*, 18(2):171–207, 2009.
- [15] H. J. Ryser. *Combinatorial Mathematics*, volume 201 of *Carus mathematical monographs*. Mathematical Association of America, 1963.

- [16] S. Toda. Classes of arithmetic circuits capturing the complexity of computing the determinant. *IEICE T. Inf. Syst.*, 75(1):116–124, 1992.
- [17] L. G. Valiant. Completeness classes in algebra. In *Proc. 11th ACM Symposium on Theory of Computing*, pages 249–261, 1979.