SIAM J. COMPUT.
Vol. 29, No. 4, pp. 1209–1222

LOWER BOUNDS FOR ($\mathrm{MOD}_p - \mathrm{MOD}_m$) CIRCUITS*

VINCE GROLMUSZ† AND GÁBOR TARDOS‡

Abstract. Modular gates are known to be immune for the random restriction techniques of Ajtai (1983), Furst, Saxe, and Sipser (1984), Yao (1985), and Håstad (1986). We demonstrate here a random clustering technique which overcomes this difficulty and is capable of proving generalizations of several known modular circuit lower bounds of Barrington, Straubing, and Thérien (1990), Krause and Pudlák (1994), and others, characterizing symmetric functions computable by small $(\mathrm{MOD}_p, \mathrm{AND}_t, \mathrm{MOD}_m)$ circuits.

Applying a degree-decreasing technique together with random restriction methods for the AND gates at the bottom level, we also prove a hard special case of the constant degree hypothesis of Barrington, Straubing, and Thérien (1990) and other related lower bounds for certain $(\mathrm{MOD}_p, \mathrm{MOD}_m, \mathrm{AND})$ circuits.

Most of the previous lower bounds on circuits with modular gates used special definitions of the modular gates (i.e., the gate outputs one if the sum of its inputs is divisible by m or is *not* divisible by m) and were not valid for more general MOD_m gates. Our methods are applicable, and our lower bounds are valid for the most general modular gates as well.

Key words. lower bounds, modular gates, composite modulus

AMS subject classifications. 68Q05, 68Q15, 68Q22

PII. S0097539798340850

1. Introduction.

Boolean circuits are perhaps the most widely examined models of computation. They gain application in diverse areas as VLSI design, complexity theory, and the theory of parallel computation.

A majority of the strongest and deepest lower bound results for computational complexity were proved using the Boolean circuit model of computation (for example, [12], [16], [8], [13], [14], or see [4] for a survey).

Unfortunately, lots of questions, even for very restricted circuit classes, have been unsolved for a long time.

Bounded depth and polynomial size is a natural restriction. Ajtai [1] and Furst, Saxe, and Sipser [6] proved that no polynomial sized, constant-depth circuit can compute the PARITY function. Yao [16] and Håstad [8] generalized this result for sublogarithmic depths. Their technique involved a sophisticated use of *random restriction techniques*, in which randomly assigned 0-1 values to the input variables fixed the output of large fan-in AND and OR Boolean gates.

Since the modular gates are very simple to define, and they are immune to the random restriction techniques in lower bound proofs for the PARITY function, the following natural question was asked by Barrington, Smolensky and others: How powerful will the Boolean circuits be if, beside the standard AND, OR, and NOT gates, MOD_m gates are also allowed in the circuit? Here, a MOD_m^A gate outputs 1 if the sum of its inputs is in a set $A \subset \{0, 1, 2, \ldots, m-1\}$ modulo m.

*Received by the editors June 24, 1998; accepted for publication (in revised form) May 18, 1999; published electronically February 10, 2000. This work was supported in part by grants OTKA F014919, FKFP 0835, and AKP 97-56 2,1.

http://www.siam.org/journals/sicomp/29-4/34085.html

†Department of Computer Science, Eötvös University, Rákóczi út 5, H-1088 Budapest, Hungary (grolmusz@cs.elte.hu).

‡Rényi Institute of the Hungarian Academy of Science, Reáltanoda u. 13-15, H-1055 Budapest, Hungary (tardos@cs.elte.hu).

Razborov [13] showed that for computing MAJORITY with AND, OR, NOT, and MOD_2 gates, exponential size is needed with constant depth. This result was generalized by Smolensky [14] for MOD_p gates instead of MOD_2 gates, where p denotes a prime.

We know very little, however, if both MOD_p and MOD_q gates are allowed in the circuit for different primes p, q, or if the modulus is a nonprime power composite, e.g., 6. For example, it is consistent with our present knowledge that depth-3, linear-sized circuits with MOD_6 gates *only* recognize an NP-complete language (see [2]).

It is not difficult to see that constant-depth circuits with MOD_p gates only (p prime) cannot compute even very simple functions—the n-fan-in OR or AND functions—since they can compute only constant degree polynomials of the input variables over GF_p (see [14]).

But depth-2 circuits with MOD_2 and MOD_3 gates or MOD_6 gates can compute the n-fan-in OR and AND functions [9], [2]. Consequently, these circuits are more powerful than circuits with MOD_p gates only. The sketch of the construction is as follows: we take a MOD_3 gate at the top of the circuit and 2^n MOD_2 gates on the next level, where each subset of the n input variables is connected to exactly one MOD_2 gate, and then this circuit computes the n-fan-in OR, since if at least one of the inputs is 1, then exactly half of the MOD_2 gates evaluate to 1.

Barrington, Straubing, and Thérien [2] conjectured that any $(\mathrm{MOD}_p^B, \mathrm{MOD}_m^A, \mathrm{AND}_d)$ circuit needs exponential size to compute the n-fan-in AND function, where the prime p and the positive integers m and d are fixed and AND_d denotes the fan-in d AND function. They called it the *constant degree hypothesis* (CDH) and proved the $d = 1$ case, with highly nontrivial algebraic techniques. Their proof also works for depth-$(\ell + 1)$

$$(\overbrace{\mathrm{MOD}_{p^k}^B, \mathrm{MOD}_{p^k}^B, \ldots, \mathrm{MOD}_{p^k}^B}^{\ell}, \mathrm{MOD}_m^A) \tag{1.1}$$

circuits, computing the AND function.

Yan and Parberry [15], using Fourier-analysis, also proved the $d = 1$ case for $(\mathrm{MOD}_p^{\{1,2,\ldots,p-1\}}, \mathrm{MOD}_2^{\{1\}})$ circuits, but their method also works for the special case of the CDH where the sum of the degrees of the monomials g_i on the input-level satisfies

$$\sum_{\deg(g_i)\geq 1} (\deg(g_i) - 1) \leq \frac{n}{2(p-1)} - O(1).$$

Krause and Waack [11] applied communication-complexity techniques to show that any $(\mathrm{MOD}_m^{\{1,2,\ldots,m-1\}}, \mathrm{SYMMETRIC})$ circuit, computing the ID function

$$\mathrm{ID}(x, y) = \begin{cases} 1 \text{ if } x = y, \\ 0 \text{ otherwise} \end{cases}$$

for $x, y \in \{0,1\}^n$, should have size at least $2^n/\log m$, where SYMMETRIC is a gate, computing an arbitrary symmetric Boolean function. Since (non-weighted) MOD_m gates are also SYMMETRIC gates, this lower bound is valid for $(\mathrm{MOD}_m^{\{1,2,\ldots,m-1\}}, \mathrm{MOD}_m^A)$ circuits. When mod m coefficients (or multiple wires) are allowed on the input-level, then the MOD_m gates are *not* SYMMETRIC gates, but the same proof techniques remain applicable. Caussinus [5] proved that the result of

[2] also implies a similar lower bound for the AND function. Unfortunately, results [11], [5] do not generalize for the more general MOD_m^A gates at the top.

Krause and Pudlák [10] proved that any $(\mathrm{MOD}_{p^k}^{\{0\}}, \mathrm{MOD}_q^{\{0\}})$ circuit which computes the $\mathrm{MOD}_r^{\{0\}}$ function has size at least 2^{cn}, for some $c > 0$, where p and r are different primes and q is not divisible by either of them.

Our main result is a characterization of those symmetric Boolean functions which are computable by quasi-polynomial-size

$$(\overbrace{\mathrm{MOD}_{p^k}^B, \mathrm{MOD}_{p^k}^B, \ldots, \mathrm{MOD}_{p^k}^B}^{\ell}, \mathrm{MOD}_m^A)$$

circuits. We prove (Theorem 2.5) that the *only* symmetric functions that are computable by such circuits are the MOD_{mp^j} functions with small j. Consequently, the nontrivial threshold functions (and thus also AND and OR) and the $\mathrm{MOD}_r^{\{0\}}$ functions if r does not divide $p^j m$ need exponential size on that circuits. Even MOD_4 requires exponential size $(\mathrm{MOD}_{3^r}, \mathrm{AND}_t, \mathrm{MOD}_2)$ circuits for constant t and r. Note the asymmetry: MOD_4 is easy to compute with a polynomial size $(\mathrm{MOD}_2, \mathrm{AND}_3)$ circuit. These results generalize the theorems of Barrington, Straubing, and Thérien [2] and Krause and Pudlák [10] and give a characterization of the computable symmetric functions, instead of singular lower bounds.

Grolmusz [7] generalized the results of [2], [15], [11], [10] for $(\mathrm{MOD}_q, \mathrm{MOD}_p, \mathrm{AND}_{cn})$ circuits, where the input-polynomials of each MOD_p gate are constructible from linear terms using at most $cn - 1$ multiplications (or, equivalently, can be computed by an arithmetic circuit of an arbitrary number of mod p additions and at most $cn-1$ fan-in 2 multiplications). In particular, one can allow the sum of *an arbitrary function* of cn variables and a linear polynomial of the n variables as inputs for each MOD_p gate. We generalize this result, too (Lemma 3.12). The main tool of the proof of [7] is a degree decreasing lemma, which we also generalize here for nonprime moduli (Lemma 3.9), and we use it both for lower and upper bound proofs.

Here we generalize the results of [7]: we prove a lower bound on the size of the $(\mathrm{MOD}_p, \mathrm{MOD}_m, \mathrm{AND})$ circuits computing AND_n if m is a positive integer, p is a prime, and each MOD_m gate has not-too-many AND gates as inputs and those AND gates have low fan-in. For the exact statement see Theorem 2.6. This is an important special case of the CDH of [2]. The lower bound also applies to circuits computing some other functions besides AND.

2. Our results.

2.1. Ideas. MOD_m gates are immune to random restriction techniques, since these gates remain MOD_m gates on the remaining variables after an arbitrary restriction, and thus (unless less than m variables remain unrestricted) the complexity does not decrease.

We overcome this difficulty by a *random clustering* technique, which forces some randomly chosen variables to be equal. Each equivalence class (or cluster) will make a new variable of the MOD_m gate, and each new variable will be invisible (i.e., its coefficient will be a multiple of m) for the gate with a constant probability (Lemma 3.4).

We use this for $(\mathrm{MOD}_p, \mathrm{AND}_t, \mathrm{MOD}_m)$ circuits, computing symmetric functions. Suppose that the equivalence classes are of size m; then the resulting function of the new, clustered variables is a unique symmetric function.

Almost all symmetric functions (except the $\mathrm{MOD}_{p^k m}$ functions) have large restrictions, whose unique factor resulting from the clustering above cannot be expressed as a modulo p sum of functions, none of which depends on all variables. An exponential lower bound follows for the number of AND gates on level 2 (Theorem 2.4).

If we have $o(n^2/\log n)$ constant-degree monomials as inputs for each MOD_m gates on level 2, then by random restrictions, one can essentially decrease their number, and a small number of low-degree monomials can be converted to linear polynomials with the help of the degree-decreasing lemma (Lemma 3.9), and we can apply Theorem 2.4 to get lower bounds. (Theorem 2.6)

2.2. Preliminaries.

DEFINITION 2.1. *A fan-in n gate is an n-variable Boolean function. Let $G_1, G_2, \ldots, G_\ell$ be gates of unbounded fan-in. Then a $(G_1, G_2, \ldots, G_\ell)$-circuit denotes a depth-ℓ circuit with a G_1-gate on the top, G_2 gates on the second level, G_3 gates on the third level from the top,..., and G_ℓ gates on the last level.* AND_t *denotes the fan-in t* AND *gate. The size of a circuit is defined to be the total number of the gates in the circuit.*

All of our modular gates are of unbounded fan-in, and we allow for connecting inputs to gates or gates to gates with multiple wires. Note that, by this definition, our modular gates are not symmetric gates in general.

In the literature MOD_m gates are sometimes defined to be 1 iff the sum of their inputs is divisible by m, and sometimes they are defined to be 1 iff the sum of their inputs is not divisible by m. The following, more general definition covers both cases.

DEFINITION 2.2. *We say that gate G is a* MOD_m-gate *if there exists $A \subset \{0, 1, \ldots, m-1\}$ such that*

$$G(x_1, x_2, \ldots, x_n) = \begin{cases} 1 \text{ if } \sum_{i=1}^{n} x_i \bmod m \in A, \\ 0 \text{ otherwise.} \end{cases}$$

A is called the 1-set of G. MOD_m *gates with 1-set A are denoted by* MOD_m^A.

NOTATION 2.3. *Let* $\Sigma_p(x_1, x_2, \ldots, x_s) = \sum_{i=1}^{s} x_i \bmod p$.

In general, Σ_p is not a Boolean gate, since its value is from $\{0, 1, \ldots, p-1\}$. However, in all of our statements, its value will be guaranteed to be 0 or 1.

2.3. Theorems. Here we list the three main results of this paper. To be concise we use $((\mathrm{MOD}_{p^k}^B)^\ell, \mathrm{MOD}_m^A)$ to denote circuits of type (1.1). Note that standard techniques (see Lemma 3.2) show that these circuits are equivalent to $(\sum_p, \mathrm{AND}_t, \mathrm{MOD}_m^A)$ circuits, and we could have stated Theorems 2.4 and 2.5 for those circuits instead.

THEOREM 2.4. *Suppose that a circuit of type $((\mathrm{MOD}_{p^k}^B)^\ell, \mathrm{MOD}_m^A)$ with p prime computes a symmetric Boolean function f on n variables, such that $f \neq \mathrm{MOD}_{p^j m}^A$ for any A. Then its size S is exponential in p^j; i.e., there exists a number $c > 1$ depending on p, m, k, and ℓ such that $S > c^{p^j}$.*

As a special case we get that the size S of an n-variable circuit of type $((\mathrm{MOD}_{p^k}^B)^\ell, \mathrm{MOD}_m^A)$ with p prime computing any of the nontrivial threshold functions (including AND *and* OR*) or the $\mathrm{MOD}_r^{\{0\}}$ function (where r does not divide mp^j for any j) is exponential in n. We have $S > c^n$ for a number $c > 1$ depending only on p, m, k, and ℓ.*

THEOREM 2.5. *Let the prime p and the positive integers m, k, and ℓ be fixed with m not a power of p. The symmetric functions computed by a type $((\mathrm{MOD}_{p^k}^B)^\ell, \mathrm{MOD}_m^A)$*

circuit of quasi-polynomial size are exactly the functions $\mathrm{MOD}^C_{mp^j}$ *with* $j = O(\log\log n)$ *and* $C \subset \{0, 1, \ldots, mp^j - 1\}$.

On the other hand, all the functions $\mathrm{MOD}^C_{mp^j}$ *with* $j = O(\log\log n)$ *can be computed by quasi-polynomial size* $(\sum_p, \mathrm{AND}_2, \mathrm{MOD}_m)$ *circuits.*

Our final result proves a special case of the CDH.

THEOREM 2.6. *Let* p *be prime and* m *a fixed positive integer. Suppose that a* $(\mathrm{MOD}^B_p, \mathrm{MOD}^A_m, \mathrm{AND})$ *circuit computes* AND_n. *If each* MOD_m *gate has fan-in* $o(n^2/\log n)$ *and each* AND *gate has constant fan-in, then the size of the circuit is super-polynomial.*

We remark that this result is a consequence of the tradeoff between the size of $(\Sigma_p, \mathrm{MOD}_m, \mathrm{AND})$ circuits computing AND and a new measure introduced here, the number of pairs of input variables the MOD_m gates relate (see Theorems 3.13 and 3.14). Note that similar bounds can be proved for circuits computing many other natural functions, like threshold or MOD_r functions.

3. The proofs.

3.1. Eliminating the top gate.

The top-gate elimination is widely used in the literature (cf. [10, Lemma 5.2] or [3]). It replaces the top MOD_{p^r} gate with constant fan-in AND gates and a simple summation modulo p with a polynomial increase in the size.

LEMMA 3.1. *Let* p *be a prime,* k *a positive integer, and* $A \subset \{0, 1, \ldots, p^k - 1\}$. *There is a modulo* p *polynomial of degree* $p^k - 1$ *computing the* $\mathrm{MOD}^A_{p^k}$ *function.* ∎

One can repeatedly use this lemma to eliminate a constant-depth subcircuit of MOD_{p^r} gates from the top of any circuit, as stated by the next lemma.

LEMMA 3.2. *Suppose that* $f : \{0,1\}^n \to \{0,1\}$ *is computed by a depth-*$(\ell + 1)$

$$\overbrace{(\mathrm{MOD}^A_{p^k}, \ldots, \mathrm{MOD}^A_{p^k}}^{\ell}, G)$$

circuit, where p *is a prime and on the input level we have arbitrary gates (or subcircuits)* G. *Suppose the number of these gates* G *is* S. *Then* f *can also be computed from the same gates* G *by a* $(\Sigma_p, \mathrm{AND}_t, G)$ *circuit, with* $t < p^{k\ell}$ *and at most* $S^{p^{k\ell}}$ AND_t *gates on the middle level.*

Proof. By Lemma 3.1 all $\mathrm{MOD}^A_{p^k}$ can be replaced by a modulo p polynomial of degree less than p^k; thus f is degree $< p^{k\ell}$ polynomial of the output of the G gates. The bound on the size comes from counting all the possible monomials in such a polynomial. ∎

Note that the size of the new circuit is still polynomial in S, and the fan-in of the AND gates is constant if the depth ℓ and the modulus p^k are constants. Note also that AND_t gates with $t < p^k$ can be considered as special MOD_{p^k} gates, and thus AND_t gates can be eliminated the same way.

3.2. Random clustering.

DEFINITION 3.3. *Let* $\sim$ *be an equivalence relation on the variables of a function* f. *By the factor* $f/\sim$ *of* f *we mean the function obtained from* f *by identifying variables according to* $\sim$. *The variables of* $f/\sim$ *correspond to the equivalence classes of* $\sim$. *For an integer* m *we call the* $f/\sim$ *an* m*-factor of* f *if each equivalence class in* $\sim$ *consists of* m *variables.*

We say that the Boolean function f is p-simple (p is a positive integer) if it can be expressed as a modulo p sum of functions, none of which depend on all of the variables.

Example. Suppose that f has six variables, and $x_1 \sim x_2, x_3 \sim x_4, x_5 \sim x_6$. Then $f/\sim$ is a 2-factor of f, has three variables, and is defined as

$$f/\sim(y_1, y_2, y_3) = f(y_1, y_1, y_2, y_2, y_3, y_3).$$

Notice that any factor of the AND function is again an AND function. The m-factor of a symmetric function is unique and it is also a symmetric function. Note that for prime numbers p a function f is p-simple iff it can be expressed as a modulo p polynomial of degree less than the number of its variables.

Implicitly, a random clustering technique was used in the paper of Krause and Pudlák [10]. However, our method gives stronger results more directly.

The following lemma is about a special type of three level circuits. It is stated in a more general way, but the reader may think of polynomial size, $(\sum_p, \mathrm{AND}_t, \mathrm{MOD}_m^A)$ circuits with constant t.

LEMMA 3.4. *Let p, m, and t be positive integers, $1 \geq \varepsilon > 0$, and suppose the Boolean function f on n variables satisfies $f \equiv \sum_{i=1}^{S} f_i \pmod p$, where each f_i is computed in an arbitrary way from t of the functions f_{ij} and from $(1-\varepsilon)n$ of the input variables. Each of the functions f_{ij} is in turn a modulo m linear combination of the input variables. Here the functions f_i output modulo p values while f_{ij} output modulo m values. If n is large enough and divisible by m and $S < c^n$, then there exists a p-simple m-factor of f, where the constant $c > 1$ depends only on m, t, and ε.*

Proof. The idea is to observe that $f_{ij}/\sim$ is a modulo m linear combination of its variables, and the coefficient of a variable, corresponding to an equivalence class in a random $\sim$, is equal to zero with a positive constant probability. Thus $f_i/\sim$ depends on all of its variables with exponentially small probability. Then, with high probability, all the functions $f_i/\sim$ have an invisible variable, and thus $f/\sim$ is p-simple.

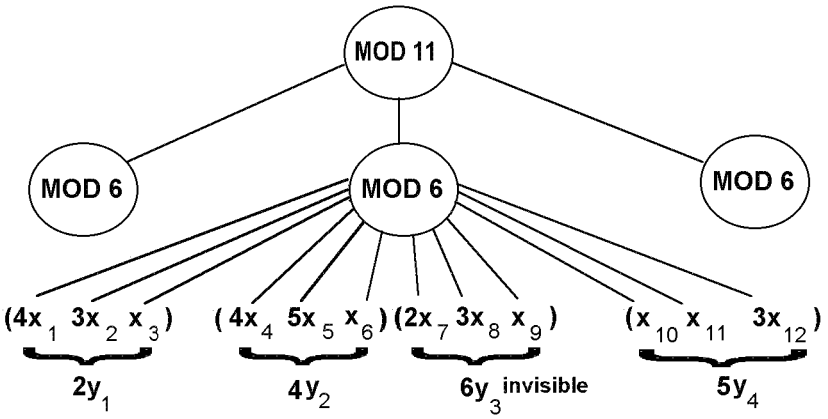

FIG. 3.1. *Random clustering in the simplest case: $t = 1, \varepsilon = 1$, and every f_{i1} is a* MOD_6 *gate.*

Let us choose $\sim$ uniformly at random from all the partitions of the variables into classes of size m. Consider choosing the equivalence classes one by one. Consider a fixed $1 \leq i \leq S$ and one of the first $\lceil \varepsilon n/(2m) \rceil$ classes. When we choose the

variables of this class, there are less than $\varepsilon n/2$ variables already in previous classes and at most $(1-\varepsilon)n$ variables directly seen by f_i, so the set H of the remaining variables has at least $\varepsilon n/2$ elements. Each variable in H has a coefficient in each of the combinations f_{ij}. Let $(a_j)_{j=1}^t$ be a list of coefficients that are most popular, and call a variable in H good if its coefficient in f_{ij} is a_j for each value of j. There are at least $|H|/m^t \geq \epsilon n/(2m^t)$ good variables. When choosing the variables for our equivalence class each has a probability of at least $\epsilon/(2m^t)$ to be good. Despite the slight dependence among these events, the probability that each of them are good is still at least $(\epsilon/(4m^t))^m$ for large enough n. If this is the case, $f_i/\sim$ does not depend on the variable corresponding to this class, since f_i does not see it directly, and the coefficient of this variable in f_{ij} is $ma_j = 0 \bmod m$. Thus (using $(1-u)^k < e^{-uk}$) with probability at most $e^{-(\varepsilon n/(2m))\cdot(\varepsilon m^{-t}/4)^m}$ does $f_i/\sim$ depend on each of its variables. We choose $\ln c = (\varepsilon/4)^{m+1}/m^{tm+1}$. If $S < c^n$, then with positive probability none of the functions $f_i/\sim$ depend on all of the variables; consequently, $f/\sim \equiv \sum_{i=1}^S f_i/\sim \pmod p$ is p-simple. ∎

We remark here that the same proof gives that if S in the lemma is bounded by another exponential function of n, then a random m-factor of f can almost always be expressed as a modulo p sum of functions; none of the functions depends on an m^{-mt} fraction of their variables.

NOTATION 3.5. *Let $w(x)$ denote the weight of a zero-one vector x, i.e., the number of ones in x. Then $f(i)$ denotes the value of the symmetric Boolean function f on inputs of weight i.*

LEMMA 3.6. *Let p be a prime. If f is a symmetric Boolean function on p^k variables with $f(0) \neq f(p^k)$, then f is not p-simple.*

Proof. Notice that

$$\sum_{x\in\{0,1\}^n} (-1)^{w(x)} f(x) \equiv 0 \pmod p$$

for p-simple functions f. The left-hand side is zero for functions not depending on one of the input variables; thus it is divisible by p for a modulo p sum of such functions.

For a symmetric function on $n = p^k$ variables, the left-hand side of the last equation is

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} f(i) \equiv f(0) - f(n) \pmod p,$$

since p divides $\binom{p^k}{i}$ unless $i = 0$ or $i = p^k$. Thus $f(0) \neq f(n)$ implies that f has full p-degree as claimed. ∎

THEOREM 3.7. *Let p be a prime, m, t, k, and S positive integers, and $1 \geq \varepsilon > 0$. Suppose the symmetric Boolean function f on n variables is the modulo p sum of S of the functions f_i, where each of the f_i is computed in an arbitrary way from t of the functions f_{ij} and from $(1-\varepsilon)n$ of the input variables. Each of the functions f_{ij} is in turn a modulo m linear combination of the input variables. Here the functions f_i output modulo p values while f_{ij} output modulo m values. Suppose f is not equal to any MOD_{mp^k} gate. Then $S > c^{p^k}$ for a constant $c > 1$ depending only on m, t, and ε.*

Proof. Since f is not a MOD_{mp^k} gate, there exist numbers $0 \leq i < i+mp^k = j \leq n$ such that $f(i) \neq f(j)$. Restrict the function f by assigning 0 to $n-j$ of its variables

and assigning 1 to i of them. The resulting function f' is a symmetric function of its mp^k variables satisfying $f'(0) \neq f'(mp^k)$. Notice that the restriction does not increase the size of the circuit computing the function. The unique m-factor of f' is a symmetric function f'' on p^k variables satisfying $f''(0) \neq f''(p^k)$. By Lemma 3.6, f'' is not p-simple. Thus Lemma 3.4 gives the claimed bound on S. □

We are ready now to prove Theorem 2.4.

Proof of Theorem 2.4. We apply Lemma 3.2 to get rid of the MOD_{p^k} gates and get a $(\sum_p, \mathrm{AND}_t, \mathrm{MOD}_m)$ circuit for our symmetric function. The size of the circuit blows up polynomially, i.e., it is bounded by S^b, where b and t depend on p, m, k, and ℓ. Then Theorem 3.7 bounds S. Notice that we did not use the feature of Theorem 3.7 that the middle gates can directly depend on many input variables.

The statement on the specific functions follows from the observation that every function mentioned there satisfies that it is not of the form $\mathrm{MOD}^A_{mp^j}$ unless $mp^j > n$. □

The following lemma nicely complements Theorem 3.7.

LEMMA 3.8. *Consider the Boolean function* $f(x_1, x_2, \ldots, x_n) = \mathrm{MOD}^A_{mp^k}(x_1, x_2, \ldots, x_n)$. *If* m *is not a power of the prime* p, *then* f *can be computed by a* $(\sum_p, \mathrm{AND}_2, \mathrm{MOD}_m)$ *circuit of size at most* $(mn)^{2p^{k'}}$, *where* $p^{k'}$ *is the largest power of* p *dividing* mp^k.

Notice that the assumption that m is not a power of p is necessary. Otherwise, if $m = p^\ell$, arbitrary size constant depth circuits of constant fan-in AND and arbitrary MOD_p and MOD_m gates could only compute Boolean functions expressible as constant degree modulo p polynomials, and that constant degree does not depend on k. Consequently, it cannot compute f, which is a degree-$(p^k - 1)$ polynomial.

Proof. Suppose first that all elements of the 1-set A are congruent to a single number a modulo m. There is a degree $p^{k'} - 1$ polynomial on the input computing $\mathrm{MOD}^A_{p^{k'}}$ modulo p (Lemma 3.1). This polynomial can be implemented by a modulo p sum of AND gates of at most $p^{k'} - 1$ variables. The number of AND gates is bounded by $n^{p^{k'}-1}$. Let q be prime factor of m different from p, and place a redundant $\mathrm{MOD}^{\{1\}}_q$ gate above each AND gate. Apply the degree decreasing lemma (Lemma 3.9) to replace each AND gate by a collection of at most $(2q)^{p^{k'}-2}$ MOD_q gates summing to the same value modulo p. First replace each MOD_q gate by a MOD_m gate computing the same function, then replace each MOD_m gate G by the AND of G and the $\mathrm{MOD}^{\{a\}}_m$ gate on all the inputs. The resulting circuit computes the AND of the $\mathrm{MOD}^{\{a\}}_m$ and the $\mathrm{MOD}^A_{p^{k'}}$ functions; thus it computes the $\mathrm{MOD}^A_{mp^k}$ function as desired.

To remove our assumption on A, notice that every set A can be decomposed into m sets A_i satisfying this assumption. The equation $\mathrm{MOD}^A_{mp^k} = \sum_i \mathrm{MOD}^{A_i}_{mp^k}$ proves the lemma. □

Consider the smallest $(\sum_p, \mathrm{AND}_t, \mathrm{MOD}_m)$ circuit computing the function $\mathrm{MOD}^{\{0\}}_{mp^j}$, and notice that the lower bound on the circuit size for this function in Theorem 3.7 is c^{p^j}, while the upper bound in Lemma 3.8 is $n^{c'p^j}$. The gap is too wide to characterize polynomial size circuits, but we can characterize quasi-polynomial-size circuits as in Theorem 2.5.

Proof of Theorem 2.5. Apply Lemma 3.2 as in Theorem 2.4 to eliminate the MOD_{p^k} gates. Use Theorem 3.7 and Lemma 3.8 to get the two sides of the characterization. □

3.3. The degree-decreasing lemma. Lemma 3.9 exploits a surprising property of (MOD_s, MOD_m)-circuits, which ($\mathrm{MOD}_p, \mathrm{MOD}_p$) circuits lack, since constant-depth circuits with MOD_p gates and arbitrary size are only capable of computing constant-degree modulo p polynomials of the input. Here we generalize the original version [7] of the degree-decreasing lemma for nonprime moduli.

LEMMA 3.9 (degree-decreasing lemma). *Let p be a prime and $s, m > 1$ be integers, satisfying* $\gcd(s,p) = \gcd(s,m) = 1$. *Let x_1, x_2, x_3 be variables taking values from $\{0, 1, \ldots, p-1\}$, $x_1' \in \{0,1\}$. Then*

$$\mathrm{MOD}_p^A(x_1x_2 + x_3) \equiv H_0 + H_1 + \cdots + H_{p-1} + \beta \pmod{s}, \tag{3.1}$$

$$\mathrm{MOD}_m^A(x_1'x_2 + x_3) \equiv H_0' + H_1' + \beta' \pmod{s}, \tag{3.2}$$

where H_i abbreviates

$$H_i = \alpha \sum_{j=0}^{p-1} \mathrm{MOD}_p^A(ix_2 + x_3 + j(x_1 + (p-i)))$$

for $i = 0, 1, \ldots, p-1$; α is the multiplicative inverse of p modulo s: $\alpha p \equiv 1 \pmod{s}$; β is a positive integer satisfying $\beta = -|A|(p-1)\alpha \bmod s$; H_i' abbreviates

$$H_i' = \alpha' \sum_{j=0}^{m-1} \mathrm{MOD}_m^A(ix_2 + x_3 + j(x_1' + (m-i)))$$

for $i = 0, 1$; and α' is the multiplicative inverse of m modulo s: $\alpha' m \equiv 1 \pmod{s}$; and β' is a positive integer satisfying $\beta' = -|A|\alpha \bmod s$.

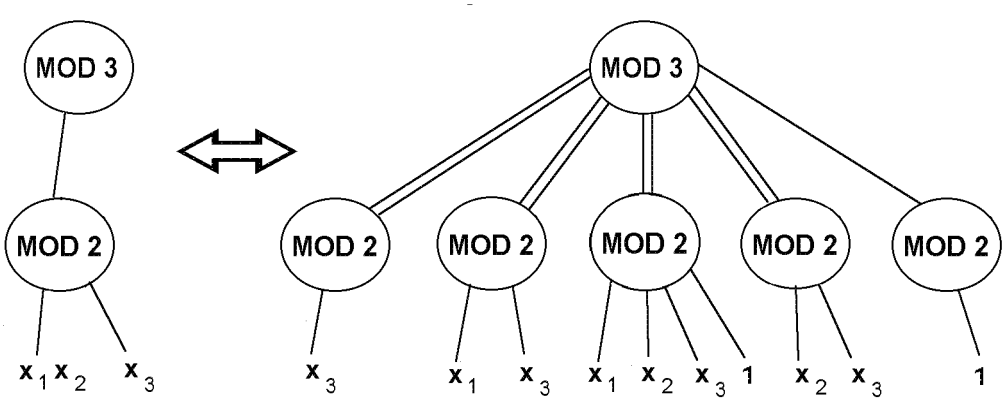

FIG. 3.2. *Degree decreasing in* ($\mathrm{MOD}_3, \mathrm{MOD}_2^{\{1\}}$) *case. On the left the input is a degree-2 polynomial, and on the right the input consists of linear polynomials.*

Proof. Let $x_1 = k$, and let $0 \le i \le p-1,\ k \ne i$. Then

$$H_k = \alpha \sum_{j=0}^{p-1} \mathrm{MOD}_p^A(kx_2 + x_3) = \alpha p \mathrm{MOD}_p^A(kx_2 + x_3) \equiv \mathrm{MOD}_p^A(x_1x_2 + x_3) \pmod{s}$$

and

$$H_i = \alpha \sum_{j=0}^{p-1} \mathrm{MOD}_p^A(ix_2 + x_3 + j(k-i)) = \alpha|A|,$$

since for any fixed x_2, x_3, i, k expression $kx_2+x_3+j(k-i)$ takes on every value exactly once modulo p while $j = 0, 1, \ldots, p-1$; so $\mathrm{MOD}_p^A(ix_2+x_3+j(k-i))$ equals 1 exactly $|A|$ times. Consequently,

$$H_0 + H_1 + \cdots + H_{p-1} + \beta \equiv \mathrm{MOD}_p^A(x_1x_2+x_3) + (p-1)\alpha|A| + \beta \equiv \mathrm{MOD}_p^A(x_1x_2+x_3) \pmod{s}.$$

Similarly, let $x_1' = k \in \{0,1\}$, and let $i \in \{0,1\}$, $k \neq i$. Then

$$H_k' = \alpha' \sum_{j=0}^{m-1} \mathrm{MOD}_m^A(kx_2+x_3) = \alpha' m \mathrm{MOD}_m^A(kx_2+x_3) \equiv \mathrm{MOD}_p^A(x_1'x_2+x_3) \pmod{s}$$

and

$$H_i' = \alpha' \sum_{j=0}^{m-1} \mathrm{MOD}_m^A(ix_2 + x_3 + j(k-i)) = \alpha'|A|,$$

since for any fixed x_2, x_3, i, k, for $i \neq k$ $|i-k| = 1$, so expression $kx_2 + x_3 + j(k-i)$ takes on every value exactly once modulo m while $j = 0, 1, \ldots, m-1$; so $\mathrm{MOD}_m^A(ix_2 + x_3 + j(k-i))$ equals 1 exactly $|A|$ times. Consequently,

$$H_0' + H_1' + \beta' \equiv \mathrm{MOD}_m^A(x_1'x_2+x_3) + \alpha'|A| + \beta' \equiv \mathrm{MOD}_p^A(x_1'x_2+x_3) \pmod{s}. \qquad \square$$

3.4. Random restriction. The CDH of [2] states that any $(\sum_p, \mathrm{MOD}_m, \mathrm{AND}_d)$ circuit computing AND has superpolynomial size if p is a prime and m and d are constants. We make progress toward this statement by proving Theorem 2.6 stating that AND requires superpolynomial size circuits of this type, if each MOD_m gate has fan-in $o(n^2/\log n)$. A stronger form of this statement (see Theorem 3.13) can be based on the following definition.

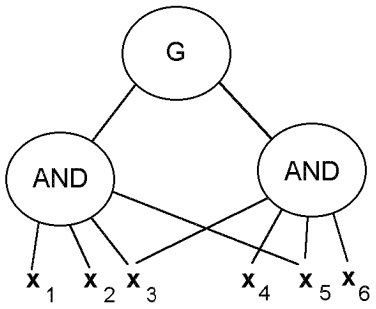

FIG. 3.3. *Gate G relates, e.g., x_1 and x_2 or x_3 and x_6 but does not relate x_1 and x_4.*

DEFINITION 3.10. *Let G be a gate of a circuit on the second level from the inputs computing some function of* AND*'s of variables. We say that G relates two input variables if they appear as inputs in a common* AND *gate below G.*

*We say that a gate G is H-*linear *if H is a subset of the input-variables, such that G does not relate two input variables outside H; i.e., the input of G is linear in the variables outside H with coefficients that are arbitrary functions of the variables in H. We call a gate ε-*linear *if it is H-linear with a set H containing at most an ε-fraction of all variables.*

We start with a simple application of the degree decreasing lemma (Lemma 3.9).

LEMMA 3.11. *Let p and m be relatively prime integers, and consider an n variable Boolean function f computed by a $(\mathrm{MOD}_m^B, \mathrm{AND})$ circuit, where the top MOD_m^B gate is H-linear. Then f can be computed by a $(\sum_p, \mathrm{MOD}_m)$ circuit with $(2m)^{|H|}$ MOD_m gates.*

Proof. We use induction on $|H|$. In the $|H| = 0$ case the AND gates have fan-in 1; thus they can be removed.

We can translate the AND gates to multiplications on the 0-1 variables. Consequently, the input of the MOD_m gate is a polynomial P of the input variables with all of its monomials having at most a single variable outside H. We may suppose that P is multilinear. If $x_i \in H$ for some $1 \le i \le n$, we can write this input in the form $P = Qx_i + R$, where the polynomials Q and R do not depend on x_i, and all their monomials contain at most a single variable outside H. We apply Lemma 3.9 to replace our MOD_m gate with the modulo p sum of $2m$ MOD_m gates. The inputs of these MOD_m gates are linear combinations of x_i, Q, and R. To finish the proof, we apply the inductive hypothesis with $H \setminus \{x_i\}$ to replace each of these new MOD_m gates with the modulo p sum of $(2m)^{|H|-1}$ MOD_m gates on the input variables. ∎

LEMMA 3.12. *Let the prime p and the positive integer m be fixed. Then there exist constants $c > 1$ and $\varepsilon > 0$ such that if a circuit $((\mathrm{MOD}_{p^k}^A)^\ell, \mathrm{MOD}_m^B, \mathrm{AND})$ computes AND_n, and every MOD_m gate is an ε-linear gate, then the size of the circuit is $S > c^n$.*

The proof of this lemma is simpler for the case when p is not dividing m. We need Theorem 3.7 in its full generality for the remaining case.

Proof. Suppose first that p does not divide m.

We apply Lemma 3.11 for the MOD_m gates. The resulting circuit computes a modulo p polynomial of degree less than $p^{k\ell}$ of the at most $S(2m)^{\varepsilon n}$ MOD_m gates (Lemma 3.2). The size is therefore at most $(S(2m)^{\varepsilon n})^{p^{k\ell}}$. But Theorem 2.4 claims an exponential lower bound on this size, thus for a small enough ε, size S must be exponential in n.

In the general case where p may divide m, we write $m = p^a m_0$ where p does not divide m_0. First we decompose each MOD_m^B gate into the sum of $\mathrm{MOD}_m^{\{b_i\}}$ gates for $B = \{b_1, b_2, \ldots, b_t\}$. Then $\mathrm{MOD}_m^{\{b_i\}}$ gates are converted to $\mathrm{MOD}_m^{\{0\}}$ gates, connecting bit 1 with multiple wires to the gate. Next, we exchange $\mathrm{MOD}_m^{\{0\}}$ gates to AND of the $\mathrm{MOD}_{m_0}^{\{0\}}$ and $\mathrm{MOD}_{p^a}^{\{0\}}$ gates. (We used a similar decomposition in the proof of Lemma 3.8.) We have increased the size of the circuit by a factor of at most $2m$ so far. We apply Lemma 3.11 to the MOD_{m_0} gates. This increases the size by a factor of at most $(2m)^{\varepsilon n}$. The resulting circuit has MOD_{m_0} and AND gates at the bottom level and MOD_{p^a}, MOD_p, $\sum_p$, and AND_2 gates everywhere else. As the last two types can be replaced with MOD_p gates, we can apply Lemma 3.2. We get a three level circuit computing AND_n with a $\sum_p$ gate on top and AND_t gates in the middle (with

a constant t depending on m, p, k, and ℓ). The bottom gates are MOD_m and AND gates. Notice that the number S_2 of the gates in the middle level is at most S_1^t, where S_1 is the number of gates on the bottom level, and $S_1 \le (2m)^{\varepsilon n+1} S$.

The fan-in of these bottom AND gates is bounded by $\varepsilon n+1$. We choose $\varepsilon < 1/4t$. Merging the bottom AND gates with the middle AND gates, one gets that AND_n is the modulo p sum of AND functions on at most $n/2$ inputs and at most t MOD_m gates. Applying Theorem 3.7, one gets that $S_2 > c^n$ with some $c > 1$ depending on p, m, k, and ℓ. Thus $S^t > c^n/(2m)^{t(\varepsilon n+1)}$ proving an exponential lower bound on S if $c > (2m)^{\varepsilon t}$. ∎

Now we turn to prove Theorem 2.6. It is a special case of the following result proving an optimal tradeoff between size and the new measure of the maximal number of related pairs.

THEOREM 3.13. *Let p be a prime and m, k, and ℓ positive integers. Suppose that a $((\mathrm{MOD}_{p^k}^B)^\ell, \mathrm{MOD}_m^A, \mathrm{AND})$ circuit computes AND_n. If each MOD_m gate in the circuit relates at most $X \ge n$ pairs of input variables then the size of the circuit is at least $c_0^{n^2/X}$, with a constant $c_0 > 1$ depending on p, m, k, and ℓ.*

Proof. We fix the values c and ε claimed in Lemma 3.12. We take a restriction on the circuit by leaving a variable unrestricted with probability $P = \varepsilon n/(22X)$ independently for each of the variables. We assign 1 to the rest of the variables. Clearly, the restricted circuit computes the AND of the remaining variables.

With probability of at least $1/2$, the number of the remaining variables is at least $n_0 = \lfloor Pn/2 \rfloor = \lfloor \varepsilon n^2/(44X) \rfloor$.

Exactly those pairs remained related in a MOD_m gate in the restricted circuit, whose both variables remained unrestricted.

The expected number of pairs related by a single gate in the restricted circuit is at most $XP^2 = \varepsilon n_0/11$. Unfortunately, the deviation can be large, it is easy to construct n gates, relating $n-1$ pairs each, such that *any* restriction to n' variables has a gate relating $n'-1$ pairs. Thus, it is important that, when using Lemma 3.12, we need not bound the number of related pairs, only the size of a set, covering each pair.

Lemma 3.12 easily implies the Theorem if there is a restriction leaving n_0 variables unrestricted, such that every MOD_m gate is ε-linear. In other words, for each G, we need the existence of a set H (depending on G) of size at most εn_0, which contains at least one of every pair related by G in the restricted circuit.

Let us bound the probability that this is not the case for a fixed MOD_m gate G. Take a maximal matching on pairs of unrestricted variables that are related by G. The set H of the endpoints of the matching-edges satisfies that no pair is related outside H, since otherwise, adding that pair to the matching would yield a larger matching. Thus it suffices to bound the probability that $|H| \ge \varepsilon n_0$, i.e., that all the variables involved in some $j = \lceil \varepsilon n_0/2 \rceil$ pairs of G-related variables forming a matching remain unrestricted. We bound this probability by the product of the number of choices for the matching and the probability that the variables remain unrestricted for a fixed matching. For a fixed gate G we get that the probability that at least n_0 variables remain unrestricted but G is not ε-linear after the restriction is at most $\binom{X}{j} P^{2j}$. Hence if $S\binom{X}{j} P^{2j} < 1/2$, where S is the size of our circuit, then Lemma 3.12 proves our theorem. The alternative is $S \ge (2\binom{X}{j} P^{2j})^{-1} \ge \left(\frac{j}{eXP^2}\right)^j$, which proves the same bound. ∎

Next we show that the logarithmic order of magnitude of the bound in Theorem 3.13 is tight.

THEOREM 3.14. *If m is not a power of the prime p, and $X > 0$ is arbitrary, then the n variable* AND *function is computable by a $(\sum_p, \mathrm{MOD}_m, \mathrm{AND})$ circuit of size $(2m)^{n^2/(2X)}$ such that the total number of pairs of variables related by any MOD_m gate in the circuit is at most X.*

Proof. Compute AND of the variables in two levels with AND gates, first computing the AND of $\lceil n^2/(2X) \rceil$ classes of at most $\lceil 2X/n \rceil$ variables each. Then place a $\mathrm{MOD}_m^{\{1\}}$ gate of fan-in 1 onto the top. Apply Lemma 3.11 to replace the top two levels by the modulo p sum of $(2m)^{n^2/(2X)}$ MOD_m gates. The inputs of these new gates are linear combinations of the outputs of the gates computing AND for a single class.

Note that Lemma 3.11 works only if m is not a multiple of p. Otherwise use that MOD_m gates can simulate MOD_q gates if q divides m. ∎

We remark that the proofs of Lemma 3.12 and Theorem 3.13 use Theorem 3.7 for the lower bound, so they apply to circuits computing OR or MOD_r with r not dividing mp^s, not just for AND. The upper bound in Theorem 3.14 can also be applied to these functions.

Acknowledgment. We are grateful to Zoltán Király for helpful discussions.

REFERENCES

[1] M. AJTAI, $\sum_1^1$ *formulae on finite structures*, Ann. Pure Appl. Logic, 24 (1983), pp. 1–48.

[2] D. A. M. BARRINGTON, H. STRAUBING, AND D. THÉRIEN, *Non-uniform automata over groups*, Inform. and Comput., 89 (1990), pp. 109–132.

[3] R. BEIGEL AND J. TARUI, *On ACC*, in Proceedings of the 32nd Annual IEEE Symposium on Foundations of Computer Science, IEEE Computer Society, Los Alamitos, CA, 1991, pp. 783–792.

[4] R. B. BOPPANA, AND M. SIPSER, *The complexity of finite functions*, Handbook of Theoretical Computer Science Vol. A: Algorithms and Complexity, J. van Leeuwen, ed., Elsevier, Amsterdam, MIT, Cambridge, MA, 1990.

[5] H. CAUSSINUS, *A note on a theorem of Barrington, Straubing and Thérien*, Inform. Process. Lett., 58 (1996), pp. 31–33.

[6] M. L. FURST, J. B. SAXE, AND M. SIPSER, *Parity, circuits and the polynomial time hierarchy*, Math. Systems Theory, 17 (1984), pp. 13–27.

[7] V. GROLMUSZ, *A degree-decreasing lemma for (mod q,mod p) circuits*, in Proceedings ICALP'98, Aalborg, Denmark, Lecture Notes in Comput. Sci. 1443, Springer-Verlag, New York, 1998, pp. 215–222.

[8] J. HÅSTAD, *Almost optimal lower bounds for small depth circuits*, in Proceedings of the 18th Annual ACM Symposium on the Theory of Computing, Berkley, CA, 1986, pp. 6–20.

[9] J. KAHN AND R. MESHULAM, *On mod p transversals*, Combinatorica, 10 (1991), pp. 17–22.

[10] M. KRAUSE AND P. PUDLÁK, *On the computational power of depth* 2 *circuits with threshold and modulo gates*, in Proceedings of the 26th Annual ACM Symposium on the Theory of Computing, Montreal, Canada, 1994, pp. 48–57.

[11] M. KRAUSE AND S. WAACK, *Variation ranks of communication matrices and lower bounds for depth-two circuits having nearly symmetric gates with unbounded fan-in*, Math. Systems Theory, 28 (1995), pp. 553–564.

[12] A. RAZBOROV, *Lower bounds for the monotone complexity of some Boolean functions*, Sov. Math. Dokl., 31 (1985), pp. 354–357.

[13] A. RAZBOROV, *Lower bounds on the size of bounded depth networks over a complete basis with logical addition*, Mat. Zametki, 41 (1987), pp. 598–607 (in Russian).

[14] R. SMOLENSKY, *Algebraic methods in the theory of lower bounds for Boolean circuit complexity*, in Proceedings of the 19th Annual ACM Symposium on the Theory of Computing, New

York, NY, 1987, pp. 77–82.
[15] P. Yan and I. Parberry, *Exponential size lower bounds for some depth three circuits*, Inform. and Comput., 112 (1994), pp. 117–130.
[16] A. C. Yao, *Separating the polynomial-time hierarchy by oracles*, in Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science, IEEE Computer Society, Los Alamitos, CA, 1985, pp. 1–10.