On the Complexity of Powering in Finite Fields

Swastik Kopparty*
Institute for Advanced Study
swastik@ias.edu

ABSTRACT

We study the complexity of computing the k^{th}-power of an element of $\mathbb{F}_{2^n}$ by constant depth arithmetic circuits over $\mathbb{F}_2$ (also known as $\text{AC}^0(\oplus)$). Our study encompasses the complexity of basic arithmetic operations such as computing cube-root and computing cubic-residuosity of elements of $\mathbb{F}_{2^n}$. Our main result is that these problems require exponential size circuits.

We also derive strong average-case versions of these results. For example, we show that no subexponential-size, constant-depth, arithmetic circuit over $\mathbb{F}_2$ can correctly compute the cubic residue symbol for more than $1/3+o(1)$ fraction of the elements of $\mathbb{F}_{2^n}$. As a corollary, we deduce a character sum bound showing that the cubic residue character over $\mathbb{F}_{2^n}$ is uncorrelated with all degree-d n-variate $\mathbb{F}_2$ polynomials (viewed as functions over $\mathbb{F}_{2^n}$ in a natural way), provided $d \ll n^{\epsilon}$ for some universal $\epsilon > 0$. Classical methods (based on van der Corput differencing and the Weil bounds) show this only for $d \ll \log(n)$.

Our proof revisits the classical Razborov-Smolensky method for circuit lower bounds, and executes an analogue of it in the land of univariate polynomials over $\mathbb{F}_{2^n}$. The tools we use come from both $\mathbb{F}_{2^n}$ and $\mathbb{F}_2^n$. In recent years, this interplay between $\mathbb{F}_{2^n}$ and $\mathbb{F}_2^n$ has played an important role in many results in pseudorandomness, property testing and coding theory.

Categories and Subject Descriptors

F.2.1 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity—*Numerical Algorithms and Problems, Computations in finite fields*

General Terms

Theory

*The author is supported by National Science Foundation Grant No. DMS-0835373.

STOC'11, June 6–8, 2011, San Jose, California, USA.

Keywords

Finite fields, Circuit Complexity, Pseudorandomness, Razborov-Smolensky method

1. INTRODUCTION

In this paper, we study the complexity of powering over $\mathbb{F}_{2^n}$. In the powering by k problem, where k is an integer between 0 and $2^n - 1$, we are given as input $x \in \mathbb{F}_{2^n}$, and we want to compute x^k. Powering contains as special cases several basic arithmetic operations. Powering x by $2^n - 2$ is the same as computing the inverse of x. For n odd (so that every element of $\mathbb{F}_{2^n}$ is a perfect cube), powering x by $(2 \cdot 2^n - 1)/3$ is the same as computing the cube-root of x. Similarly, for n even (so that not every element of $\mathbb{F}_{2^n}$ is a perfect cube), powering x by $(2^n - 1)/3$ is the problem of computing the cubic residue symbol of x (which tells us whether x is a cubic residue or not).

The powering problem is well known to be solvable in polynomial time by the repeated-squaring method. The focus of this paper is the parallel complexity of the problem, or equivalently, the complexity of this problem for low-depth circuits. Specifically, the model of computation we consider is the class of arithmetic circuits over $\mathbb{F}_2$ with gates of unbounded fan-in. Here the wires carry elements of $\mathbb{F}_2$, and we are allowed addition and multiplication gates with unbounded fan-in. Equivalently, it is the class $\text{AC}^0(\oplus)$, which consists of polynomial-size, bounded-depth circuits with unbounded fan-in AND, OR, NOT and PARITY gates. Throughout we will work with non-uniform circuit families.

Elements of $\mathbb{F}_{2^n}$ are represented in some pre-decided basis for $\mathbb{F}_{2^n}$ over $\mathbb{F}_2$ (for example, one could use the basis $\{1, y, \ldots, y^{n-1}\}$, working mod an irreducible polynomial $e(y)$ of degree n). In this setting, $\text{AC}^0(\oplus)$ is very well suited for computation over $\mathbb{F}_{2^n}$. Addition, multiplication and even iterated addition of elements of $\mathbb{F}_{2^n}$ can be trivially done by $\text{AC}^0(\oplus)$ circuits. Although there is an arbitrary choice of basis involved in representing elements of $\mathbb{F}_{2^n}$, $\text{AC}^0(\oplus)$ can switch representations from one basis to another, and this makes the question "Can C(X) be computed by polynomial size $\text{AC}^0(\oplus)$ circuits?" have an answer that is invariant under the choice of representation.

Our main result is a lower bound, which states that $\text{AC}^0(\oplus)$ circuits of depth d for powering over $\mathbb{F}_{2^n}$ must have size at least $\exp(n^{\Omega(1/d)})$. This lower bound also holds for the problems of computing q^{th}-root and q^{th}-residuosity (but we do not know if it also holds for the problem of computing the inverse). These latter lower bounds admit a self-amplification, and this enables us to deduce strong average-

case lower bounds, as well as bounds on some new kinds of character sums.

Related Work.

There have been two lines of work that considered problems of this kind.

Fich and Tompa [7], Eberly [6], and von Zur Gathen [21] study the problem of whether small, low-depth arithmetic circuits over $\mathbb{F}_p$ can power elements of $\mathbb{F}_{p^n}$. The main result here is that for constant p, powering has arithmetic circuits of bounded fan-in with size $\text{poly}(n)$ and depth $O(\log(n))$. For circuits with bounded fan-in, this depth is clearly optimal.

Our main result implies that even for arithmetic circuits over $\mathbb{F}_2$ with unbounded fan-in, there are no polynomial size circuits for powering over $\mathbb{F}_{2^n}$ of depth smaller than $\frac{\log(n)}{\log\log n}$.

Another line of work has focussed on finding the lowest uniform Boolean circuit complexity class in which one can perform basic operations over $\mathbb{F}_{2^n}$ (and other natural algebraic structures).

Most relevant for us is the work of Healy and Viola [10], who considered various questions about implementing arithmetic over $\mathbb{F}_{2^n}$ in $\text{AC}^0(\oplus)$. Although the main results there were on uniform upper bounds, [10] also proved, via reduction from Majority, that some problems closely related to powering are hard for $\text{AC}^0(\oplus)$. They showed that there are no $\text{AC}^0(\oplus)$ circuits of polynomial size that compute iterated multiplication over $\mathbb{F}_{2^n}$. The more general problem of exponentiation, where we are given as input both $x \in \mathbb{F}_{2^n}$ and $k \in \mathbb{Z}$, and we want to compute x^k was also shown to be hard for $\text{AC}^0(\oplus)$ by [10]. On the other hand, results of [10, 11] imply that both these problems are in TC^0.

One of the motivations behind this work was a suggestion, from [3], that the algebraic world of $\mathbb{F}_{2^n}$ is a good place to find functions that are hard on average for $\text{AC}^0(\oplus)$ circuits. In [3], it was shown that therein one can find functions which are hard on average for the (weaker) class of $\mathbb{F}_2$-polynomials of degree $\ll \log(n)$. Our results validate this suggestion to some extent, but it seems likely that there is much more to be said about this.

Residuosity.

It is of much interest to understand what kind of pseudorandom objects can be constructed in low complexity classes. In this context, Gutfreund and Viola [9] showed that certain arithmetic constructions over $\mathbb{F}_{2^n}$, which could be implemented in $\text{AC}^0(\oplus)$, were powerful enough to express some interesting pseudorandom phenomena.

One very powerful source of pseudorandom constructions comes from characters over finite fields. Multiplicative characters of finite fields display highly pseudorandom properties, and yet are tractable for analysis (through tools like the Weil-bound for character sums). These pseudorandom properties of characters have been used in the construction of some remarkable pseudorandom graphs (the Paley graphs, see [8, 15]), as well as other combinatorial objects like ϵ-biased sets [1, 2]. It is therefore natural to ask whether pseudorandom constructions based on multiplicative characters can be executed in $\text{AC}^0(\oplus)$.

Shparlinski [18] showed, via a reduction from Majority, that the quadratic-residue character over $\mathbb{F}_p$, for prime p, cannot be computed in $\text{AC}^0(\oplus)$. However, it remained open whether $\text{AC}^0(\oplus)$ could compute multiplicative characters in fields of small characteristic.

Our results imply that the cubic-residue (and higher-residue) multiplicative characters over $\mathbb{F}_{2^n}$ cannot be computed in $\text{AC}^0(\oplus)$. In particular, the adjacency relation of the Paley graphs over $\mathbb{F}_{2^n}$ cannot be computed in $\text{AC}^0(\oplus)$.

In the process, we show yet another pseudorandom property of the cubic (and higher) residue characters: they do not correlate with low-degree polynomials over $\mathbb{F}_2$. Specifically, let $\eta : \mathbb{F}_{2^n} \to \mathbb{F}_2^n$ be an $\mathbb{F}_2$-linear isomorphism. We show that there is an $\epsilon > 0$ such that for every polynomial $p : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most n^ϵ,

$$\left|\mathbb{E}_{x\in\mathbb{F}_{2^n}}\left[\chi(x)(-1)^{p(\eta(x))}\right]\right| < \frac{1}{n^\epsilon}.$$

Previously, such correlation bounds for the cubic character χ were only known for polynomials p of degree less than $\log n$. Curiously, the proof of this improved number-theoretic character sum bound is based on complexity-theoretic considerations: representations of circuits by polynomials, the "versatility argument" (to be explained), and random self-reducibility. These tools seem to be new to this domain (for a survey of classical, elementary techniques for bounding character sums, see [13]).

1.1 Methods

The heart of our method is to translate the whole setting into the language of univariate polynomials over $\mathbb{F}_{2^n}$. At first glance, this seems untenable; operations in $\text{AC}^0(\oplus)$-circuits inherently depend on the individual bits that are being carried around on the wires. Nevertheless, some sort of translation is possible. The Razborov-Smolensky approximation method tells us that everything computed by an $\text{AC}^0(\oplus)$ circuit is almost everywhere described by a low-degree polynomial over $\mathbb{F}_2$. For circuits with n output bits (such as a circuit computing the cube-root of an element of $\mathbb{F}_{2^n}$), it guarantees that each output bit has such a polynomial. In recent years, a number of works in pseudorandomness, property testing and coding theory have looked at precisely this situation: they take an n-tuple of n-variate polynomials over $\mathbb{F}_2$, view the entire tuple as a mapping from the big field $\mathbb{F}_{2^n}$ to the big field $\mathbb{F}_{2^n}$, represent this mapping as a univariate polynomial over $\mathbb{F}_{2^n}$, and reason about it. Via this connection, we are able to work with $\text{AC}^0(\oplus)$ circuits, in a basis-independent fashion, entirely in terms of univariate polynomials over $\mathbb{F}_{2^n}$.

In this land of univariate polynomials over $\mathbb{F}_{2^n}$, we execute the second step of the Razborov-Smolensky argument. This involves showing that if $X^{1/3}$ or the cubic-residue character can be computed by $\text{AC}^0(\oplus)$, then a large number of other functions can be computed "easily", and this violates the basic counting bound that most functions cannot be computed easily.

1.2 Results

We now formally state our main result. It shows that for certain (explicit) $\alpha \in [2^n - 1]$ (these are the α relevant for computing roots and residuosity), the map sending $x \in \mathbb{F}_{2^n}$ to x^α is hard on average for $\text{AC}^0(\oplus)$ circuits.

THEOREM 1 ($\text{AC}^0(\oplus)$ COMPLEXITY OF POWERING). *Let $a, b, q \in \mathbb{Z}$ be constants with q odd, $q > 1$ and $0 < |a|, |b| < q$. Let n be such that $\alpha = (a2^n + b)/q$ is an integer.*

Define $\Lambda : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ by $\Lambda(X) = X^\alpha$.

Then for every $\mathrm{AC}^0(\oplus)$ *circuit* $C : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ *of depth* d *and size* $M \leq 2^{n^{1/5d}}$*, for sufficiently large* n*, we have:*

$$\Pr_{x \in \mathbb{F}_{2^n}} [C(x) = \Lambda(x)] \leq 1 - \epsilon_0,$$

where $\epsilon_0 > 0$ *depends only on* q *and* d*.*

Using this theorem, along with a random self-reducibility argument, we show that taking roots is very hard on average for $\mathrm{AC}^0(\oplus)$ circuits.

THEOREM 2 ($\mathrm{AC}^0(\oplus)$ COMPLEXITY OF TAKING ROOTS). *Let* $q > 1$ *be odd. Let* n *be such that* q *is relatively prime to* $2^n - 1$*. Let* α *be the multiplicative inverse of* q *in* $\mathbb{Z}_{2^n-1}$*.*

Define $\Lambda : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ *by* $\Lambda(X) = X^\alpha$*.*

Then for every $\mathrm{AC}^0(\oplus)$ *circuit* $C : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ *of depth* d *and size* $M \leq 2^{n^{1/6d}}$*, for sufficiently large* n*, we have:*

$$\Pr_{x \in \mathbb{F}_{2^n}} [C(x) = \Lambda(x)] \leq 2^{-n^\epsilon},$$

where $\epsilon > 0$ *depends only on* q *and* d*.*

Next we state a similar hardness-amplified version of the main theorem for computing the q-th residue symbol. Let $q | 2^n - 1$ (so that $\mathbb{F}_{2^n}^\times$ has a subgroup G of index q). For our purposes, a function $\Lambda : \mathbb{F}_{2^n} \to \{0,1\}^t$ is a q^{th} residue symbol if it sends elements of distinct cosets of G in $\mathbb{F}_{2^n}^\times$ to distinct images $\in \{0,1\}^t$.

THEOREM 3. ($\mathrm{AC}^0(\oplus)$ COMPLEXITY OF THE q-TH RESIDUE SYMBOL). *Let* $q > 1$ *be an odd prime. Let* n *be sufficiently large with* $q | (2^n - 1)$*.*

Let $t = \lceil \log q \rceil$*. Let* $\Lambda : \mathbb{F}_{2^n} \to \{0,1\}^t$ *be a* q^{th} *residue symbol.*

Then for every $\mathrm{AC}^0(\oplus)$ *circuit* $C : \mathbb{F}_{2^n} \to \{0,1\}^t$ *of constant depth* d *and size* $M \leq 2^{n^{1/20d}}$*, we have:*

$$\Pr_{x \in \mathbb{F}_{2^n}} [C(x) = \Lambda(x)] \leq \frac{1}{q} + O(n^{-\frac{1}{20d}}).$$

Finally, we give a new character sum bound that these results lead to.

THEOREM 4. (BOUNDS ON A MULTIPLICATIVE CHARACTER SUM). *There exists a constant* $\epsilon > 0$ *such that the following holds. Let* q *be a constant. Let* n *be sufficiently large with* $q | (2^n - 1)$*.*

Let $p(x_1, \ldots, x_n) \in \mathbb{F}_2[x_1, \ldots, x_n]$ *with* $\deg(p) \leq n^\epsilon$*. Let* $\eta : \mathbb{F}_{2^n} \to \mathbb{F}_2^n$ *be an* $\mathbb{F}_2$*-linear isomorphism. Let* $\chi : \mathbb{F}_{2^n} \to \mathbb{C}$ *be a nontrivial* q^{th}*-residue character. Then:*

$$\left| \mathbb{E}_{x \in \mathbb{F}_{2^n}} \left[\chi(x)(-1)^{p(\eta(x))} \right] \right| \leq n^{-\epsilon}.$$

Organization of this paper.

In the next section, we give an overview of our methods. In Sections 3, 4 and 5, we prove the main lower bound, Theorem 1. In Section 6, we amplify this lower bound to prove Theorem 2, Theorem 3 and Theorem 4. We conclude with some open questions.

2. OVERVIEW OF THE PROOF

A basic principle that lies behind many unconditional lower bounds today is the fact that not all functions can be computed efficiently. While this does not itself give us an explicit function that is hard to compute, it lends itself to that task quite nicely.

Informally, we call a function Λ versatile if for every function f, one can easily compute f when given oracle access to Λ. If Λ is versatile, this automatically implies that Λ cannot be computed efficiently (provided that "efficiently", "easily" etc. are all defined suitably), for otherwise we could easily compute every function f. Depending on the actual setting, implementing such an argument can require varying degrees of sophistication.

Our lower bounds for finite field operations over $\mathrm{AC}^0(\oplus)$ will also follow the versatility argument, modelled along lines of the original arguments of Razborov and Smolensky [16, 19] showing $\mathrm{AC}^0(\oplus)$ lower bounds for the modular counting (Mod_q) and Majority (Maj) functions (see also the discussion of these arguments in [17]). Before giving an outline of our proof, we quickly recall this short and elegant argument.

2.1 The Razborov-Smolensky proof of $\mathrm{Maj} \notin \mathrm{AC}^0(\oplus)$

The basis for all our understanding of the power of $\mathrm{AC}^0(\oplus)$ comes from the Razborov-Smolensky method of polynomial approximation, which says that $\mathrm{AC}^0(\oplus)$ circuits are computed by low-degree polynomials over $\mathbb{F}_2$ on a large fraction of their domain.

THEOREM 5 (RAZBOROV-SMOLENSKY [16, 19]). *Let* $C : \{0,1\}^n \to \{0,1\}^m$ *be an* $\mathrm{AC}^0(\oplus)$ *circuit of size* M *and depth* d*. Then for every* $t > 0$*, there exist polynomials* $p_1, \ldots, p_m \in \mathbb{F}_2[x_1, \ldots, x_n]$ *and a set* $S \subseteq \{0,1\}^n$ *such that*

- $\deg(p_i) \leq O(t^d)$,
- $|S| \geq \left(1 - \frac{M}{2^t}\right) 2^n$,
- *For every* $x \in S$*, we have* $(p_1(x), \ldots, p_m(x)) = C(x)$,

We begin by reviewing [20] which shows that $\mathrm{Maj} \notin \mathrm{AC}^0(\oplus)$ (and which also shows that Maj does not correlate with $\mathrm{AC}^0(\oplus)$ circuits). This will use the Razborov-Smolensky polynomial approximation in conjunction with a lemma that shows that Maj is versatile (in a certain sense).

LEMMA 6 (MAJORITY IS VERSATILE).
For every $f : \{0,1\}^n \to \{0,1\}$*, there exist polynomials* $g, h \in \mathbb{F}_2[x_1, \ldots, x_n]$ *of degree at most* $n/2$ *such that for every* $x \in \mathbb{F}_2^n$

$$f(x) = g(x) \cdot Maj(x) + h(x).$$

We include a short proof of this in the appendix.

Now suppose Maj had a small $\mathrm{AC}^0(\oplus)$ circuit. By the Razborov-Smolensky method of polynomial approximations (Theorem 5), there is a set $S \subseteq \{0,1\}^n$ of cardinality $(1 - o(1)) \cdot 2^n$ and a polynomial p of degree at most $\mathrm{polylog}(n)$ such that $\mathrm{Maj}(x) = p(x)$ for each $x \in S$. By Lemma 6, every function from S to $\mathbb{F}_2$ can be computed by a polynomial of the form $g(x) \cdot p(x) + h(x)$, which is an $\mathbb{F}_2$ polynomial of degree at most $n/2 + \mathrm{polylog}(n)$. However, there are many functions from S to $\mathbb{F}_2$, and not that many polynomials of degree at most $n/2 + \mathrm{polylog}(n)$. This is a contradiction.

This proof even shows that for any $\mathrm{AC}^0(\oplus)$ circuit C, the fraction of inputs $x \in \{0,1\}^n$ on which $C(x) = \mathrm{Maj}(x)$ is at most $\frac{1}{2} + \frac{\mathrm{polylog}(n)}{\sqrt{n}}$.

2.2 Hardness of Cubic residuosity

Following the previous argument as a model, we will show that if detecting whether an element of $\mathbb{F}_{2^n}$ (where n is even) is a cubic residue or not is easy, then this will help us compute a large number of functions easily. The core of our work is a precise version of this statement that is sufficient for the subsequent lower bound argument, and also true.

We now work towards presenting this statement. Given a function $g : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$, we will be interested in viewing it as a map from n-bit strings to n-bit strings. Fix arbitrary representations of $\mathbb{F}_{2^n}$ as $\mathbb{F}_2^n$ (in an $\mathbb{F}_2$-linear manner). View g as n-tuple of functions $(g_1, \ldots, g_n) : \mathbb{F}_2^n \to \mathbb{F}_2^n$, where each $g_i : \mathbb{F}_2^n \to \mathbb{F}_2$. We define the $\mathbb{F}_2$-degree of the function g, denoted $\deg_{\mathbb{F}_2}(g)$, to be $\max_i \deg(g_i)$ (where $\deg(g_i)$ is the degree of g_i as a polynomial in $\mathbb{F}_2[x_1, \ldots, x_n]$).

Now define $\Lambda : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ by $\Lambda(x) = x^{(2^n-1)/3}$. This is the ($\mathbb{F}_{2^n}$-valued, as opposed to $\mathbb{C}$-valued) cubic residue character. The key properties of Λ are (i) for $x \in \mathbb{F}_{2^n}^\times$, $\Lambda(x) \in \{1, \omega, \omega^2\}$, where ω is a nontrivial cube-root of unity in $\mathbb{F}_{2^n}$, (ii) $\Lambda(xy) = \Lambda(x)\Lambda(y)$, and (iii) computing Λ is $\mathrm{AC}^0(\oplus)$-equivalent to the problem of deciding whether an element is a cubic residue or not.

The following lemma now quantifies what we mean when we say that computing $\Lambda(x)$ easily allows one to compute all functions from $\mathbb{F}_{2^n}$ to $\mathbb{F}_{2^n}$ easily.

LEMMA 7 (CUBIC RESIDUOSITY IS QUITE VERSATILE). *There exists a $\mathbb{F}_{2^n}$-linear space V of functions from $\mathbb{F}_{2^n}$ to $\mathbb{F}_{2^n}$, such that for every $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$, there exist functions $g, h : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ with $\mathbb{F}_2$-degree $\leq n/2$ and a function $e : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ such that*

$$f(x) = g(x) + \Lambda(x) \cdot h(x) + e(x), \tag{1}$$

where:

- $e \in V$
- $\dim(V) \leq \frac{1}{4} \cdot 2^n$.

The 3 main differences between this Lemma and the version used in the previous subsection are (1) the space of functions that we work with are $\mathbb{F}_{2^n}$-valued, (2) correspondingly, the multiplication "$\cdot$" refers to multiplication over the big field $\mathbb{F}_{2^n}$, and (3) the presence of the helper function e which comes from a somewhat low dimensional space of helper functions V.

Given Lemma 7, we deduce the hardness of computing cubic residuosity for $\mathrm{AC}^0(\oplus)$ as follows. Suppose there was a circuit $C : \{0,1\}^n \to \{0,1\}^2$ computing cubic residuosity (in the sense that elements of different cosets of the group of cubic residues get mapped to different strings in $\{0,1\}^2$).

Using the cube-roots of unity $\{1, \omega, \omega^2\} \subseteq \mathbb{F}_{2^n}$, we can modify this circuit to produce another $\mathrm{AC}^0(\oplus)$ circuit $C' : \{0,1\}^n \to \{0,1\}^n$ such that for all $x \in \mathbb{F}_{2^n}$, $C'(x) = \Lambda(x)$ (here we identify $\{0,1\}^n$ and $\mathbb{F}_{2^n}$ in an $\mathbb{F}_2$-linear fashion). We then apply the Razborov-Smolensky polynomial approximation to $C'(x)$ and get a set $S \subseteq \{0,1\}^n$ with $|S| = (1 - o(1)) \cdot 2^n$, and a collection of polynomials $(p_1, \ldots, p_n)$ of degree at most polylog(n) such that for every $x \in S$, $(p_1, \ldots, p_n)(x) = C'(x) = \Lambda(x)$.

Now pick an arbitrary function $f : S \to \{0,1\}^n$ and invoke the lemma. We get that f can be computed as $g + \Lambda \cdot h + e$. For $x \in S$, we know that $\Lambda(x) = (p_1, \ldots, p_n)(x)$, and thus on S, f can be computed as $r + e$, where $r : \{0,1\}^n \to \{0,1\}^n$ is an n-tuple of polynomials of degree at most $n/2 +$ polylog(n) (here we used the fact that the $\mathbb{F}_{2^n}$-multiplication "$\cdot$" appearing in Equation (1) is a bilinear operation over $\mathbb{F}_2$).

Now, there are many functions from S to $\{0,1\}^n$, but the number of functions that can be written as $r+e$, where r is an n-tuple of $\mathbb{F}_2$-polynomials of degree at most $n/2+$polylog(n), and e is an element of a small dimensional space of functions, is not that large. This is the desired contradiction. Thus, computing the cubic residue character Λ cannot be done by small $\mathrm{AC}^0(\oplus)$ circuits.

We now briefly comment on how Lemma 7 is proved. Lemma 7 deals with the interaction of the function Λ, whose definition is intrinsic to the $\mathbb{F}_{2^n}$ setting, with functions of low $\mathbb{F}_2$-degree, which is a notion best understood by viewing $\mathbb{F}_{2^n}$ as $\mathbb{F}_2^n$.

Such interaction between $\mathbb{F}_{2^n}$ and $\mathbb{F}_2^n$ is facilitated by a simple characterization of the $\mathbb{F}_2$-degree of a function $g : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ in terms of the representation of g as a univariate polynomial over $\mathbb{F}_{2^n}$. Via this characterization, the proof of Lemma 7 goes by analyzing the univariate polynomial representation of an arbitrary function $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$, and explicitly determining from it the functions g, h of low $\mathbb{F}_2$-degree. The crucial step is to bound the dimension of V, which reduces to a certain additive-combinatorial problem on $\mathbb{Z}_{2^n-1}$: we need to understand the joint distribution of Hamming weights of base-2 representations of the integers x and $(x + (2^n - 1)/3) \mod (2^n - 1)$, when x is picked uniformly at random from $\{0, 1, \ldots, 2^n - 1\}$. We do this in Section 4 by studying the the Markov chain implementing random integer addition in base 2.

In recent years, the interplay between $\mathbb{F}_{2^n}$ and $\mathbb{F}_2^n$ has played an important role in many results in pseudorandomness, property testing and coding theory [5, 12, 4, 3]. Our application to circuit complexity seems to be a natural setting for this: the powering problem over $\mathbb{F}_{2^n}$ inherently deals with $\mathbb{F}_{2^n}$, while the power of $\mathrm{AC}^0(\oplus)$ is best understood through polynomials over $\mathbb{F}_2^n$.

3. THE VERSATILITY OF POWERING

In this section, we show that powering is versatile. We begin by reviewing the dictionary that translates between multivariate polynomials over $\mathbb{F}_2^n$ and univariate polynomials over $\mathbb{F}_{2^n}$.

3.1 $\mathbb{F}_{2^n}$ vs. $\mathbb{F}_2^n$

We have the basic fact that every function from $\mathbb{F}_{2^n}$ to $\mathbb{F}_{2^n}$ can be computed by a univariate polynomial over $\mathbb{F}_{2^n}$.

LEMMA 8. *Every $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ can be written uniquely as $f(X) = \sum_{i=0}^{2^n-1} a_i X^i$, where each $a_i \in \mathbb{F}_{2^n}$.*

We will now relate the notion of $\mathbb{F}_2$-degree of a function $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ to the representation of f as a univariate polynomial.

Let us recall the definition of $\mathbb{F}_2$-degree. Pick an arbitrary $\mathbb{F}_2$-linear isomorphism $\eta : \mathbb{F}_{2^n} \to \mathbb{F}_2^n$, and consider the the map $(f_1, \ldots, f_n) : \mathbb{F}_2^n \to \mathbb{F}_2^n$ given by:

$$(f_1, \ldots, f_n)(x) = \eta(f(\eta^{-1}(x))).$$

Then the $\mathbb{F}_2$-degree of f, denoted $\deg_{\mathbb{F}_2}(f)$, is defined to be $\max_{i\in[n]} \deg(f_i)$, where $\deg(f_i)$ refers to the degree of the unique multilinear $\mathbb{F}_2$-polynomial representation of $f_i : \mathbb{F}_2^n \to \mathbb{F}_2$.

Given a function $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$, consider its representation as a univariate polynomial of degree $\leq 2^n - 1$ over $\mathbb{F}_{2^n}$. Suppose $f(X) = \sum_{i=0}^{2^n-1} a_i X^i$, with $a_i \in \mathbb{F}_{2^n}$. In terms of this representation, the $\mathbb{F}_2$-degree of f has a simple expression: amongst all the i for which a_i is nonzero, let d be the largest value of the Hamming weight of the base-2 representation of i. Then the $\mathbb{F}_2$-degree of f equals d (for a proof, see [12, 14]).

Let $\mathrm{wt}(i)$ denote the Hamming weight of the integer i. Define $M_d = \{i \in \{0, 1, \ldots, 2^n - 2\} \mid \mathrm{wt}(i) \leq d\}$. Thus we have:

LEMMA 9. *Let $0 \leq d \leq n-1$. A function $g : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ has $\mathbb{F}_2$-degree at most d if and only if $g(X)$ can be written as an $\mathbb{F}_{2^n}$-linear combination of the monomials $\{X^i \mid i \in M_d\}$.*

We view M_d as a subset of $\mathbb{Z}_{2^n-1}$. We have the following observations:

- $|M_d + \alpha| = |M_d|$ (this is clear).
- $2 \cdot M_d = M_d$ (because multiplying by 2 in $\mathbb{Z}_{2^n-1}$ does not change the Hamming weight).

3.2 Versatility of powering

We are now ready to show that the map $x \to x^\alpha$ is versatile whenever α satisfies a suitable pseudorandomness condition. In the next section, we verify that the α's we are interested in satisfy this condition (a random α does).

THEOREM 10. *Let $\alpha \in \mathbb{Z}_{2^n-1}$ such that:*

$$|M_{n/2} \cup (\alpha + M_{n/2})| \geq (\frac{1}{2} + \epsilon_0)2^n, \tag{2}$$

where $M_{n/2}$ is viewed as a subset of $\mathbb{Z}_{2^n-1}$.

Let $\Lambda : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ be given by $\Lambda(x) = x^\alpha$.

Then there exists an $\mathbb{F}_{2^n}$-linear space V of functions from $\mathbb{F}_{2^n}$ to $\mathbb{F}_{2^n}$ such that for every $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$, there are functions $g, h, e : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ such that for every $x \in \mathbb{F}_{2^n}$:

$$f(x) = g(x) + \Lambda(x) \cdot h(x) + e(x),$$

where:

- $\deg_{\mathbb{F}_2}(g), \deg_{\mathbb{F}_2}(h) \leq n/2$,
- $e \in V$,
- $\dim(V) \leq (\frac{1}{2} - \epsilon_0) \cdot 2^n + 2$.

PROOF. Let $J = M_{n/2} \cup (\alpha + M_{n/2}) \subseteq \mathbb{Z}_{2^n-1}$. By assumption, $|J| \geq (\frac{1}{2} + \epsilon_0) \cdot 2^n$.

Now view J as a subset of $\{0, 1, \ldots, 2^n - 1\}$. Let

$$J^* = \{0, 2^n - 1\} \cup \{0, 1, \ldots, 2^n - 1\} \setminus J.$$

We define V to be the $\mathbb{F}_{2^n}$-span of the functions $\{X^i \mid i \in J^*\}$. Notice that V has the right dimension.

Notice that the set of all functions f that can be written in the form $g + \Lambda \cdot h + e$, where $\deg_{\mathbb{F}_2}(g), \deg_{\mathbb{F}_2}(h) \leq n/2$ and $e \in V$ is an $\mathbb{F}_{2^n}$-vector space. Thus, by Lemma 8, it suffices to show that for each $i \in \{0, 1, \ldots, 2^n - 1\}$, the function $f(X) = X^i$ is of the form $g(X) + \Lambda(X) \cdot h(X) + e(X)$.

If $i \in J^*$, then we may write the function $f(X) = X^i$ in the form $f = g + \Lambda \cdot h + e$, where $g = h = 0$ and $e(X) = X^i$ is in V. Thus the function $f(X) = X^i$ is of the required form.

Otherwise $i \in J \setminus \{0, 2^n - 1\}$. So either $i \in M_{n/2}$ or $i \in M_{n/2} + \alpha$.

If $i \in M_{n/2}$, then we may write $f(X) = g + \Lambda \cdot h + e$, where $g(X) = X^i$ has $\deg_{\mathbb{F}_2}(g) \leq n/2$, and $h = e = 0$.

To treat the final case, suppose $i \in \alpha + M_{n/2}$ with $i \notin \{0, 2^n - 1\}$ (recall that here we treated $M_{n/2}$ as a subset of $\mathbb{Z}_{2^n-1}$; thus i is of the form $\alpha + j \mod 2^n - 1$, where $j \in M_{n/2}$). In this case, the function $f(X) = X^i$ can be written as

$$f = g + \Lambda \cdot h + e,$$

where $h(X) = X^j$ has $\deg_{\mathbb{F}_2}(h) \leq n/2$, and $g = e = 0$. To see why this representation is valid, we have that for every $x \in \mathbb{F}_{2^n}$,

$$\begin{aligned} g(x) + \Lambda(x) \cdot h(x) + e(x) &= x^\alpha \cdot x^j \\ &= x^{\alpha+j} \\ &= x^{(\alpha+j) \mod (2^n-1)} \\ &= x^i \\ &= f(x), \end{aligned}$$

where we used the basic fact that if r is not divisible by $2^n - 1$, then for every $x \in \mathbb{F}_{2^n}$ we have:

$$x^r = x^{r \mod (2^n-1)},$$

and the fact that $i = \alpha + j$ is not divisible by $2^n - 1$.

Thus for every function $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$, we conclude that there are g, h, e with $f(x) = g + \Lambda \cdot h + e$, where g, h have $\mathbb{F}_2$-degree at most $n/2$ and $e \in V$. □

4. ADDING INTEGERS

Because of the the result of the previous section, we are interested in α for which $|M_{n/2} \cup (\alpha + M_{n/2})|$ is large. We now give a simple criterion on α for this to occur.

THEOREM 11. *Let t, t' be constants. Let $\sigma_0 \in \{0,1\}^t$ be a bit-string with $\sigma_0 \neq 0^t, 1^t$, and let $\sigma' \in \{0,1\}^{t'}$.*

Let $\sigma \in \{0,1\}^n$ be a bit-string of the form $\sigma'\sigma_0^\ell$ (thus $n = t' + \ell t$). Let α be the positive integer whose base 2 representation is σ. We view α as an element of $\mathbb{Z}_{2^n-1}$.

Let $M_{n/2} \subseteq \mathbb{Z}_{2^n-1}$ be the set of all x whose binary representation has Hamming weight at most $n/2$.

Then for all sufficiently large (depending on σ_0, σ') n, we have

$$|M_{n/2} \cup (\alpha + M_{n/2})| \geq (\frac{1}{2} + \epsilon_0) \cdot 2^n,$$

where $\epsilon_0 > 0$ depends only on σ_0, σ'.

PROOF. Note that $|M_{n/2}| = |\alpha + M_{n/2}| = \left(\frac{1}{2} + O(\frac{1}{n^{1/2}})\right) \cdot 2^n$.

We want to show that the union of $M_{n/2}$ with $\alpha + M_{n/2}$ is noticeably larger than $M_{n/2}$. This is equivalent to showing that the intersection of $M_{n/2}$ with $\alpha + M_{n/2}$ is noticeably smaller than $M_{n/2}$. Equivalently, we want to show that a constant fraction of the elements of $M_{n/2}$ are not elements of $\alpha + M_{n/2}$.

The proof proceeds by studying the following experiment. Pick $R \in \mathbb{Z}_{2^n-1}$ uniformly at random. Let $S = \alpha + R \in$

$\mathbb{Z}_{2^n-1}$. Let W_R and W_S be the Hamming weights of R and S respectively. We will show that

$$\Pr[R \in M_{n/2} \wedge S \in M_{n/2}] = \Pr[W_R < n/2 \wedge W_S < n/2] < \frac{1}{2} - \epsilon_0, \tag{3}$$

for some $\epsilon_0 > 0$ depending only on σ_0, which implies the theorem.

We will now try to understand the joint distribution of (W_R, W_S). It will turn out that the distribution of (W_R, W_S), suitably translated and scaled, is very close to a (correlated) bivariate Gaussian distribution. By estimating the correlation of W_R and W_S, we will be able to deduce (3).

The key observation is that addition of the random number R to the periodic number α can be described by a simple Markov chain. Let $\ell = \lfloor n/t \rfloor$. Recall that σ is of the form $\sigma'\sigma_0\sigma_0 \dots \sigma_0\sigma_0$ (where there are ℓ blocks of σ_0). Write $R = R'R_{\ell-1}R_{\ell-2}\dots R_1R_0$, where each $R_i \in \{0,1\}^t$, and $S = S'S_{\ell-1}S_{\ell-2}\dots S_1S_0$, where each $S_i \in \{0,1\}^t$.

Let us carefully understand how the S_i are generated from the R_i, following the elementary-school addition algorithm. First R_0 is added to σ_0 to get S_0 and a carry bit c_1. Then c_1, R_1 and σ_0 are added to get S_1 and a carry bit c_2, and so on. until $c_{\ell-1}, R_{\ell-1}$ and σ_0 are added to get the carry bit c'. Finally, c', R' and σ' are added to get S' and a carry bit c. However, since c is the bit that would land in the 2^n's place of the binary representation of S, and since we are working mod $2^n - 1$, this is equivalent to adding c to the integer S. Our analysis of this will mainly study $R_{\ell-1}R_{\ell-2}\dots R_0$ and $S_{\ell-1}S_{\ell-2}\dots S_0$ (before the carry bit c gets added on), and we will later describe how to account for R', S' and the carry bit c. The core of the argument is that after conditioning on the carry bits c_i, the different (R_i, S_i) become independent. In order to formalize this, we now view this process of addition as a Markov chain.

Consider a Markov chain with state-space $\{(x,b) \mid x \in \{0,1\}^t, b \in \{0,1\}\}$. In state (x,b), the transition rule is as follows: if adding b, x and σ_0 produces carry bit b', then jump to a uniformly random element of $\{(x',b') \mid x' \in \{0,1\}^t\}$. We start the Markov chain at a uniform element of $\{(x,0) \mid x \in \{0,1\}^t\}$.

Consider ℓ steps of this Markov chain (x_0,b_0), (x_1,b_1), …, $(x_{\ell-1},b_{\ell-1})$. For each $i \in \{0,1,\dots,\ell-1\}$, define y_i to equal the t-bit sum (without carry) of b_i, x_i and σ_0. By design, the distribution of $(x_{\ell-1}x_{\ell-2}\dots x_0, y_{\ell-1}y_{\ell-2}\dots y_0)$ is identical to the distribution of $(R_{\ell-1}\dots R_0, S_{\ell-1}\dots S_0)$.

Let α_0 be the integer whose base-2 representation is σ_0; by assumption $0 < \alpha_0 < 2^t - 1$.

We now look at ℓ steps of the Markov chain and expose only at the sequence of b's $b_0, b_1, \dots, b_{\ell-1}$. The process that generates b is itself a very simple 2-state Markov chain: $b_0 = 0$; if $b_i = 0$, then $b_{i+1} = 0$ with probability $\frac{2^t-\alpha_0}{2^t}$ and $b_{i+1} = 1$ with probability $\frac{\alpha_0}{2^t}$; if $b_i = 1$, then $b_{i+1} = 0$ with probability $\frac{2^t-1-\alpha_0}{2^t}$ and $b_{i+1} = 1$ with probability $\frac{\alpha_0+1}{2^t}$. Since $0 < \alpha_0 < 2^t - 1$, all these transition probabilities are positive; this tells us that the Markov chain behaves "well". Let π be the stationary distribution for this 2-state Markov chain. For $b, b' \in \{0,1\}$, let $\pi_{bb'}$ be $\pi(b)$ times the probability of jumping from state b to state b' (thus in a random walk, we expect to see about $\pi_{bb'}$ occurrences of the transition $b \to b'$).

Therefore, with probability $1 - \ell^{-\omega(1)}$, the sequence $b_0, b_1, \dots, b_{\ell-1}$ will be such that for all $b, b' \in \{0,1\}$, the number of i with $b_ib_{i+1} = bb'$ is $\pi_{bb'} \cdot \ell \pm \sqrt{\ell} \cdot \mathrm{polylog}\ell$.

We now choose $x_0, \dots, x_{\ell-1}$ given these b_i. Observe that the distribution of the x_i given the sequence $b_0, \dots, b_{\ell-1}$ depends only on b_i and b_{i+1} (in particular, the x_i are all independent given the sequence $b_0, \dots, b_{\ell-1}$. For $b, b' \in \{0,1\}$, define

$$U_{bb'} = \{x \in \{0,1\}^t \mid \text{ adding } b, x \text{ and } \sigma_0 \text{ produces carry bit } b'\}.$$

Then the distribution of x_i given the sequence $b_0, \dots, b_{\ell-1}$ is uniform over the set $U_{b_ib_{i+1}}$. Given x_i, then y_i is simply the t-bit sum of b_i, x_i and σ_0 (without the carry bit).

To summarize, given the sequence $b_0, \dots, b_{\ell-1}$, the tuple $(W_x, W_y) = (\mathrm{wt}(x_{\ell-1}\dots x_0), \mathrm{wt}(y_{\ell-1}\dots y_0))$ can be expressed as the sum of ℓ independent random variables η_i. Each η_i is distributed according to one of the four distributions:

$$\begin{aligned}\eta_{bb'} = &\text{distribution of } (\mathrm{wt}(x), \mathrm{wt}(y)),\\ &\text{where } x \text{ is uniformly picked from } U_{bb'}\\ &\text{and } y = b + x + \sigma_0 \text{ (without carry)}.\end{aligned}$$

By the earlier observation, with very high probability over the choice of b_i, the number of copies of $\eta_{bb'}$ in this sum is very close to $\pi_{bb'} \cdot \ell$.

We already know the mean of (W_x, W_y): it is simply $\ell t/2$ (because W_x, W_y are the Hamming weights of random ℓt-bit integers).

Thus, by the 2-dimensional central limit theorem, the distribution of $\frac{1}{\sqrt{\ell}}(W_x - \ell t/2, W_y - \ell t/2)$ is close to the distribution of a 2-dimensional Gaussian distribution with mean 0 and covariance matrix $\sum_{b,b' \in \{0,1\}} \pi_{bb'} \mathrm{Cov}[\eta_{bb'}]$. We know the variance of W_x and the variance of W_y: they both equal $\ell t/4$ (again because W_x, W_y are the Hamming weights of random ℓt-bit integers).

The next lemma shows that W_x and W_y are not perfectly correlated or perfectly anticorrelated (by showing that at least one of η_{00} and η_{01} does not have perfect correlation or anticorrelation between its coordinates).

LEMMA 12. *There strings $x', x'' \in \{0,1\}^t$ with the same Hamming weight, such that the t-bit $x' + \sigma_0$ (without carry) and the t-bit $x'' + \sigma_0$ (without carry) have different Hamming weights.*

PROOF. σ_0 has some bit i equal to 0. Let x' have a 1 in coordinate i and be 0 in all other coordinates.

σ_0 has some bit j equal to 1. Let x'' have a 1 in coordinate j and be 0 in all other coordinates.

Then $\mathrm{wt}(\sigma_0 + x') = \mathrm{wt}(\sigma_0) + 1$, but $\mathrm{wt}(\sigma_0 + x'') \le \mathrm{wt}(\sigma_0)$. □

Recall the elementary fact that for a 2-dimensional Gaussian (G_1, G_2) with mean 0, and with each coordinate having positive variance, but both coordinates not perfectly correlated or anticorrelated, we must have

$$\Pr[G_1 < 0 \wedge G_2 < 0] < 1/2.$$

By smoothness of the density, we even have that there exists $\epsilon > 0$ such that

$$\Pr[G_1 < \epsilon \wedge G_2 < \epsilon] < 1/2 - \epsilon.$$

We can thus conclude the same (for a slightly smaller ϵ) for the distribution of $\frac{1}{\sqrt{\ell}}(W_x - \ell t/2, W_y - \ell t/2)$ for sufficiently

larger ℓ (or equivalently, sufficiently large n):

$$\Pr[W_x < \ell t/2 + \epsilon\sqrt{\ell} \wedge W_y < \ell t/2 + \epsilon\sqrt{\ell}] < 1/2 - \epsilon. \quad (4)$$

We now address the leftover issue of dealing with R', S' and the carry bit c. The number of bits in R' and S' is at most t', which is a constant, and hence the Hamming weights of x and R differ only by at most a constant. Similarly for y and S. Equation (4) leaves enough margin to let us conclude that $W_R < n/2 + O(\sqrt{n})$ and $W_S < n/2 + O(\sqrt{n})$ still occurs with probability at most $1/2 - \epsilon/2$. For the carry bit c, observe that this can only reduces the Hamming weight of S by k if the last k bits of S are all 1's. Taking $k = \Theta(\log^2 n)$, we see that the probability that the carry bit reduces the Hamming weight of S by more that $\log^2 n$ is negligible, and thus even taking the effect of the carry bit into account we have

$$\Pr[R \in M_{n/2} \wedge S \in M_{n/2}] = \Pr[W_R < n/2 \wedge W_S < n/2] < \frac{1}{2} - \epsilon_0.$$

REMARK The previous lemma assumed that α has base-2 representation of the form $\sigma'\sigma_0^\ell$. This implies the lemma also applies when α has base-2 representation of the form $\sigma_0^\ell\sigma'$. Indeed, recall that multiplying an element of $\mathbb{Z}_{2^n-1}$ by 2 rotates its base-2 representation cyclically. Thus $M_{n/2} = 2 \cdot M_{n/2} = 2^i \cdot M_{n/2}$ for any i, and for any set $A \subseteq \mathbb{Z}_{2^n-1}$, $|A| = |2A| = |2^i \cdot A|$.

Taking an $\alpha \in \mathbb{Z}_{2^n-1}$ with base-2 representation of the form $\sigma_0^\ell\sigma'$, we have that $\alpha' = 2^{n-t'}\alpha$ has base-2 representation of the form $\sigma'\sigma_0^\ell$, and thus by the previous lemma:

$$\begin{aligned} |M_{n/2} \cup (\alpha + M_{n/2})| &= |2^{n-t'} \cdot (M_{n/2} \cup (\alpha + M_{n/2}))| \\ &= |M_{n/2} \cup (\alpha' + M_{n/2})| \\ &\geq \left(\frac{1}{2} + \epsilon_0\right) \cdot 2^n. \end{aligned}$$

5. PROOF OF THE MAIN THEOREM

In this section, we complete the proof of the Main Theorem, Theorem 1.

Summing up what we have proved in the earlier sections: whenever α looks periodic in its base-2 representation, we have $|M_{n/2} \cup (\alpha + M_{n/2})|$ large. This implies that the map $x \to x^\alpha$ is versatile (in the sense that an oracle for x^α can be used to compute every function from $\mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ in a simple form). This will be used to place a lower bound on the complexity of computing the map $x \to x^\alpha$.

PROOF OF THEOREM 1: Recall the basic fact that the binary representation of a/q is periodic non-terminating bit string of the form $\pm 0.\sigma_0\sigma_0\ldots$, where $\sigma_0 \in \{0,1\}^t$ and t equals the multiplicative order of 2 mod q. Since $q \neq 1$, we have $\sigma_0 \neq 0^t, 1^t$.

Thus the binary representation of the integer $\beta = (a \cdot 2^n + b)/q$ (reduced mod $2^n - 1$ so that it lies in the interval $\{0, 1, \ldots, 2^n - 2\}$) is of the form $\sigma_0\sigma_0\ldots\sigma_0\sigma'$, where $\sigma' \in \{0,1\}^{t'}$ for some constant t'.

By Theorem 11 and the remark after it,

$$|M_{n/2} \cup (\beta + M_{n/2})| \geq (\frac{1}{2} + \epsilon_0) \cdot 2^n.$$

Given this, we may invoke Theorem 10 and conclude that there is an $\mathbb{F}_{2^n}$-linear space of functions V, with $\dim(V) \leq (\frac{1}{2} - \epsilon_0) \cdot 2^n$, such that for every $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$, there exists g, h of $\mathbb{F}_{2^n}$-degree at most $n/2$ and $e \in V$ with:

$$f(x) = g(x) + \Lambda(x) \cdot h(x) + e(x), \quad (5)$$

for all $x \in \mathbb{F}_{2^n}$.

Now let $C : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be an $\mathrm{AC}^0(\oplus)$-circuit of depth d and size M. Then by the Razborov-Smolensky polynomial approximation, Theorem 5 (invoked with $t = \log^2(M)$), we know that there is a set $S \subseteq \mathbb{F}_{2^n}$ with $|S| \geq \left(1 - \frac{1}{M^{\omega(1)}}\right) \cdot 2^n$, and an n-tuple of polynomials $(p_1, \ldots, p_n)$ over $\mathbb{F}_2$, each of degree at most $\log^{2d}(M)$, such that for every $x \in S$, we have $C(x) = (p_1(x), \ldots, p_n(x))$.

Now let T be the set of $x \in \mathbb{F}_{2^n}$ where $C(x) = \Lambda(x)$ (again, we identified $\mathbb{F}_2^n$ with $\mathbb{F}_{2^n}$ in an arbitrary $\mathbb{F}_2$-linear way). Thus for every $x \in S \cap T$, we have $(p_1(x), \ldots, p_n(x)) = C(x) = \Lambda(x)$. Now view $(p_1, \ldots, p_n) : \mathbb{F}_2^n \to \mathbb{F}_2^n$ as a function $p : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$; we know that p has $\mathbb{F}_2$-degree at most $\log^{2d}(M)$. We just showed that $p(x) = \Lambda(x)$ for all $x \in S \cap T$.

We will now show that $S \cap T$ is small, which will then imply that T is small. Consider an arbitrary function from $S \cap T$ to $\mathbb{F}_{2^n}$. By (5), we see that this function can be represented in the form $g + \Lambda \cdot h + e$. Using the fact that $p = \Lambda$ on $S \cap T$, we get that every function from $S \cap T$ to $\mathbb{F}_{2^n}$ can be represented in the form

$$g + p \cdot h + e.$$

We will now study how many functions can be represented in this form, and thus get an upper bound on the number of functions from $S \cap T$ to $\mathbb{F}_{2^n}$. g has $\mathbb{F}_2$-degree at most $n/2$. Now by definition, $p \cdot h$ has $\mathbb{F}_2$-degree at most $n/2 + \log^{2d}(M)$. e lies in an $\mathbb{F}_{2^n}$-linear space of dimension $(\frac{1}{2} - \epsilon_0) \cdot 2^n$. The sum of the first two terms is a function of $\mathbb{F}_2$-degree at most $n/2 + \log^{2d}(M)$, and thus by Lemma 9 lies in an $\mathbb{F}_{2^n}$-linear space of dimension at most $\left(\frac{1}{2} + \frac{\log^{2d}(M)}{\sqrt{n}}\right) \cdot 2^n$. Thus the collection of all possible functions representable in form $g + p \cdot h + e$ lies in an $\mathbb{F}_{2^n}$-linear space of dimension at most $\left(1 - \epsilon_0 + \frac{\log^{2d}(M)}{\sqrt{n}}\right) \cdot 2^n$.

On the other hand, we know that the space of all functions from $S \cap T$ to $\mathbb{F}_{2^n}$ is an $\mathbb{F}_{2^n}$-linear space of dimension $|S \cap T|$.

Thus

$$|S \cap T| \leq \left(1 - \epsilon_0 + \frac{\log^{2d}(M)}{\sqrt{n}}\right) \cdot 2^n,$$

and so

$$|T| \leq \left(1 - \epsilon_0 + \frac{\log^{2d}(M)}{\sqrt{n}} + \frac{1}{M^{\omega(1)}}\right) \cdot 2^n.$$

Thus, for $M \leq 2^{n^{1/5d}}$, we have

$$\Pr_{x \in \mathbb{F}_{2^n}}[C(x) = \Lambda(x)] \leq 1 - \frac{\epsilon_0}{2}.$$

□

6. RANDOM SELF-REDUCTIONS

In this section, we observe that the problems of computing the q-th root and computing the q-th residue symbol are random self-reducible in $\mathrm{AC}^0(\oplus)$, and we thus derive improved average-case hardness for these functions.

6.1 Roots

We first consider the maps $x \to x^{1/q}$. Recall that we are working in a field $\mathbb{F}_{2^n}$ where $2^n - 1$ is relatively prime to q, so that this map is always defined.

PROOF OF THEOREM 2: Recall that we have $\Lambda(x)^q = x$ for all $x \in \mathbb{F}_{2^n}$.

Let $C : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ be an $\mathrm{AC}^0(\oplus)$ circuit of size M and depth d. Suppose $\Pr_{x \in \mathbb{F}_{2^n}}[C(x) = \Lambda(x)] \geq 2^{-n^{1/(6d)}}$. Let $S \subseteq \mathbb{F}_{2^n}$ be the set of points where C is correct:

$$S = \{x \in \mathbb{F}_{2^n} \mid C(x) = \Lambda(x)\}.$$

Thus, $|S| \geq 2^{-n^{1/(6d)}} \cdot 2^n$.

Construct the $\mathrm{AC}^0(\oplus)$ circuit C' as follows. Fix $t = 2^{2n^{1/(6d)}}$. C' will take as advice certain $b_1, \ldots, b_t \in \mathbb{F}_{2^n}$, along with $c_1, \ldots, c_t \in \mathbb{F}_{2^n}$ where $c_i = \frac{1}{\Lambda(b_i)}$.

Circuit C', given input $x \in \mathbb{F}_{2^n}$:

- For each $i \in [t]$, compute $a_i = C(b_i \cdot x)$.
- If for any $i \in [t]$, we have $(c_i \cdot a_i)^q = x$, output $c_i \cdot a_i$.

The size of C' is $t \cdot M$, and it had depth $d + 1$.

We will show that there exist $b_1, \ldots, b_t \in \mathbb{F}_{2^n}$ such that the resulting circuit C' satisfies:

$$\Pr_{x \in \mathbb{F}_{2^n}}[C'(x) = \Lambda(x)] \geq 1 - o(1).$$

Observe that C' will correctly compute the q^{th}-root of x iff there is some $i \in [t]$ for which $b_i x \in S$. For any fixed x, if $b_1, \ldots, b_t$ are picked uniformly at random from $\mathbb{F}_{2^n}$, the probability of this occurring is $1 - (1 - |S|/2^n)^t \geq 1 - o(1)$. Thus, there exists $b_1, \ldots, b_t$ such that for at least $(1 - o(1))$-fraction of the elements $x \in \mathbb{F}_{2^n}$, C' correctly computes the q^{th}-root of its input x. This is a contradiction to Theorem 1. □

6.2 Residuosity

We now consider residuosity.

PROOF OF THEOREM 3: The proof is similar to the proof of Theorem 2. The main difference is that the q^{th}-residue symbol cannot be easily checked (unlike the q^{th}-root). Instead the random self-reduction will involve a suitable plurality vote which will be correct with high probability.

Let $C : \mathbb{F}_{2^n} \to \{0,1\}^t$ be an $\mathrm{AC}^0(\oplus)$ circuit of size M and depth d. Suppose

$$\Pr_{x \in \mathbb{F}_{2^n}}[C(x) = \Lambda(x)] \geq \frac{1}{q} + n^{-1/(10d)}.$$

We begin by replacing Λ with the more algebraically pleasant function $\Lambda : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ given by $\Lambda(x) = x^{(2^n-1)/q}$ (which always outputs a q^{th}-root of unity in $\mathbb{F}_{2^n}$). It is clear that one can suitably modify C without increasing its size by more that $O(nq)$, so that C also maps $\mathbb{F}_{2^n}$ to $\mathbb{F}_{2^n}$, C always outputs a q^{th}-root of unity in $\mathbb{F}_{2^n}$, and C still predicts Λ on a $\frac{1}{q} + n^{-1/(10d)}$ fraction of $\mathbb{F}_{2^n}$.

Construct the $\mathrm{AC}^0(\oplus)$ circuit C' as follows. Fix $t = n^{1/(10d)}$. C' will take as advice certain $b_1, \ldots, b_t \in \mathbb{F}_{2^n}$, certain $e_1, \ldots, e_t \in \{1, 2, \ldots, q-1\}$ as well as $c_1, \ldots, c_t \in \mathbb{F}_{2^n}$, where $c_i = \frac{1}{\Lambda(b_i)}$.

Circuit C', given input $x \in \mathbb{F}_{2^n}$:

- Let $e_i' \in \{1, 2, \ldots, q-1\}$ equal the multiplicative inverse of $e_i \mod q$.
- For each $i \in [t]$, compute $a_i = C(b_i \cdot x^{e_i'})^{e_i}$.
- Output the most frequent value of $a_i \cdot c_i^{e_i}$, as i varies in $[t]$.

The size of C' is $t \cdot M + O(2^t)$ (the $t \cdot M$ term is for the t invocations of C, and the 2^t term is for computing the most frequent value), and it had depth $d + 1$. Furthermore, C' always outputs a q^{th}-root of unity.

Again, we will show that there exists a choice of $b_1, \ldots, b_t \in \mathbb{F}_{2^n}$ and $e_1, \ldots, e_t \in \{1, \ldots, q-1\}$ such that the resulting circuit C' satisfies:

$$\Pr_{x \in \mathbb{F}_{2^n}}[C'(x) = \Lambda(x)] \geq 1 - o(1).$$

To show this, pick $b_1, \ldots, b_t$ independently and uniformly at random from $\mathbb{F}_{2^n}$ and $e_1, \ldots, e_t$ independently and uniformly at random from $\{1, \ldots, q-1\}$. Fix any $x \in \mathbb{F}_{2^n}$. We now show that the distribution of a_i has the following properties:

$$\Pr[a_i = \Lambda(b_i)^{e_i} \cdot \Lambda(x)] \geq \frac{1}{q} + n^{-1/(10d)},$$

and for every q^{th}-root of unity $\omega \neq 1$, we have:

$$\Pr[a_i = \omega \cdot \Lambda(b_i)^{e_i} \Lambda(x)] \leq \frac{1}{q} - \frac{(q-1)n^{-1/(10d)}}{q}.$$

Indeed,

$$\begin{aligned}
\Pr[a_i = \Lambda(b_i)^{e_i} \cdot \Lambda(x)] &= \Pr_{b_i, e_i}[C(b_i x^{e_i'})^{e_i} = \Lambda(b_i^{e_i} x)] \\
&= \Pr_{b_i, e_i}[C(b_i x^{e_i'})^{e_i} = \Lambda(b_i x^{e_i'})^{e_i}] \\
&= \Pr_{y \in \mathbb{F}_{2^n}}[C(y) = \Lambda(y)] \geq \frac{1}{q} + n^{-1/(10d)} \\
&\quad \text{(since } b_i \text{ is uniformly distributed in } \mathbb{F}_{2^n}\text{)}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Pr[a_i = \omega \cdot \Lambda(b_i)^{e_i} \cdot \Lambda(x)] &= \Pr_{b_i, e_i}[C(b_i x^{e_i'})^{e_i} = \omega \cdot \Lambda(b_i^{e_i} x)] \\
&= \Pr_{b_i, e_i}[C(b_i x^{e_i'})^{e_i} = \omega \cdot \Lambda(b_i x^{e_i'})^{e_i}] \\
&= \Pr_{y \in \mathbb{F}_{2^n}, e_i}[C(y) = \omega^{e_i'} \cdot \Lambda(y)] \\
&= \frac{1}{q-1} \cdot \Pr_{y \in \mathbb{F}_{2^n}}[C(y) \neq \Lambda(y)] \\
&\leq \frac{1}{q} - \frac{(q-1)n^{-1/(10d)}}{q}.
\end{aligned}$$

Given this claim about the distribution of the a_i, it follows that the most frequent value (as i varies over $[t]$) of $a_i \cdot c_i^{e_i}$ will equal $\Lambda(x)$ with probability at least $1 - o(1)$.

Thus, there is a fixed choice of the b_i and the e_i, such that for at least $(1 - o(1))$-fraction of $x \in \mathbb{F}_{2^n}$ the circuit C' on input x computes $\Lambda(x)$. This is a contradiction to Theorem 1. □

6.3 Bounds on multiplicative character sums

We now use the average-case hardness of residuosity that we just derived to prove the multiplicative character sum bound, Theorem 4. This is an instantiation of the general principle that distinguishability implies predictability.

PROOF OF THEOREM 4: We will show the Theorem with $\epsilon = 1/40$. Suppose not; i.e., there is a polynomial $p(x_1, \ldots, x_n) \in \mathbb{F}_2[x_1, \ldots, x_n]$ of degree at most n^ϵ such that:

$$\left|\mathbb{E}_{x \in \mathbb{F}_{2^n}}\left[\chi(x)(-1)^{p(\eta(x))}\right]\right| > n^{-\epsilon}.$$

Here p can be thought of as distinguishing between the q^{th}-residue character and a uniformly random q^{th}-root of unity. We will use this "distinguisher" p to construct an $\mathrm{AC}^0(\oplus)$ circuit C which predicts the value of χ with probability noticeably larger than $1/q$, thus contradicting Theorem 3.

Let $\theta' = \mathbb{E}_{x \in \mathbb{F}_{2^n}}\left[\chi(x)(-1)^{p(\eta(x))}\right]$ (recall that we assume that $|\theta'| > n^{-\epsilon}$). Set $\theta = \frac{\overline{\theta'}}{|\theta'|}$. Therefore

$$\mathbb{E}_{x \in \mathbb{F}_{2^n}}[\theta \cdot \chi(x)(-1)^{p(\eta(x))}] = |\theta'|,$$

and so

$$\mathbb{E}_{x \in \mathbb{F}_{2^n}}[\Re(\theta \cdot \chi(x))(-1)^{p(\eta(x))}] = |\theta'| > n^{-\epsilon}.$$

We now construct the circuit C.
Randomized $\mathrm{AC}^0(\oplus)$ **circuit** $C(x, r)$
Input $x \in \mathbb{F}_{2^n}$, random bits r

- Compute $p(\eta(x))$. p is a polynomial with at most $\binom{n}{\leq n^\epsilon}$ monomials, and hence can be computed by a depth 2 $\mathrm{AC}^0(\oplus)$ circuit of size $2^{n^{2\epsilon}}$.

- Pick $I \in \{0, 1, \ldots, q-1\}$ according to the probability distribution

  $$\Pr[I = i] = \frac{1}{q}\left(1 + \Re(\theta \cdot \omega^i)(-1)^{p(\eta(x))}\right).$$

  (it suffices that these probabilities are accurate upto $\pm\exp(-n)$ additive error, which can be done by a $\mathrm{poly}(n)$-size $\mathrm{AC}^0(\oplus)$ circuit).

- Output ω^I.

Observe that C is a circuit of depth ≤ 3 and size $\leq 2^{n^{2\epsilon}}$. By construction, for each $x \in \mathbb{F}_{2^n}$, we have:

$$\Pr_r[C(x, r) = \chi(x)] = \frac{1}{q}\left(1 + \Re(\theta \cdot \chi(x))(-1)^{p(\eta(x))}\right).$$

Thus

$$\begin{aligned}\mathbb{E}_{x \in \mathbb{F}_{2^n}}\mathbb{E}_r[\mathbf{1}_{C(x,r)=\chi(x)}] &= \mathbb{E}_x[\frac{1}{q}\left(1 + \Re(\theta \cdot \chi(x))(-1)^{p(\eta(x))}\right)] \\ &= \frac{1}{q}(1 + |\theta'|) \\ &\geq \frac{1}{q}(1 + n^{-\epsilon}).\end{aligned}$$

Therefore, there is a choice of r_0 such that:

$$\Pr_{x \in \mathbb{F}_{2^n}}[C(x, r_0) = \chi(x)] = \frac{1}{q}(1 + n^{-\epsilon}).$$

This contradicts Theorem 3 (for ϵ small enough, $\epsilon = 1/40$ suffices), as desired. □

7. OPEN PROBLEMS

We now conclude with some open problems.

1. Can the inverse over $\mathbb{F}_{2^n}$ be computed in $\mathrm{AC}^0(\oplus)$? This seems like a very basic question. Computing the inverse is a special case of powering, but our techniques fall short of proving its hardness (in particular, the inverse map is not versatile in the sense of Lemma 7).

2. Is the cubic residue character Λ exponentially hard on average for $\mathrm{AC}^0(\oplus)$? Concretely, is it the case that for every $\mathrm{AC}^0(\oplus)$ circuit of polynomial size:

   $$\Pr_{x \in \mathbb{F}_{2^n}}[C(x) = \Lambda(x)] \leq \frac{1}{3} + n^{-\omega(1)}?$$

   We conjecture that it is so.

   A well known conjecture asserts the same for the Mod_3 function. Perhaps the algebraic setting of $\mathbb{F}_{2^n}$ makes more tools available to us and will be easier to handle?

   Such hardness on average would be relevant to the construction of pseudorandom generators against $\mathrm{AC}^0(\oplus)$.

8. ACKNOWLEDGEMENTS

I am very grateful to Eric Allender, Noga Alon, Paul Beame, Eli Ben-Sasson, Jean Bourgain, Shubhangi Saraf, Srikanth Srinivasan, Madhu Sudan and Emanuele Viola for useful feedback and encouragement.

9. REFERENCES

[1] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k-wise independent random variables. In IEEE, editor, *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, pages 544–553, St. Louis, MS, Oct. 1990. IEEE Computer Society Press.

[2] Y. Azar, R. Motwani, and J. Naor. Approximating probability distributions using small sample spaces. *Combinatorica*, 18(2):151–171, 1998.

[3] E. Ben-Sasson and S. Kopparty. Polynomials uncorrelated with polynomials, 2008. Unpublished Manuscript.

[4] E. Ben-Sasson and S. Kopparty. Affine dispersers from subspace polynomials. In *STOC*, pages 65–74, 2009.

[5] J. Bourgain. On the construction of affine extractors. *Geometric And Functional Analysis*, 17:33–57, 2007.

[6] W. Eberly. Very fast parallel polynomial arithmetic. *SIAM J. Comput.*, 18(5):955–976, 1989.

[7] F. E. Fich and M. Tompa. The parallel complexity of exponentiating polynomials over finite fields. *J. ACM*, 35(3):651–667, 1988.

[8] R. L. Graham and J. H. Spencer. A constructive solution to a tournament problem. *Canad. Math. Bull.*, 14:45–48, 1971.

[9] D. Gutfreund and E. Viola. Fooling parity tests with parity gates. In K. Jansen, S. Khanna, J. D. P. Rolim, and D. Ron, editors, *APPROX-RANDOM*, volume 3122 of *Lecture Notes in Computer Science*, pages 381–392. Springer, 2004.

[10] A. Healy and E. Viola. Constant-depth circuits for arithmetic in finite fields of characteristic two. In B. Durand and W. Thomas, editors, *STACS*, volume

3884 of *Lecture Notes in Computer Science*, pages 672–683. Springer, 2006.

[11] W. Hesse, E. Allender, and D. A. M. Barrington. Uniform constant-depth threshold circuits for division and iterated multiplication. *J. Comput. Syst. Sci*, 65(4):695–716, 2002.

[12] T. Kaufman and M. Sudan. Algebraic property testing: the role of invariance. In *STOC*, pages 403–412, 2008.

[13] S. V. Konyagin and I. E. Shparlinski. *Character sums with exponential functions and their applications*, volume 136 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1999.

[14] S. Kopparty. *Algebraic Methods in Randomness and Pseudorandomness*. PhD thesis, MIT, 2010.

[15] M. Naor, A. Nussboim, and E. Tromer. Efficiently constructible huge graphs that preserve first order properties of random graphs. In J. Kilian, editor, *TCC*, volume 3378 of *Lecture Notes in Computer Science*, pages 66–85. Springer, 2005.

[16] A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *MATHNASUSSR: Mathematical Notes of the Academy of Sciences of the USSR*, 41, 1987.

[17] A. A. Razborov and S. Rudich. Natural proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997.

[18] I. E. Shparlinski. *Number theoretic methods in cryptography: complexity lower bounds*, volume 17 of *Progress in computer science and applied logic*. Birkhäuser Verlag, 1999.

[19] R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *STOC*, pages 77–82, 1987.

[20] R. Smolensky. On representations by low-degree polynomials. In *FOCS*, pages 130–138, 1993.

[21] J. von zur Gathen. Efficient exponentiation in finite fields (extended abstract). In *FOCS*, pages 384–391, 1991.

APPENDIX

A. VERSATILITY OF MAJORITY

PROOF OF LEMMA 6: Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be any function.

Define $S_0 = \{x \in \mathbb{F}_2^n \mid \mathrm{Maj}(x) = 0\}$. Define $S_1 = \{x \in \mathbb{F}_2^n \mid \mathrm{Maj}(x) = 1\}$.

The crucial fact is that S_0 and S_1 are interpolating sets for polynomials of degree at most $n/2$ (namely, for every $f_0 : S_0 \to \mathbb{F}_2$, there is a unique polynomial $p_0 : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most $n/2$ such that $p_0|_{S_0} = f_0$; and similarly for every $f_1 : S_1 \to \mathbb{F}_2$ there exists a p_1).

This fact has a simple proof. We focus on S_0 (the case of S_1 being identical). For every $I \subseteq [n]$, $|I| < n/2$, let x_I be the monomial $\prod_{i \in I} x_i$, and let $v_I \in \mathbb{F}_2^n$ be the point with 1's in coordinates in to I and 0's in the remaining coordinates. Then the $2^{n-1} \times 2^{n-1}$ matrix with rows and columns indexed by subsets $I \subseteq [n]$ with $|I| < n/2$, and whose I, J entry is $x_I(v_J)$ is upper triangular with 1s on the diagonal (where the subsets are in order of nondecreasing size). Thus the functions $x_I|_{S_0}$ are linearly independent, and by dimension counting, they form a basis for all functions on S_0.

Thus there exists a $p_0 : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most $n/2$ such that $p_0|_{S_0} = f|_{S_0}$, and a p_1 such that $p_1|_{S_1} = f|_{S_1}$. Then for all $x \in \mathbb{F}_2^n$,

$$\begin{aligned} f(x) &= p_0(x) \cdot (1 - \mathrm{Maj}(x)) + p_1(x) \cdot \mathrm{Maj}(x) \\ &= (p_1 - p_0)(x) \cdot \mathrm{Maj}(x) + p_0(x). \end{aligned}$$

□