ON ACC

RICHARD BEIGEL AND JUN TARUI

Abstract. We show that every language L in the class ACC can be recognized by depth-two deterministic circuits with a symmetric-function gate at the root and $2^{\log^{O(1)} n}$ AND gates of fan-in $\log^{O(1)} n$ at the leaves, or equivalently, there exist polynomials $p_n(x_1, \ldots, x_n)$ over $\mathbf{Z}$ of degree $\log^{O(1)} n$ and with coefficients of magnitude $2^{\log^{O(1)} n}$ and functions $h_n : \mathbf{Z} \to \{0, 1\}$ such that for each n and each $x \in \{0, 1\}^n$, $\chi_L(x) = h_n(p_n(x_1, \ldots, x_n))$. This improves an earlier result of Yao (1985). We also analyze and improve modulus-amplifying polynomials constructed by Toda (1991) and Yao (1985).

1. Introduction and Overview

1.1. The ACC Problem. Strong lower bounds have been established for the size of constant-depth circuits that compute explicit Boolean functions in the case that the allowable gates are NOT, OR, AND, and MOD_q, where q is a fixed prime power. (For Boolean variables $y_1, \ldots, y_\nu$, $\mathrm{MOD}_m(y_1, \ldots, y_\nu) = 1$ if $\sum y_i \equiv 0 \pmod{m}$, and 0 otherwise.) A series of work by Furst *et al.* (1984), Ajtai (1983), Yao (1985), and Håstad (1986) has established a near-optimal exponential lower bound for the size of constant-depth circuits with NOT, OR, and AND gates that compute PARITY. Razborov (1987) and Smolensky (1987) have shown that to compute the MOD_q function, constant-depth circuits with NOT, OR, AND, and $\mathrm{MOD}_{q'}$ gates require exponential size if q and q' are powers of distinct primes. For more information about these results and this line of research including history, motivations, and applications, see Sipser (1992) and Boppana & Sipser (1990).

It remains an open problem, however, to show a limitation of constant-depth circuits with MOD_m gates, where m is a fixed composite: The class ACC

— defined by Barrington (1989) and considered further by Barrington (1989), Barrington & Thérien (1988), McKenzie & Thérien (1989), and Yao (1985) — consists of languages recognized by a family of constant-depth polynomial-size circuits with NOT and unbounded fan-in OR, AND, and MOD_m gates, where m is fixed for the family. It is an open problem to show some explicit language (e.g., a language in NP) is not in ACC.

1.2. Boolean Functions and Polynomials. Considering representability (in several senses) of a Boolean function by a polynomial has provided many insights in the theory of shallow circuits. Work along this line can give (1) a positive result by showing that (a) predicates in a certain class can be represented in a certain sense by polynomials of some restricted type (e.g., polynomials of degree polylogarithmic in the number of variables) and that (b) such polynomials can be simulated by certain circuits. It can also give (2) a negative result that some function f is outside a class $\mathcal{C}$ by showing, in addition to (a), that (c) function f cannot be represented by polynomials used for class $\mathcal{C}$ in (a).

The work of Razborov (1987) and Smolensky (1987) mentioned above is a beautiful example of type (2). Celebrated work of Toda (1991) is of type (1): Although Toda's work is in the context of PH, the polynomial hierarchy, (for a definition, see, e.g., Johnson (1990)), it can be translated into the context of shallow circuits because of the well-known connection established by Furst *et al.* (1984) between PH and AC^0 (the class of languages recognized by constant-depth polynomial-size circuits with NOT, OR, and AND gates; more accurately, corresponding to PH is the class qAC^0 obtained by taking the size bound to be quasipolynomial, i.e., $2^{\log^{O(1)} n}$). Indeed, corresponding results in terms of shallow circuits and their improvements have been shown in a series of subsequent work by Allender (1989), Allender & Hertrampf (1994), Beigel *et al.* (1991), Kannan *et al.* (1993), Tarui (1993), and Toda & Ogiwara (1992). Many other results obtained by considering polynomial representations are explained by Beigel (1993).

1.3. Results. Yao (1985) obtained the first nontrivial upper bound on the computing power of ACC circuits. In this paper we simplify Yao's proof and improve his result (thus both contributions are of type (1) above).

For a polynomial $p(x_1, \ldots, x_n)$ over $\mathbf{Z}$, the ring of integers, define the *norm* of p to be the sum of the absolute values of the coefficients of p. (Beigel & Tarui (1991) and respectively Yao (1985) use the word "size" to denote what we call norm and the logarithm of what we call norm.) Define SYM^+ to be the class of languages L for which there exist a family $\{r_n(x_1, \ldots, x_n)\}$ of degree-$\log^{O(1)} n$

norm-$2^{\log^{O(1)} n}$ polynomials over $\mathbf{Z}$ and a family $\{h_n\}$ of functions from $\mathbf{Z}$ to $\{0,1\}$ such that for each n and each $x \in \{0,1\}^n$, $\chi_L(x) = h_n(r_n(x_1, \ldots, x_n))$, where χ_L denotes the characteristic function of L. (Beigel & Tarui (1991) called the class SYM$^+$ by the name SYMMC. Here we have adapted the notation proposed by Beigel *et al.* (1991) and Barrington (1992).) By standard techniques of Beigel *et al.* (1994) it is immediate that a language L is in SYM$^+$ if and only if L can be recognized by depth-two size-$2^{\log^{O(1)} n}$ circuits with a symmetric-function gate at the root (*top*) and AND gates of fan-in $\log^{O(1)} n$ at the leaves (*bottom*). (A symmetric-function gate computes some symmetric Boolean function, i.e., a Boolean function that only depends on the number of inputs that are 1. A NOT gate can appear in a circuit as a negated input literal $\overline{x_i}$.)

Yao (1985) showed that ACC is contained in a probabilistic version of SYM$^+$: If L is in ACC, there exist finite sets S_n of degree-$\log^{O(1)} n$ norm-$2^{\log^{O(1)} n}$ polynomials in n variables, "simple" probability distributions ρ_n on S_n, and functions $h_n : \mathbf{Z} \to \{0,1\}$ such that for each n and each $x \in \{0,1\}^n$, when $r \in S_n$ is randomly chosen according to ρ_n, $\chi_L(x) = h_n(r_n(x_1, \ldots, x_n))$ with high probability.

In this paper we show that ACC is in fact contained in SYM$^+$:

THEOREM 1.1.

$$\mathrm{ACC} \subseteq \mathrm{SYM}^+.$$

We can think of Yao's result and our improvement as exhibiting somewhat surprising representational power of low-degree polynomials or as raising the new problem of showing that some explicit language is outside SYM$^+$ or some subclass of SYM$^+$ by analyzing (maybe a restricted class of) low-degree polynomials algebraically or combinatorially (i.e., showing (c) above for SYM$^+$). In both senses it seems more useful to think of SYM$^+$ in terms of low-degree polynomials as opposed to depth-two circuits (hence our definition of SYM$^+$ above).

Actually we obtain a "uniform" version of Theorem 1 by using the Valiant–Vazirani method due to Toda (1991) for probabilistic simulations of OR and AND. If we do not care about uniformity, we can instead use a simple nonconstructive argument together with the Razborov–Smolensky method, and obtain polynomials of lower degree in the end. We include a full explanation of how to do this, and the paper is totally self-contained as to proving Theorem 1 as stated (without uniformity). (As for the proof of the uniform version, we refer the reader to the literature for a discussion of the Valiant–Vazirani method.)

We also show the following extension of Theorem 1.1, in which we allow an output gate to be any symmetric-function gate, not just a MOD_m gate.

PROPOSITION 1.2. *Let L be a language recognized by a family of constant-depth size-$2^{\log^{O(1)} n}$ circuits having a symmetric-function gate at the root and* NOT, OR, AND *and* MOD_m *gates elsewhere, where m is fixed for the family. Then L is in* SYM^+.

We prove Theorem 1.1 by showing how one can convert an ACC circuit to an equivalent *modular polynomial circuit* in which each gate evaluates a low-degree polynomial modulo some prime p, and showing how one can collapse such a modular polynomial circuit using *modulus-amplifying* polynomials. In this way, we can present all arguments explicitly in terms of polynomials, which is the way that they are best seen.

1.4. Modulus-Amplifying Polynomials. Say that an integer polynomial $P(x)$ in one variable is *k-modulus-amplifying* if for all integers N and all integers $m \geq 2$,

$$N \equiv 0 \pmod{m} \implies P(N) \equiv 0 \pmod{m^k};$$
$$N \equiv 1 \pmod{m} \implies P(N) \equiv 1 \pmod{m^k}.$$

Toda (1991) was the first to discover a construction and an application of low-degree modulus-amplifying polynomials. Toda constructed a k-modulus-amplifying polynomial of degree $\Theta(k^2)$ and used it in proving that $\mathrm{BP} \cdot \oplus\mathrm{P} \subseteq \mathrm{P}^{\#\mathrm{P}[1]}$. (The polynomials actually constructed by Toda have modulus-amplifying property of a slightly different kind as will be explained in Section 2.3.) Yao (1985) discovered a new application of modulus-amplifying polynomials and obtained the result mentioned above; also he noted that a k-modulus-amplifying polynomial of degree $\Theta(k^{\log_2 3})$ can be obtained by a slightly different construction. Both Toda and Yao used a recursive construction. We put these polynomials that seem somewhat magical in better perspective and obtain a k-modulus-amplifying polynomial of degree $2k-1$, which is optimal. Modulus-amplifying polynomials of lower degree yield polynomials of lower degree in the proof of Theorem 1.1, but are not essential for the proof.

2. Proof of Theorem 1.1

As usual, we assume without loss of generality that NOT gates in a circuit only appear as negated input literals $\overline{x_i}$'s. All polynomials in the paper are over $\mathbf{Z}$. We let $\mathbf{Z}[x_1, \ldots, x_l]$ denote the ring of polynomials over $\mathbf{Z}$ in variables $x_1, \ldots, x_l$.

Throughout the paper we will be interested in producing low-degree small-norm polynomials. It turns out that for polynomials that we deal with, the norm is always at most exponential in the degree, and that checking this is usually easy. Thus the reader may mostly pay attention to the degree.

2.1. Representing $\mathbf{MOD}_{p^e}$, OR, and AND Modulo p. In this subsection we show how the MOD_{p^e} function, where p is prime, and the OR/AND functions can be represented, modulo p, by low-degree small-norm polynomials over $\mathbf{Z}$.

2.1.1 $\mathbf{MOD}_{p^e}$

We include our own proof of the following lemma, which seems to be folklore (the earliest use we can find is due by Chandra *et al.* (1984)).

LEMMA 2.1. *Let p be a prime and let $e \geq 1$. Then, there is a polynomial $r(x_1, \ldots, x_n)$ of degree $p^e - 1$ and norm $n^{O(p^e)}$ such that for each $x \in \{0,1\}^n$,*

$$\mathrm{MOD}_{p^e}(x_1, \ldots, x_n) = r(x_1, \ldots, x_n) \bmod p.$$

To prove the lemma, we use the following fact. A proof of this fact using Kummer's Theorem was given by Beigel & Gill (1992); a proof using Lucas's theorem was given by Beigel & Tarui (1991). Here we give a simple, direct proof.

FACT 2.2. *For a prime p, a positive integer e, and an integer x,*

$$x \equiv 0 \pmod{p^e} \iff \forall\, i \in \{0, \ldots, e-1\} \quad \binom{x}{p^i} \equiv 0 \pmod{p}.$$

PROOF. Write

$$\binom{x}{p^i} = \frac{x(x-1)\cdots(x-p^i+1)}{p^i(p^i-1)\cdots 1}.$$

The factors in both numerator and denominator take the values $0, 1, \ldots, p^i - 1$ modulo p^i. Thus $\binom{x}{p^i} \equiv 0 \pmod{p}$ if and only if the unique factor in the numerator that is a multiple of p^i is in fact a multiple of p^{i+1}. From this the conclusion follows by a simple induction on i. $\square$

Proof of Lemma 2.1. By Fermat's little theorem, for integers $y_1, \dots, y_k$,

$$\prod_{i=1}^{k}(1 - y_i^{p-1}) \equiv \begin{cases} 1 \pmod p & \text{if } \forall\, i \in \{1, \dots, k\} \;\; y_i \equiv 0 \pmod p; \\ 0 \pmod p & \text{otherwise.} \end{cases}$$

From this and Fact 2.2, it is easy to see that the following polynomial satisfies the conclusion.

$$\begin{aligned} r(x_1, \dots, x_n) &= \prod_{i=0}^{e-1}\left(1 - \binom{\sum_{j=1}^{n} x_j}{p^i}^{p-1}\right) \\ &= \prod_{i=0}^{e-1}\left(1 - \left(\sum_{S \subseteq \{1,\dots,n\}, |S| = p^i} \; \prod_{j \in S} x_j\right)^{p-1}\right). \;\square \end{aligned}$$

2.1.2 OR and AND

A *probabilistic polynomial* $p(x_1, \dots, x_n)$ is a random variable that is uniformly distributed over some finite multiset $\Omega = \{p_1, \dots, p_s\}$, where $p_i \in \mathbf{Z}[x_1, \dots, x_n]$. The *degree* and the *norm* of a probabilistic polynomial p are, respectively, the maximum degree and the maximum norm of p_i $(1 \le i \le s)$.

The Valiant–Vazirani method due to Toda (1991), as reinterpreted and extended by Allender (1989, Allender & Hertrampf (1994), Beigel *et al.* (1991), Kannan *et al.* (1993), and Tarui (1993), and Toda & Ogiwara (1992), yields the following. (For a proof of the particular version stated below, see Tarui (1993) or Beigel *et al.* (1991); the condition (2) below is a technical one that makes our subsequent proofs simpler and can easily be satisfied by raising a polynomial to the $(p-1)$-th power.)

Lemma 2.3. *Let p be a prime and let $\varepsilon > 0$. Then there is a probabilistic polynomial $r(x_1, \dots, x_n)$ that has degree $d = O(\log(1/\varepsilon)\log n)$, norm $n^{O(d)}$, and "easily" constructible sample space of size $2^{O(\log(1/\varepsilon)\log^2 n)}$, and satisfies the following: (1) For each $x \in \{0,1\}^n$, $r(x) \bmod p = \mathrm{OR}(x)$ with probability at least $1 - \varepsilon$. (2) For each $x \in \{0,1\}^n$, $r(x) \bmod p \in \{0,1\}$ with probability 1. A similar probabilistic polynomial for* AND *also exists.*

Remark 2.4. *To obtain "uniform" versions of our results we need "easy" constructibility of a sample space. If we do not care about uniformity, we can alternatively proceed as follows. Let*

$$\Gamma = \{(a_1x_1 + \cdots + a_nx_n)^{p-1} : (a_1, \dots, a_n) \in \{0, \dots, p-1\}^n\}.$$

If we take $\ell = O(\log(1/\varepsilon))$ *large enough and let*

$$\Lambda = \{1 - \prod_{i=1}^{\ell}(1 - q_i) : (q_1, \ldots, q_\ell) \in \Gamma^\ell\},$$

then for each $x \in \{0,1\}^n$*, a randomly chosen* $r \in \Lambda$ *satisfies* $(r(x) \bmod p) = \mathrm{OR}(x)$ *with probability at least* $1 - \frac{1}{2}\varepsilon$ *as was noted by Razborov (1987) and Smolensky (1987).*

By a simple probabilistic argument involving the Chernoff bound, the existence of a small subset $\Omega \subseteq \Lambda$ *that computes OR with probability at least* $1 - \varepsilon$ *can be shown. Fix* $x \in \{0,1\}^n$*. For large enough* $N = O((1/\varepsilon) \cdot n)$*, if we sample a polynomial from* Λ *independently* N *times and obtain* $r_1, \ldots, r_N$*, then the probability that* $r_i(x) \neq \mathrm{OR}(x)$ *for more than* εN r_i*'s is less than* 2^{-n} *by the Chernoff bound on tail of Bernoulli trials. Thus, there exists a multiset* $\Omega \subseteq \Lambda$ *of size* N *such that for* every $x \in \{0,1\}^n$, $(r(x) \bmod p) = \mathrm{OR}(x)$ *except for at most* ε *fraction of* r*'s in* Ω*.*

The multiset Ω *considered as a probabilistic polynomial satisfies the conditions of Lemma 2.3 and has lower degree* $d = O(p \log(1/\varepsilon))$*, norm* $n^{O(d)}$*, and smaller sample space of size* $O((1/\varepsilon) \cdot n)$*. Lower degree and smaller sample space yield polynomials of lower degree in our subsequent proofs. Actually, to achieve low degree, using a Chernoff-bound argument above at the later stage is more effective as will be explained in Remark 2.6.*

2.2. Modular Polynomial Circuits. In what follows we denote a gate in a circuit by a lower-case letter, e.g., g_i, and the Boolean function that a gate computes by the corresponding upper-case letter, e.g., G_i. For gates $g_1, \ldots, g_l$, we also consider a polynomial in variables $g_1, \ldots, g_l$, thus letting $g_1, \ldots, g_l$ denote both gates and formal variables in a polynomial. We assume, without loss of generality for all our purposes, that between any pair of gates in a circuit, there is at most one edge. For a gate g in a circuit, $\mathrm{input}(g)$ denotes the set of gates g_i such that there is an edge from g_i to g.

A *modular polynomial circuit* C for n Boolean variables is similar to a standard circuit except that each gate is labeled by a polynomial instead of AND, OR, etc. Each *non*output gate g with $\mathrm{input}(g) = \{g_1, \ldots, g_l\}$ is associated with some polynomial $r \in \mathbf{Z}[g_1, \ldots, g_l]$ and a positive integer m called its *modulus*. We require that each such pair of polynomial r and modulus m has the property that for each $(g_1, \ldots, g_l) \in \{0,1\}^l$, $r(g_1, \ldots, g_l) \bmod m \in \{0,1\}$. Each such gate is interpreted to compute the Boolean function

$$G(x_1, \ldots, x_n) = r(G_1(x_1, \ldots, x_n), \ldots, G_l(x_1, \ldots, x_n)) \bmod m.$$

An *output* gate g with input(g)= $\{g_1,\ldots,g_l\}$ is associated with some polynomial $r \in \mathbf{Z}[g_1,\ldots,g_l]$ *and* some function $h : \mathbf{Z} \to \{0,1\}$ (no modulus is associated with an output gate), and is interpreted to compute the Boolean function $h(r(G_1(x),\ldots,G_l(x)))$.

A modular polynomial circuit C is *stratified* if each wire in C is between gates of depth d and $d+1$ for some d and all the gates at depth i are associated with a common single modulus m_i. For a modular polynomial circuit: the *size* and the *depth* are, as in a standard circuit, the number of vertices and the depth of its underlying graph respectively; the *degree* and the *norm* are the maximum degree and norm, respectively, over polynomials associated with its gates; its *modulus size* is the maximum of moduli associated with its gates.

LEMMA 2.5. *For any depth-c size-s $\{\mathrm{AND}, \mathrm{OR}, \mathrm{MOD}_m\}$ circuit for n variables, there exists an equivalent stratified modular polynomial circuit of depth $c' = O(c)$, size $s' = 2^{O(\log^3 s)}$, degree $d = O(m\log^2 s)$, norm $t = s^{O(d)}$, and modulus size $\le m$. In particular, for $s = 2^{\log^{O(1)} n}$ and $m = \log^{O(1)} n$, we have $s' = 2^{\log^{O(1)} n}$, $d = \log^{O(1)} n$, $t = 2^{\log^{O(1)} n}$.*

PROOF. Let C be a circuit as above and let $m = p_1^{e_1}\cdots p_\mu^{e_\mu}$ be the prime factorization of m. A MOD_m gate is equivalent to the AND of a $\mathrm{MOD}_{p_1^{e_1}}$ gate, …, a $\mathrm{MOD}_{p_\mu^{e_\mu}}$ gate. Thus, by adding some "dummy" gates if necessary, and increasing size and depth by only a constant factor, we can convert C into an equivalent stratified ACC circuit C' in which all the MOD gates at the same depth i share a single modulus $p_i^{e_i}$.

Fix depth i and let $p = p_i$ and $e = e_i$. A gate at depth i is either an AND, an OR, or a MOD_{p^e} gate. If there is no MOD gate at depth i, take $p = 2$. For each gate g with input(g) = $\{g_1,\ldots,g_l\}$, associate a modulus p, and associate a polynomial $r \in \mathbf{Z}[g_1,\ldots,g_l]$ as follows.

- Case 1: g is a MOD_{p^e} gate. Associate the polynomial given in the proof of Lemma 2.1.

- Case 2: g is an OR gate or an AND gate. Associate a probabilistic polynomial given in Lemma 2.3 (take $\varepsilon = 1/(3s)$) that has degree $d = O(p\log^2 s)$, norm $s^{O(d)}$, and sample space of size $s' = 2^{O(\log^3 s)}$, and computes OR/AND on $\{0,1\}^\ell$ with probability at least $1-1/(3s)$. (Note that $\ell \le \mathrm{size}(C') = O(s)$.)

(At the bottom level proceed similarly as above using $(1-x_i)$ for each negative literal $\overline{x_i}$.)

Now C' has been transformed to a "modular *probabilistic* polynomial circuit" that, for each x, computes C'_n with probability at least $2/3$. We can assume that all probabilistic polynomials used have underlying sample space of the same size $S = 2^{O(\log^3 n)}$ and that each sample space Ω is indexed by the set $\{1, \ldots, S\}$ (i.e., $\Omega = \{r_i\}_{i=1}^{S}$). By fixing $i = 1, \ldots, S$, thus "fixing" every probabilistic polynomial to each of the S ordinary polynomials in its sample space, obtain S modular (ordinary) polynomial circuits, and connect their S output gates $g_1, \ldots, g_S$ to a new output gate g. Associate with g the linear polynomial $g_1 + \cdots + g_S$ and the function $h : \mathbf{Z} \to \{0, 1\}$ that computes the majority among g_i's: $h(y) = 1$ if $y \geq \lceil S/2 \rceil$, and 0 otherwise. We have obtained a desired modular polynomial circuit. $\square$

REMARK 2.6. *If we use appropriately amplified Razborov–Smolensky polynomials (use Λ in Remark 2.4) in Case 2 and apply a Chernoff-bound argument as in Remark 2.4 to independent copies of C' thus obtained, we can produce an equivalent modular polynomial circuit of size $O(sn)$ and degree $O(m \log s)$.*

2.3. Modulus-Amplifying Polynomials. Recall that a k-modulus-amplifying polynomial P_k satisfies the following conditions for all integers N:

$$\begin{aligned} N \equiv 0 \pmod{m} &\implies P_k(N) \equiv 0 \pmod{m^k}; \\ N \equiv 1 \pmod{m} &\implies P_k(N) \equiv 1 \pmod{m^k}. \end{aligned}$$

Toda (1991) constructed polynomials $\hat{P}_k$ that satisfy the following slightly different conditions and can be used for his applications (and for proving our theorem also) just as polynomials P_k:

$$\begin{aligned} N \equiv 0 \pmod{m} &\implies \hat{P}_k(N) \equiv 0 \pmod{m^k}; \\ N \equiv -1 \pmod{m} &\implies \hat{P}_k(N) \equiv -1 \pmod{m^k}. \end{aligned}$$

Toda obtained his polynomials by letting $\hat{P}_2(x) = 3x^4 + 4x^3$ and $\hat{P}_{2^i}(x) = \hat{P}_2(\hat{P}_{2^{i-1}}(x))$ for $i > 1$. (For $2^{i-1} < k \leq 2^i$, let $\hat{P}_k(x) = \hat{P}_{2^i}(x)$.) Thus Toda's $\hat{P}_k$ has degree $\Theta(k^2)$. Noting that positive coefficients are not necessary for his applications (In retrospect, positive coefficients are not necessary for Toda's applications either), Yao (1985) constructed k-modulus-amplifying polynomials P_k starting with $P_2(x) = 3x^2 - 2x^3$ and defining P_{2^i} by the same recurrence. Yao's P_k has degree $\Theta(k^{\log_2 3})$. Now we put modulus-amplifying polynomials in better perspective and construct optimal-degree modulus-amplifying polynomials.

The conditions for P_k above are equivalent to the following congruences in $\mathbf{Z}[x]$, the ring of polynomials in one variable over $\mathbf{Z}$.

$$P_k(x) \equiv 0 \pmod{x^k}; \tag{2.1}$$
$$P_k(x) \equiv 1 \pmod{(x-1)^k}. \tag{2.2}$$

The polynomials x^k and $(x-1)^k$ are relatively prime in $\mathbf{Z}[x]$; i.e., there are polynomials $f(x)$, $g(x) \in \mathbf{Z}[x]$ such that $f(x)x^k + g(x)(x-1)^k = 1$. This follows from the solution we give below. (Alternatively we can argue as follows: In a commutative ring, two ideals (in our case (x^k) and $((x-1)^k)$ are relatively prime if and only if their radical ideals are relatively prime. (For a proof, see a textbook on commutative algebra, e.g., Atiyah & MacDonald (1969, Proposition 1.16., p.9).) But the radicals of (x^k) and $((x-1)^k)$ are (x) and $(x-1)$ respectively, and are clearly relatively prime. In fact x^k and $(x+s)^k$ are relatively prime in $\mathbf{Z}[x]$ if and only if $s = 1$ or -1.) Thus by the Chinese remainder theorem (applied for the ring $\mathbf{Z}[x]$), the equations (2.1) and (2.2) have a unique solution in $\mathbf{Z}[x]$ modulo $x^k(x-1)^k$. We explicitly solve (2.1) and (2.2) and get a degree $2k-1$ solution, thus achieving the optimal degree.

Consider $(1-x)^k$ in $\mathbf{Z}[[x]]$, the ring of formal power series over $\mathbf{Z}$. Since its constant term is 1, it is invertible in $\mathbf{Z}[[x]]$; i.e., we can find $R_k \in \mathbf{Z}[[x]]$ such that $(1-x)^k R_k = 1$. Throw away all the terms in the power series R_k of degree k and higher, and obtain the *polynomial* Q_k. Then $1-(1-x)^k Q_k$ is a solution to (2.1) and (2.2). In fact,

$$R_k = \frac{1}{(1-x)^k} = (1 + x + x^2 + \cdots)^k = \sum_{j \geq 0} \binom{k+j-1}{j} x^j.$$

Thus

$$Q_k = \sum_{j=0}^{k-1} \binom{k+j-1}{j} x^j,$$

and

$$P_k = (-1)^{k+1}(x-1)^k \left(\sum_{j=0}^{k-1} \binom{k+j-1}{j} x^j \right) + 1$$

is our solution. Note that the norm of P_k is $2^{O(k)}$. In what follows, P_k denotes the polynomial constructed above. The effect of using our P_k's instead of the polynomials constructed by Toda and Yao will be mentioned in Remark 2.9.

2.4. Collapse by Modulus Amplification. For an integer a and a positive integer m, define $a \ \overline{\mathrm{mod}} \ m$ to be the unique integer in the range $[-\alpha, \beta]$ that is congruent to a modulo m, where $\alpha = \beta = \frac{m-1}{2}$ if m is odd; and $\alpha = \frac{m}{2} - 1$, $\beta = \frac{m}{2}$ if m is even. Note that $a = a \ \overline{\mathrm{mod}} \ m$ if $m \geq 2|a| + 1$. Also note the following obvious property of the norm: If a polynomial $p(x_1, \ldots, x_n)$ has norm N, then for any $x \in \{0,1\}^n$, $-N \leq p(x) \leq N$.

Fact 2.7. *Suppose that a polynomial $r(x_1, \ldots, x_l)$ has norm N and that positive integers m and k satisfy $m^k \geq 2N + 1$. Let $a_1, \ldots, a_l$ be integers satisfying $(a_i \bmod m) \in \{0,1\}$ $(1 \leq i \leq l)$. Then,*

$$r(a_1 \bmod m, \ldots, a_l \bmod m) = r(P_k(a_1), \ldots, P_k(a_l)) \ \overline{\mathrm{mod}} \ m^k.$$

Proof.

$$\begin{aligned}
& r(a_1 \bmod m, \ldots, a_l \bmod m) \\
= \ & r(P_k(a_1) \bmod m^k, \ldots, P_k(a_l) \bmod m^k) \\
= \ & r(P_k(a_1) \bmod m^k, \ldots, P_k(a_l) \bmod m^k) \ \overline{\mathrm{mod}} \ m^k \\
= \ & r(P_k(a_1), \ldots, P_k(a_l)) \ \overline{\mathrm{mod}} \ m^k. \ \square
\end{aligned}$$

The following lemma says that we can "collapse" stratified modular polynomial circuits using modulus-amplifying polynomials P_k, and, combined with Lemma 2.5, lets us finish the proof that $\mathrm{ACC} \subseteq \mathrm{SYM}^+$. The lemma is stated in a bit more general setting than necessary: It allows moduli of order $\log^{O(1)} n$ instead of $O(1)$.

Lemma 2.8. *Let $\{C_n\}$ be a family of stratified modular polynomial circuits of depth $O(1)$, size $2^{\log^{O(1)} n}$, degree $\log^{O(1)} n$ and norm $2^{\log^{O(1)} n}$, and modulus size $\log^{O(1)} n$. Then, the language recognized by C is in SYM^+, i.e., there exist a family $\{r_n(x_1, \ldots, x_n)\}$ of degree-$\log^{O(1)} n$ norm-$2^{\log^{O(1)} n}$ polynomials and a family $\{h_n\}$ of functions from $\mathbf{Z}$ to $\{0,1\}$ such that for each n and each $x \in \{0,1\}^n$, $C_n(x) = h_n(r_n(x_1, \ldots, x_n))$.*

Proof. Let $\{C_n\}$ be as above. Fix n and let $d = \mathrm{depth}(C_n)$. The proof is by induction on d. For the base case $d = 1$, the output gate of C_n is associated with a polynomial $r_n(x_1, \ldots, x_n)$ of degree $\log^{O(1)} n$ and norm $2^{\log^{O(1)} n}$ and a function $h_n : \mathbf{Z} \to \{0,1\}$, and there is nothing to prove since by definition of modular polynomial circuit, $C_n(x) = h_n(r_n(x))$. For the case $d \geq 2$:

- Let $\{g_1, \ldots, g_l\}$ be the set of gates at depth 1 and assume that the output gate g is associated with a polynomial $r(g_1, \ldots, g_l)$ and a function $h : \mathbf{Z} \to \{0,1\}$.
- Let $\{y_1, \ldots, y_\nu\}$ be the set of gates at depth 2 and assume that the gates $g_1, \ldots, g_l$ at depth 1 are associated with polynomials $r_1(y_1, \ldots, y_\nu), \ldots, r_l(y_1, \ldots, y_\nu)$ respectively and with a common modulus $m = \log^{O(1)} n$. (r_i may be a polynomial in the variables that form a proper subset of $\{y_1, \ldots, y_\nu\}$; but such a polynomial can be regarded as a polynomial in $y_1, \ldots, y_\nu$; this simplifies notations below.)

We show that we can collapse these top two levels. Take $k = \log^{O(1)} n$ large enough so that $m^k \geq 2\,\mathrm{norm}(r) + 1$. Recall that in a modular polynomial circuit, for each $\vec{y} = (y_1, \ldots, y_\nu) \in \{0,1\}^\nu$, $(r_i(\vec{y}) \bmod m) \in \{0,1\}$; thus by Fact 2.7,

$$\begin{aligned} & r(r_1(\vec{y}) \bmod m, \ldots, r_l(\vec{y}) \bmod m) \\ = \; & r(\, P_k(r_1(\vec{y})), \ldots, P_k(r_l(\vec{y}))\,) \;\overline{\bmod}\; m^k. \end{aligned}$$

Thus let

$$r'(\vec{y}) = r(\, P_k(r_1(\vec{y})), \ldots, P_k(r_l(\vec{y}))\,)$$

and define $h' : \mathbf{Z} \to \{0,1\}$ by

$$h'(z) = h(z \;\overline{\bmod}\; m^k).$$

Then for each $\vec{y} = (y_1, \ldots, y_\nu) \in \{0,1\}^\nu$,

$$\begin{aligned} & h(\, r(r_1(\vec{y}) \bmod m, \ldots, r_l(\vec{y}) \bmod m)\,) \\ = \; & h(r'(\vec{y}) \;\overline{\bmod}\; m^k) \\ = \; & h'(r'(\vec{y})). \end{aligned}$$

It is not hard to see that the polynomial r' has degree $\log^{O(1)} n$ and norm $2^{\log^{O(1)} n}$.

Introduce a new output gate g' and associate with g' polynomial r' and function h'. Let C'_n be the new circuit thus obtained. Then C'_n is equivalent to C_n and has depth $d-1$, size smaller than that of C_n, degree $\log^{O(1)} n$, and norm $2^{\log^{O(1)} n}$. The inductive step is complete, and we have proved the lemma. □

Now the proof of Theorem 1.1 is complete.

Remark 2.9. *Let δ denote the degree of the polynomial obtained by our proof method. For a polynomial-size depth-d ACC circuit, $\delta = \log^{2^{\Theta(d)}} n$ and the norm of the polynomial is $n^{\log^{2^{\Theta(d)}} n}$. The degree and the norm correspond, respectively, to the bottom fan-in and the size of depth-two circuits that characterize SYM$^+$.*

More specifically: Let C be a stratified polynomial-size depth-d (assume $d \geq 2$) circuit having only MOD gates and such that all the MOD gates at the same depth i are MOD_{p_i} gates for some prime p_i of order $O(1)$. Then $\delta = \Theta(\log^{(\alpha+1)^{d-1}-1} n)$, where $\alpha = 1$ in our case, and α would be $\log_2 3$ and 2 if one uses Yao's and Toda's modulus-amplifying polynomials respectively.

Now consider a circuit that is similar to C, but has AND/OR gates in addition and assume that we use our degree $2k-1$ modulus-amplifying polynomials P_k. (The analysis below remains valid as far as the degree is $O(k)$.)

In this case, $\delta = \Theta(\log^{\beta} n)$, where $\beta = 2^{d+1} - 2$ if one uses Razborov–Smolensky polynomials together with a nonconstructive argument as explained above, and $\beta = 2^{d+2} - 3$ if one uses the Valiant–Vazirani method.

Remark 2.10. *The function $h_n : \mathbf{Z} \rightarrow \{0, 1\}$ obtained in the proof above has the following form.*

$$h_n(N) = \begin{cases} 1 & \text{if } (\cdots((N \overline{\bmod} M_1) \overline{\bmod} M_2) \cdots \overline{\bmod} M_c) \geq \lceil S/2 \rceil; \\ 0 & \text{otherwise,} \end{cases}$$

where each M_i is a power of prime. Recently, F.Green et al. *(1992) have shown that h_n can be taken to be a "Mid-Bit" function, whose value on an integer N is the $t(n)$-th least significant bit of the standard binary expansion of N for some $t(n) = \log^{O(1)} n$. Barrington's 1992 survey includes explanations of other recent work that is related to this paper.*

Remark 2.11. *Theorem 1.1 still holds when we allow a modulus m of MOD gates to* grow *as far as $m = \log^{O(1)} n$ and m has only $O(1)$ distinct prime factors.*

On the other hand, if we allow $O(\log n / \log\log n)$ distinct prime moduli of magnitude $O(\log n)$, any symmetric Boolean function on n variables can be computed by a circuit of the form: an OR at the root, at most n ANDs of fan-in $O(\log n)$ at the next level, and $O(\log^2 n / \log\log n)$ MODs at the bottom. Thus if a similar result holds in this case, then TC^0 *(the class of languages recognized by constant-depth polynomial-size threshold circuits) is contained in* SYM$^+$.

3. Proof of Proposition 1.2

We proceed as we did to prove ACC $\subseteq$ SYM$^+$: Show that we can convert such circuits as in the proposition to equivalent low-degree small-norm modular polynomial circuits, and use Lemma 2.8 to finish the proof.

Let C be a circuit of size $N = 2^{\log^{O(1)} n}$ with a fan-in F output gate g computing a symmetric function and F ACC subcircuits $C_1, \ldots, C_F$ below g. The symmetric-function gate g computes some Boolean function that only depends on $\sum_{i=1}^{F} C_i(x)$. Thus we want a construction that can determine, for each x, the number of i's $(1 \le i \le F)$ such that $C_i(x) = 1$. We proceed as follows:

1. Using a probabilistic polynomial of sample size $S = 2^{\log^{O(1)} n}$ that computes OR/AND with error probability at most $1/(3FN)$, convert each C_i $(1 \le i \le F)$ to a stratified modular probabilistic polynomial circuit that computes C_i with error probability at most $\varepsilon = 1/(3F)$.

2. For each C_i, by fixing a probabilistic polynomial to be each of the S ordinary polynomials in its sample space, create S stratified modular (ordinary) polynomial circuits and let $\Psi_i = \{g_i^{(1)}, \ldots, g_i^{(S)}\}$ be the set of output gates of those S circuits. Clearly, for each input x and each set Ψ_i, one of the following two cases holds:

   (a) At most εS $g_i^{(j)}$'s $(1 \le j \le S)$ output 1.

   (b) At least $(1-\varepsilon)S$ $g_i^{(j)}$'s $(1 \le j \le S)$ output 1.

   For a rational number r that is not of the form $j + \frac{1}{2}$ for some integer j, let nearest-int(r) denote the unique integer k that minimizes $|r - k|$. It is easy to see that the number of sets Ψ_i for which the case (b) holds equals

   $$\text{nearest-int}\left(\frac{\sum_{i=1}^{F}\sum_{j=1}^{S} g_i^{(j)}}{S}\right).$$

3. Connect the $g_i^{(j)}$'s to a new output gate g' and associate with g' the linear polynomial $\sum_{i=1}^{F}\sum_{j=1}^{S} g_i^{(j)}$ and the function $h(y) = \overline{h}(\text{nearest-int}(y/S))$, where $\overline{h} : \{0, \ldots, F\} \to \{0, 1\}$ is the function computed by the symmetric-function output gate of C (as expressed in terms of the number of inputs that are 1).

We have obtained an equivalent stratified modular polynomial circuit of constant depth, size $2^{\log^{O(1)} n}$, degree $\log^{O(1)} n$, and norm $2^{\log^{O(1)} n}$. We can use Lemma 2.8 to finish the proof. □

Acknowledgements

Thanks to Nick Reingold for helpful discussions, Andy Yao for describing his result to us, Uli Hertrampf for proving some neat facts about MODs, and Bill Gasarch for suggesting that we try to improve Yao's bounds. We also thank the anonymous referees for making many suggestions for improving the presentation of the paper.

A preliminary version of this paper appeared as Beigel & Tarui (1991). The first author gratefully acknowledges partial support from the U.S.A. National Science Foundation under grant CCR-8958528. The second author performed some of this work at the University of Rochester (where he was supported in part by NSF grant CDA-8822724) and at the University of Warwick (where he was supported in part by the ESPRIT II BRA Programme of the EC under contract # 7141 [ALCOM II]).

References

M. Ajtai, Σ_1^1 formulae on finite structures. *Annals of Pure and Applied Logic* **24** (1983), 1–48.

E. Allender, A note on the power of threshold circuits. In *Proceedings of the 30th Ann. Symp. Found. Comput. Sci.*, 1989, 580–584.

E. Allender and U. Hertrampf, Depth reduction for circuits of unbounded fan-in. *Inform. and Comput.* **108** (1994). To appear.

N. Atiyah and I. MacDonald, *Introduction to Commutative Algebra*. Addison-Wesley, 1969.

D. A. Barrington, Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *J. Comput. System Sci.* **38**(1) (1989), 150–164.

D. A. M. Barrington, Quasipolynomial size circuit classes. In *Proceedings of the 7th Ann. Conf. Structure in Complexity Theory*. IEEE, 1992, 86–93.

D. A. M. Barrington and D. Thérien, Finite monoids and the fine structure of NC^1. *J. Assoc. Comput. Mach.* **35**(4) (1988), 941–952.

R. Beigel, The polynomial method in circuit complexity. In *Proceedings of the 8th Ann. Conf. Structure in Complexity Theory*, 1993, 82–95.

R. Beigel and J. Gill, Counting classes: Thresholds, parity, mods, and fewness. *Theoret. Comput. Sci.* **103**(1) (1992), 3–23.

R. Beigel and J. Tarui, On ACC. In *Proceedings of the 32nd Ann. Symp. Found. Comput. Sci.*, 1991, 783–792.

R. Beigel, N. Reingold, and D. Spielman, The perceptron strikes back. In *Proceedings of the 6th Ann. Conf. Structure in Complexity Theory*, 1991, 286–291.

R. Beigel, N. Reingold, and D. Spielman, PP is closed under intersection. *J. Comput. System Sci.* **48** (1994). To appear.

R. Boppana and M. Sipser, The complexity of finite functions. In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity*, ed. J. van Leeuwen, 757–804. MIT Press and Elsevier, The Netherlands, 1990.

A. K. Chandra, L. Stockmeyer, and U. Vishkin, Constant depth reducibility. *SIAM J. Comput.* **13**(2) (1984), 423–438.

F.Green, J. Köbler, and J. Torán, The power of the middle bit. In *Proceedings of the 7th Ann. Conf. Structure in Complexity Theory*. IEEE, 1992, 111–117. An extended version has been drafted by Green, Köbler, Regan, Schwentick, and Torán.

M. Furst, J. B. Saxe, and M. Sipser, Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory* **17**(1) (1984), 13–27.

J. T. Håstad, *Computational Limitations for Small-Depth Circuits.* ACM Doctoral Dissertation Award. MIT Press, Cambridge, MA, 1986.

D. Johnson, A catalog of complexity classes. In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity*, 69–161. MIT Press and Elsevier, 1990.

R. Kannan, H. Venkateswaran, V. Vinay, and A. C. Yao, A circuit-based proof of Toda's theorem. *Inform. and Comput.* (1993). To appear.

P. McKenzie and D. Thérien, Automata theory meets circuit complexity. In *Proceedings of the 16th ICALP*. Springer-Verlag, 1989, 589–602. Lecture Notes in Computer Science 372.

A. A. Razborov, Lower bounds for the size of circuits of bounded depth with basis $\{\wedge, \oplus\}$. *Math. notes of the Academy of Science of the USSR* **41**(4) (1987), 333–338.

M. Sipser, The history and status of the P versus NP question. In *Proceedings of the 24th Ann. ACM Symp. Theor. Comput.*, 1992, 603–618.

R. Smolensky, Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th Ann. ACM Symp. Theor. Comput.*, 1987, 77–82.

J. Tarui, Probabilistic polynomials, AC^0 functions, and the polynomial-time hierarchy. *Theoret. Comput. Sci.* **113** (1993), 167–183.

S. Toda, PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput.* **20**(5) (1991), 865–877.

S. Toda and M. Ogiwara, Counting classes are at least as hard as the polynomial-time hierarchy. *SIAM J. Comput.* **21**(2) (1992), 316–328.

A. C. Yao, Separating the polynomial-time hierarchy by oracles. In *Proceedings of the 26th Ann. Symp. Found. Comput. Sci.*, 1985, 1–10.

Manuscript received March 29, 1993

Richard Beigel
Dept. of Computer Science
P.O. Box 208285
New Haven, CT 06520-8285
U.S.A.
beigel-richard@cs.yale.edu

Jun Tarui
Department of Communications & Systems Engineering
University of Electro-Communications
Chofu, Tokyo, 182
Japan
jun@ttl.cas.uec.ac.jp