

# Candidate Weak Pseudorandom Functions in $AC^0 \circ MOD_2$

Adi Akavia\*    Andrej Bogdanov†    Siyao Guo‡    Akshay Kamath§    Alon Rosen¶

## Abstract

Pseudorandom functions (PRFs) play a fundamental role in symmetric-key cryptography. However, they are inherently complex and cannot be implemented in the class  $AC^0(MOD_2)$ . Weak pseudorandom functions (weak PRFs) do not suffer from this complexity limitation, yet they suffice for many cryptographic applications.

We study the minimal complexity requirements for constructing weak PRFs. To this end

- We conjecture that the function family  $F_A(x) = g(Ax)$ , where  $A$  is a random square  $GF(2)$  matrix and  $g$  is a carefully chosen function of constant depth, is a weak PRF. In support of our conjecture, we show that functions in this family are inapproximable by  $GF(2)$  polynomials and do not correlate with any fixed Boolean function family of subexponential size.
- We study the class  $AC^0 \circ MOD_2$  that captures the complexity of our construction. We conjecture that all functions in this class have a Fourier coefficient of magnitude  $\exp(-\text{poly log } n)$  and prove this conjecture in the case when the  $MOD_2$  function is typical.
- We investigate the relation between the hardness of learning noisy parities and the existence of weak PRFs in  $AC^0 \circ MOD_2$ .

We argue that such a complexity-driven approach can play a role in bridging the gap between the theory and practice of cryptography.

---

\*Tel Aviv Academic College, Israel. Email: [akavia@mta.ac.il](mailto:akavia@mta.ac.il). Work done while visiting IDC Herzliya, supported by the ERC under the EU's Seventh Framework Programme (FP/2007-2013) ERC Grant Agreement n. 307952.

†Department of Computer Science and Engineering and Institute for Theoretical Computer Science and Communications, Chinese University of Hong Kong. Email: [andrejb@cse.cuhk.edu.hk](mailto:andrejb@cse.cuhk.edu.hk). Partially supported by RGC GRF grants CUHK410111 and 410112. Part of this work was done while visiting IDC Herzliya, supported by the ERC under the EU's Seventh Framework Programme (FP/2007-2013) ERC Grant Agreement n. 307952.

‡Department of Computer Science and Engineering, Chinese University of Hong Kong. Email: [syguo@cse.cuhk.edu.hk](mailto:syguo@cse.cuhk.edu.hk). Part of this work was done while visiting IDC Herzliya, supported by the ERC under the EU's Seventh Framework Programme (FP/2007-2013) ERC Grant Agreement n. 307952.

§Department of Computer Science, Chennai Mathematical Institute. Email: [adkamath@cmi.ac.in](mailto:adkamath@cmi.ac.in). Part of this work was done while visiting IDC Herzliya, supported by the ERC under the EU's Seventh Framework Programme (FP/2007-2013) ERC Grant Agreement n. 307952.

¶Efi Arazi School of Computer Science, IDC Herzliya, Israel. Email: [alon.rosen@idc.ac.il](mailto:alon.rosen@idc.ac.il). Supported by the ERC under the EU's Seventh Framework Programme (FP/2007-2013) ERC Grant Agreement n. 307952.

# 1 Introduction

The design of symmetric-key cryptographic primitives can be roughly classified into two main categories: (1) the “theory-oriented” approach, in which claims on the security of one’s design are supported by a reduction from a well-established hardness assumption, and (2) the “practice-oriented” approach, in which the construction is heuristically guided by practical experience and common sense. In both cases, confidence in the security of the design is gained through lack of cryptanalysis over time. In this respect, the theory-oriented approach is preferable over its practical counterpart, as it typically relies on simpler and mathematically more natural hardness assumptions. At the same time, the practice-oriented approach, not being constrained by burdensome proofs of security, results in significantly more efficient constructions.

Even though practical constructions that withstand the test of time are widely considered to be secure enough for applications, it is not clear whether the lack of cryptanalysis should be attributed to their secure design or to the fact that the design’s complexity hinders analysis and theoretical understanding. This leaves us in an unsatisfactory state of affairs, in which the most widely used cryptographic primitives lack justification for their security. But is complex design really necessary for actual security, or is it only introduced to hinder cryptanalytic efforts in practice? Is simple design, one that is easy to understand from a theoretical perspective (and moreover lends itself to efficient implementation), incompatible with security?

The tension between the theoretical and practical approaches is perhaps most apparent in the design of *pseudorandom functions* (PRFs) [GGM84], a fundamental cryptographic primitive that yields direct solutions to most central goals of symmetric cryptography (encryption, authentication, identification). While simple theory-oriented constructions of PRFs do exist, these constructions are inefficient and as a result the most widely deployed PRF in practice is AES [DR02]. It is true that, comparatively speaking, the design of AES can be considered simple. Yet, we are still quite far from understanding the role that many of its design choices play in its security as a PRF.

Our focus is on so called *weak pseudorandom functions*. An adversary for a weak PRF aims to distinguish a random member of the family from a truly random function after observing a polynomially-bounded number of samples  $(x_1, f(x_1)), \dots, (x_m, f(x_m))$ , where  $x_1, \dots, x_m$  are independent uniformly random strings from  $\{0, 1\}^n$  and  $f: \{0, 1\}^n \rightarrow \{0, 1\}$  is the function in question. Although strong PRFs (in which the  $x_i$ ’s can be adaptively chosen by the adversary) have traditionally played a more important role in cryptography, in many applications of interest they can be replaced by weak PRFs. Cryptographic applications of weak PRFs have been studied in several works [DN02, MS07, Pie09, DKPW12, LM13].

In this paper we set out to better understand what makes a weak PRF secure from a complexity theoretic perspective, and in particular whether the efficiency of such functions is inherently tied to that of strong PRFs. We propose constructions of weak pseudorandom functions whose complexity of evaluation is minimal. To this end, we study the class  $\text{AC}^0 \circ \text{MOD}_2$  of polynomial-size, constant-depth circuit families with a layer of  $\text{MOD}_2$  gates at the bottom level followed by layers of AND/OR gates.

Unlike the typical theory-oriented construction, our design will not be accompanied by a reduction from an established problem. Instead, we let the choices in our design be guided by insights on the complexity of low-depth circuits. A useful byproduct of this process is that it identifies certain complexity theoretic properties of the class  $\text{AC}^0 \circ \text{MOD}_2$  and gives rise to interesting conjectures that are of independent interest.

In addition to minimizing the computational complexity of our candidate weak pseudorandom functions, we also make deliberate effort to keep their design as simple as possible. This is done in order to make its cryptanalysis more appealing, and to examine to what extent the functions can withstand actual attacks (this approach was also explicitly advocated by Goldreich [Gol00] and Miles and Viola [MV12]). The longer they do, the more confidence we will gain in their security.

## 1.1 Low-depth pseudorandom functions

Are parallel implementations of PRFs at all achievable? For the case of strong PRFs, a complexity-theoretic viewpoint provides a fairly satisfactory answer to this question. Based on works of Razborov and Smolensky [Raz87, Smo87], Razborov and Rudich [RR94] showed that the class  $AC^0(\text{MOD}_2)$  of polynomial-size, constant-depth circuit families with unbounded fan-in AND, OR, and  $\text{MOD}_2$  gates cannot contain strong PRFs of hardness<sup>1</sup> beyond  $\exp \text{poly} \log n$ .

Several works proposed candidate, theory-guided, constructions of strong PRFs that can be implemented in the class  $TC^0$  of polynomial-size, constant-depth circuit families with unbounded fan-in threshold gates [NR97, BPR12]. The class  $TC^0$  strictly contains  $AC^0(\text{MOD}_2)$  and is believed to be of strictly lower complexity than the original construction of Goldreich et al. [GGM84].

In the context of weak PRFs, Linial, Mansour, and Nisan [LMN93] show that weak (and hence also strong) PRF implementations in the class  $AC^0$  of polynomial-size, constant-depth circuit families with unbounded fan-in AND/OR gates can have hardness at most  $\exp \text{poly} \log n$ , which is considered inadequate for cryptographic practice. With this in mind, the class  $AC^0 \circ \text{MOD}_2$ , being a slight extension of  $AC^0$ , does indeed seem to be “minimal”. Moreover, as mentioned above, this class (being contained in  $AC^0(\text{MOD}_2)$ ) does not admit strong PRFs of adequate hardness.

Linial et al. prove that  $AC^0$  function families have most of their Fourier mass concentrated on the first  $t = \text{poly} \log n$  levels of the spectrum; thus such functions can be distinguished from random by detecting a noticeable correlation with one of the linear functions that depends on at most  $t$  variables. More generally, one could consider attacks that exploits the large correlation between the function being attacked and some other function  $h$  that belongs to some fixed family  $\mathcal{H}$  of relatively small size.

These “LMN-type” attacks can be performed given access to the function at random (and in particular not adaptively chosen) inputs. Thus, at the very least a candidate weak pseudorandom function should exhibit low correlation with members of fixed families of small size.

## 1.2 A Candidate weak pseudorandom function in $AC^0 \circ \text{MOD}_2$

For parameters  $n, k \in \mathbb{N}$ , such that  $n/k = \Theta(2^k)$ , and uniformly chosen  $A \in \{0, 1\}^{(n+1) \times (n+1)}$  our candidate weak PRF is:

$$F_A(x) = g(Ax) \quad \text{where} \quad g(y, z) = \text{TRIBES}_{n/k, k}(y) \oplus z.$$

Here,  $\text{TRIBES}_{n/k, k}$  is the function defined as the OR of  $n/k$  ANDs over independent blocks with  $k$  bits in each block. In particular, for every  $A$ ,  $F_A$  can be computed by  $AC^0 \circ \text{MOD}_2$  circuits. The rationale behind the design of our candidate function takes the following considerations into account:

---

<sup>1</sup>A pseudorandom function has hardness  $h$  if no adversary of size  $h$  can distinguish it from a random one with advantage better than  $1/h$ .

- Any function of the form  $g(Ax)$  for  $g \in \text{AC}^0$  and random  $A$  has known heavy Fourier coefficients, and hence can be learned from random samples by the algorithm of Linial et al. The matrix  $A$  randomly shuffles (and aggregates) the heavy Fourier coefficients of the TRIBES function. In fact, this transformation reduces the correlation of the function with any fixed Boolean function family of subexponential size. Keeping  $A$  secret is necessary for avoiding these types of attacks.
- The choice of the TRIBES function is driven by the need to prevent approximation of the function  $F_A$  by  $GF(2)$  polynomials of low degree, which are learnable from random examples.
- We XOR the output of TRIBES with an independent bit in order to make the output of  $F_A$  unbiased.

Needless to say, it would have been preferable if we could also provide a reduction from a well-established hard problem. However, we currently do not know of such a reduction to any candidate weak pseudorandom function in  $\text{AC}^0 \circ \text{MOD}_2$ . As will become clear later, our results on the complexity class  $\text{AC}^0 \circ \text{MOD}_2$  (combined with results of Feldman et al. [FGKP09]) imply that the computational hardness of the *Learning Parity With Noise* (LPN) problem [BFKL94] is necessary for the existence of weak PRFs in  $\text{AC}^0 \circ \text{MOD}_2$ .

We now turn to give a more detailed exposition of the rationale behind the choices made in the design of our candidate weak PRF, as well as connections to the complexity class  $\text{AC}^0 \circ \text{MOD}_2$  and to the LPN problem.

### 1.3 The power of $\text{AC}^0 \circ \text{MOD}_2$ circuits

To better understand the security of our candidate weak PRFs, we study the class  $\text{AC}^0 \circ \text{MOD}_2$  of polynomial-size, constant-depth circuit families consisting of a layer of  $\text{MOD}_2$  gates at the bottom level followed by layers of AND/OR gates. Functions computed by such circuits can be represented in the form  $f(x) = g(Ax)$ , where  $A$  is an  $m(n) \times n$   $GF(2)$  matrix and  $g$  is a constant-depth circuit of size polynomial in  $n$ .

We begin by conjecturing a structural property of all functions in this class.

**Conjecture 1.** *Let  $\{f_n: \{0,1\}^n \rightarrow \{-1,1\}\}$  be a function family in  $\text{AC}^0 \circ \text{MOD}_2$ . There is a polynomial  $p$  such that for every  $n$  there exists an  $a \in \{0,1\}^n$  satisfying  $|\hat{f}_n(a)| \geq 2^{-p(\log n)}$ .*

Here we identify the outputs “true” and “false” of the circuit with the values  $-1$  and  $1$ , respectively, and we use the notation  $\hat{f}(a)$  for the  $a$ -th Fourier coefficient of  $f$  over  $\mathbb{Z}_2^n$ . In Section 2 we prove Conjecture 1 in the cases when

1.  $A$  is a random matrix (of arbitrary dimension) and  $g$  is an arbitrary  $\text{AC}^0$  function, and
2.  $A$  is an arbitrary matrix and  $g$  is a polynomial-size depth-2 function (a CNF or a DNF).<sup>2</sup>

Our conjecture implies that functions in  $\text{AC}^0 \circ \text{MOD}_2$  can be distinguished from random ones by the Goldreich-Levin and Kushilevitz-Mansour algorithms for learning Fourier coefficients. However, these algorithms require the examples  $(x, f(x))$  to be chosen by the adversary. Similarly, the distinguisher of Razborov and Rudich that separates functions in  $\text{AC}^0(\text{MOD}_2)$  from random ones

---

<sup>2</sup>This argument also appears in independent work of Servedio and Viola [SV12].

requires chosen examples. Therefore these arguments do not rule out the possibility of weak PRFs in  $AC^0 \circ MOD_2$ .

Servedio and Viola [SV12] conjecture that the inner product function modulo 2 cannot be computed by polynomial-size  $AC^0 \circ MOD_2$  circuits. This would follow from positive answer to Conjecture 1.

#### 1.4 Some natural attacks on weak pseudorandom functions

Natural attacks for weak PRFs are typically based on *weak learning algorithms* with respect to uniform distributions. Informally, an algorithm is a weak learning algorithm with respect to uniform distribution for a function family if, for any function in the family, after observing a bounded number of random samples (i.e., input output pairs where inputs are randomly chosen from the uniform distribution) of the function, the algorithm predicts the output of a random challenge input with advantage better than  $1/2$ . Because the prediction of any algorithm for random functions is  $1/2$ , a weak learning algorithm can distinguish this function family from a random one<sup>3</sup>.

To the best of our knowledge, known attacks for weak PRFs are essentially based on following two kinds of learning algorithms: statistical query learning algorithms and low-degree learning algorithms.

**Statistical query learning algorithms.** Kearns [Kea98] observed that most known learning algorithms can be converted to *statistical query learning algorithms* which use random samples in a restricted way: such algorithms only obtain approximate estimates of  $E_x[g(x, f(x))]$  for statistics  $g$  of the algorithm's choice.

In particular, the algorithm of Linial et al. [LMN93] which rules out weak PRFs in  $AC^0$  is a statistical query learning algorithm. Linial, Mansour, and Nisan proved that  $AC^0$  function families have most of their Fourier mass concentrated on the first  $t = \text{poly log } n$  levels of the spectrum; thus such functions has a noticeable correlation with one of the linear functions that depends on at most  $t$  variables. We find that one by estimating the correlation between given function in  $AC^0$  and every linear function that that depends on at most  $t$  variables. And that one gives a prediction of given function for random inputs with advantage better than  $1/2$ . Notice that correlation of given function  $f$  with linear function  $h$ , namely  $E_x[f(x)h(x)]$ , is a statistical property of  $f$ . Thus this algorithm is a statistical query learning algorithm.

In Section 4 we show that statistical query learning algorithms cannot weakly learn the function family  $F_A(x) = g_A(x)$  where  $A$  is a random  $n \times n$   $GF(2)$  matrix and  $g$  is an unbiased function.

**Low-degree learning algorithms.** Kearns [Kea98] demonstrated that parity functions cannot be learned by statistical query learning algorithms. Thus, learning algorithms for low-degree functions are different from statistical query learning algorithms.

Parity functions, and more generally low-degree polynomials over  $GF(2)$ , can be efficiently learned by Gaussian elimination. Given a degree  $d$  polynomial  $f(x) = \sum_{S \subseteq [n], |S|=d} a_S \prod_{i \in S} x_i$ , we can learn the coefficients by reducing to linear function (or parity function) case by linearization.

---

<sup>3</sup>We remark that an algorithm which distinguishes a function family from random is not necessary a weak learning algorithm for this function family. A weak learning algorithm needs to work for any function in the family. However, an algorithm that works for most functions in the family is sufficient to break the pseudorandomness of a function family

Specifically, we transform samples of  $f$  to samples of linear function  $g(y) = \sum_{S \subseteq [n], |S|=d} a_{S,y}$  by letting  $y_S = \prod_{i \in S} x_i$  and then recover coefficient of  $g$  by Gaussian elimination. The complexity is  $\text{poly}\binom{n}{d}$ .

In this work we consider function families  $F_A(x) = g(Ax)$ , where  $A$  is a random  $n \times n$   $GF(2)$  matrix and  $g$  is an almost unbiased “rounding” function of low complexity. For which choices of  $g$  is  $F_A$  a weak PRF? Clearly if  $g$  is a linear function then  $F_A$  is not weakly pseudorandom, as the algorithm for learning linear functions from random examples can be used to distinguish  $F_A$  from a random function. More generally, if  $g$  is a degree- $d$   $GF(2)$  polynomial, then the degree- $d$  polynomial learning algorithm distinguishes  $F_A$  from a random function in time  $\text{poly}\binom{n}{d}$ . Even more generally, if  $g$  is  $\varepsilon$ -close to a degree- $d$  polynomial  $p$  for  $\varepsilon \ll 1/\binom{n}{d}$  we would expect  $F_A$  to appear indistinguishable from  $p(Ax)$  from the perspective of the learning algorithm, and so  $F_A$  can again be distinguished from a random function.

Taking into account these bad choices of the function  $g$ , we propose the following conjecture regarding the pseudorandomness of  $F_A$ . We say  $g$  is *unbiased* if  $\mathbb{E}_{x \sim \{0,1\}^n} [g(x)] = 0$ .

**Conjecture 2.** *Let  $\alpha < 1$  be a constant,  $n$  be sufficiently large, and  $g: \{0,1\}^n \rightarrow \{-1,1\}$  be an unbiased function. If  $g$  is  $1/\binom{n}{d}^\alpha$ -far from all  $GF(2)$  polynomials of degree  $d$  then  $F_A(x) = g(Ax)$  is a weak PRF of hardness  $\binom{n}{d^{\Omega(1)}}$ .*

Razborov and Smolensky showed that the majority and  $\text{MOD}_q$  predicates on  $n$  input bits, where  $q$  is a power of a prime other than 2, are  $\Omega(1)$ -far from all  $GF(2)$  polynomials of degree  $\sqrt{n}$ . Conjecture 2 predicts that the resulting weak pseudorandom functions have exponential hardness.

## 1.5 Inapproximability of $\text{AC}^0 \circ \text{MOD}_2$ by low-degree polynomials

Can we instantiate Conjecture 2 with a function  $g$  in the class  $\text{AC}^0 \circ \text{MOD}_2$ ? Razborov [Raz87] shows that every  $\text{AC}^0$  function  $g$  can be approximated by a polynomial of degree  $d$  with error  $\varepsilon = \exp(-d^{\alpha'}/\log n)$ , where  $\alpha' \leq 1$  is a constant that is inverse proportional to the circuit depth. In contrast, Conjecture 2 requires that  $g$  be  $1/\binom{n}{d}^\alpha = \exp(-O(\alpha d \log n))$ -far from all such polynomials (assuming  $d \leq \sqrt{n}$ ). Thus we would like to find functions in  $\text{AC}^0 \circ \text{MOD}_2$  whose Razborov approximating polynomials have essentially optimal degree-error tradeoff.

As an example of interest we consider the  $\text{TRIBES}_{n/k,k}$  function. This function is defined as the OR of  $n/k$  ANDs over independent blocks with  $k$  bits in each block. We will assume that the bias of  $\text{TRIBES}_{n/k,k}$  is constant, which is only possible if  $n/k = \Theta(2^k)$ . The Razborov approximation method gives approximating polynomials for  $\text{TRIBES}_{n/k,k}$  of degree  $d$  and error  $\exp(-O(d/\log n))$  for every  $d$ . We show this is tight up to  $O(\log n)$  factor in the exponent.

**Theorem 3.** *Let  $\alpha > 0$  be a constant. Assume that  $\alpha k 2^k \leq n \leq (1 - \alpha)k 2^k$ . For every  $GF(2)$  polynomial  $p$  of degree  $1 \leq d \leq n - k$ ,  $\Pr_x[p(x) \neq \text{TRIBES}_{n/k,k}(x)] \geq 2^{-3d}$ .*

Despite this, the  $\text{TRIBES}_{n/k,k}$  function may not be a suitable choice for  $g$  as we are not aware of any sequence of values  $(n, k(n))$  that makes  $g$  unbiased (or of negligible bias). We therefore work with the function  $g(y, z) = \text{TRIBES}_{n/k,k}(y) \oplus z$ , which we show is unbiased yet cannot be approximated by degree  $d$  polynomials any better than  $\text{TRIBES}$ .

**Corollary 4.** *The function  $g(y, z) = \text{TRIBES}_{n/k,k}(y) \oplus z$  is unbiased and satisfies the conclusion of Theorem 3.*

## 1.6 Learning noisy parities vs. weak pseudorandom functions in $AC^0 \circ MOD_2$

The Goldreich-Levin and Kushilevitz-Mansour algorithms efficiently learn all large Fourier coefficients of a boolean function  $f$ , but require that the examples  $(x, f(x))$  be chosen by the algorithm. Feldman et al. [FGKP09] show that if noisy parities can be learned efficiently, then random examples can be used instead of chosen ones. Thus assuming Conjecture 1, the LPN assumption is necessary for the existence of weak PRFs in  $AC^0 \circ MOD_2$ . Is it also sufficient?

Banerjee et al. [BPR12] observe that under the Learning with Errors (LWE) [Reg10] assumption over a sufficiently large modulus  $q$  one can construct relatively simple weak PRFs. Their argument is based on the existence of an efficient *rounding function*  $g$  for which the distributions  $g(Ax + e)$  and  $g(Ax)$  are of negligible statistical distance. Here,  $A$  is a random  $GF(q)$  matrix and  $e$  follows the noise model of the LWE assumption. Can this argument be modified to handle smaller values of  $q$ , specifically  $q = 2$ ?

In Section 5 we show that no such reduction exists from LPN to breaking a weak PRF of the form  $g(Ax)$  for any choice of the rounding function  $g$ : If samples of the form  $g(Ax + e)$  and  $g(Ax)$  are statistically indistinguishable, then  $g$  must be significantly biased.

In a recent work, Alwen et al. [AKPW13] use a different type of reduction to show hardness of “learning with rounding” for a bounded number of samples assuming the hardness of LWE over moduli  $q$  of magnitude polynomial in the hardness parameter.

## 2 The Fourier Spectrum of $AC^0 \circ MOD_2$

In this section we prove two special cases of Conjecture 1. This conjecture postulates the existence of a large Fourier coefficient in functions of the form  $f(x) = g(Ax)$ , where  $A$  is an  $m$  by  $n$  matrix and  $g$  can be represented by a small  $AC^0$  circuit.

Linial, Mansour, and Nisan [LMN93] showed that the Fourier spectrum of a function that can be computed by  $AC^0$  circuits of size  $s$  and depth  $d$  is concentrated on the first  $O(\log s)^d$  levels. It follows that at least one of these Fourier coefficients of low weight must have large value. If the rows of  $A$  are linearly independent (i.e. the map  $A^T$  is injective), then all the Fourier coefficients of  $g$  appear as coefficients of  $f$  and the conjecture follows easily.

The scenario where the rows of  $A$  are dependent is more interesting. In particular this always happens when  $m > n$ . In this case, every Fourier coefficient of  $f$  is a scaled sum of Fourier coefficients of  $g$  over some affine subspace. The concentration property of  $g$  proved by Linial et al. is no longer sufficient to obtain the conclusion of Conjecture 1.

We make the following partial progress towards the conjecture:

1. In Proposition 5 we prove that if  $g$  has small DNF size, then  $g(Ax)$  has a large Fourier coefficient. This proof is inspired by the analysis of Jackson’s algorithm for weakly learning DNF formulas.
2. In Proposition 7 we show that if no set of  $\text{poly } \log s$  rows of  $A$  is linearly dependent, then  $g(Ax)$  has a large Fourier coefficient. Curiously, in addition to Linial et al.’s characterization of the Fourier spectrum of  $AC^0$  circuit, our proof relies on Braverman’s theorem stating that polylogarithmically independent distributions are pseudo-random for  $AC^0$  circuits.

In particular, Proposition 7 applies to the cases when the map  $A^T$  is injective and when  $A$  is a random matrix of essentially arbitrary dimension.

## 2.1 $g$ is a DNF and $A$ is arbitrary

**Proposition 5.** *Let  $g: \{0, 1\}^m \rightarrow \{-1, 1\}$  be a DNF with  $s$  terms,  $A$  be an  $m \times n$   $GF(2)$  matrix and  $f(x) = g(Ax)$ . There exists  $a \in \{0, 1\}^n$  such that  $|\hat{f}(a)| \geq 1/(2s + 1)$ .*

The proof of Prop 5 uses the following Lemma of Jackson, which he used in the analysis of his weak learner for DNF formulas.

**Lemma 6** (Jackson [Jac97]). *Let  $g: \{0, 1\}^m \rightarrow \{-1, 1\}$  be a DNF with  $s$  terms. For every distribution  $D$  on  $\{0, 1\}^m$ , there exists a vector  $b \in \{0, 1\}^m$  so that  $\mathbb{E}_{z \sim D}[g(z)\chi_b(z)] \geq 1/(2s + 1)$ .*

*Proof of Proposition 5.* By Lemma 6, for distribution  $Ax$  where  $x \sim \{0, 1\}^n$ , there exists  $b \in \{0, 1\}^m$  so that

$$\mathbb{E}_{x \sim \{0, 1\}^n}[g(Ax)\chi_b(Ax)] \geq 1/(2s + 1).$$

Notice that  $\mathbb{E}_{x \sim \{0, 1\}^n}[g(Ax)\chi_b(Ax)] = \mathbb{E}_{x \sim \{0, 1\}^n}[f(x)\chi_{A^T b}(x)] = \hat{f}(A^T b)$ . Therefore there exists  $a = A^T b$ ,  $|\hat{f}(a)| \geq 1/(2s + 1)$ .  $\square$

## 2.2 $A$ is “typical” and $g$ is arbitrary

In this section we consider the case where  $A$  is a random matrix and  $g$  is an arbitrary  $AC^0$  function. We first prove, in Proposition 7, that Conjecture 1 holds for an arbitrary  $AC^0$  function  $g$  and  $A$  with a certain property (specifically, the property that no poly  $\log s$  rows of  $A$  are linearly dependent). Then in Corollary 10, we show that a random matrix  $A$  has this property with very high probability. Therefore Conjecture 1 holds for an arbitrary  $AC^0$  function  $g$  and a random matrix  $A$ .

**Proposition 7.** *Let  $g: \{0, 1\}^m \rightarrow \{-1, 1\}$  be an AND/OR circuit of size  $s \geq n$  and depth  $d$ ,  $A$  be an  $m \times n$   $GF(2)$  matrix so that every set of  $r = (\log s)^{O(d^3)}$  rows of  $A$  is linearly independent, and  $f(x) = g(Ax)$ . Then there exists an  $a \in \{0, 1\}^n$  for which  $|\hat{f}(a)| \geq m^{-O((\log s)^d)}$ .*

Our proof relies on the following fundamental lemmas about  $AC^0$  circuits of Linial, Mansour, and Nisan and Braverman.

**Lemma 8** (Linial et al. [LMN93]). *Let  $g: \{0, 1\}^m \rightarrow \{-1, 1\}$  be an AND/OR circuit of size  $s \geq n$  and depth  $d$ . For any  $t$ ,*

$$\sum_{|b| > t} \hat{g}(b)^2 \leq 2s2^{-t^{1/d}/20}$$

**Lemma 9** (Braverman [Bra11]). *Let  $f: \{0, 1\}^n \rightarrow \{-1, 1\}$  be a function computed by a size  $s$  depth  $d$  AND/OR circuit, and  $\mathcal{D}$  be an  $r$ -wise independent distribution over  $\{0, 1\}^n$ , where  $r \geq r(s, d, \varepsilon) = (\log(s/\varepsilon))^{O(d^2)}$ . Then  $|\mathbb{E}_{z \sim \mathcal{D}}[f(z)] - \mathbb{E}_{z \sim \{0, 1\}^n}[f(z)]| \leq \varepsilon$ .*

*Proof of Proposition 7.* We apply Lemma 8 with  $t = (20 \log 4s)^d$ . In particular, there must exist at least one  $b \in \{0, 1\}^m$  with Hamming weight at most  $t$  for which

$$\hat{g}(b)^2 \geq 1/\left(2 \sum_{i=0}^t \binom{m}{i}\right) = m^{-O(\log s)^d}.$$

We now show that for  $a = A^T b$ ,  $|\hat{f}(a)| \geq |\hat{g}(b)|/2 = m^{-O(\log s)^d}$ .

Consider the function  $h(z) = g(z)\chi_b(z)$ , where  $\chi_b(z) = (-1)^{\sum z_i}$ . Håstad [Hås87] observes that  $\chi_b$  can be computed by a circuit of depth  $d$  and size  $2^{t^{1/(d-1)}}$ . Therefore  $h$  is computable by a circuit of size  $s + 2^{O(t^{1/(d-1)})} = 2^{O(\log s)^{d/(d-1)}}$  and depth  $d + 1$ . Instantiating Lemma 9 with the appropriate parameters, we get that

$$|\mathbb{E}_{z \sim \mathcal{D}}[h(z)] - \mathbb{E}_{z \sim \{0,1\}^m}[h(z)]| \leq |\hat{g}(b)|/2, \quad (1)$$

as long as  $\mathcal{D}$  is  $r$ -wise independent, where

$$r = r(2^{O(\log s)^{d/(d-1)}}, d + 1, |\hat{g}(b)|/2) \geq r(2^{O(\log s)^{d/(d-1)}}, d + 1, m^{-O(\log s)^d}) = (\log s)^{O(d^3)}.$$

By our assumption on the matrix  $A$ , the distribution  $Ax$ , where  $x \sim \{0,1\}^n$  is  $r$ -wise independent. Substituting into (1) we get

$$|\mathbb{E}_{x \sim \{0,1\}^n}[g(Ax)\chi_b(Ax)] - \mathbb{E}_{z \sim \{0,1\}^m}[g(z)\chi_b(z)]| \leq |\hat{g}(b)|/2.$$

This inequality can be rewritten as

$$|\hat{f}(A^T b) - \hat{g}(b)| \leq |\hat{g}(b)|/2$$

so by the triangle inequality,  $|\hat{f}(A^T b)| \geq |\hat{g}(b)|/2$ .  $\square$

**Corollary 10.** *Let  $A$  be a random  $m \times n$   $GF(2)$  matrix. With probability  $1 - 2^{-\Omega(n)}$  over the choice of  $A$ , for every AND/OR circuit  $g: \{0,1\}^m \rightarrow \{-1,1\}$  of size  $n \leq s \leq \exp((n/\log m)^{1/O(d^3)})$  and depth  $d$ , and  $f(x) = g(Ax)$ , there exists an  $a \in \{0,1\}^n$  for which  $|\hat{f}(a)| \geq m^{-O((\log s)^d)}$ .*

*Proof.* It is sufficient to show that with probability  $1 - 2^{-\Omega(n)}$  over the choice of  $A$ , every  $r = (\log s)^{\text{poly}d}$  rows of  $A$  are linearly independent. By a union bound, the probability that there exists a linear dependence between  $r$  or fewer rows is at most

$$\frac{1}{2^n} \sum_{i=0}^r \binom{m}{i} \leq \frac{(r+1)m^r}{2^n} \leq 2^{-\Omega(n)},$$

where the last inequality follows from our assumption that  $s \leq \exp((n/\log m)^{1/O(d^3)})$ .  $\square$

### 3 $AC^0$ functions inapproximable by low-degree polynomials

In order to apply Conjecture 2 towards obtaining a weak PRF in the class  $AC^0 \circ \text{MOD}_2$ , we need a function  $g$  in the class  $AC^0 \circ \text{MOD}_2$  that (1) is unbiased and (2) is inapproximable by  $GF(2)$  polynomials of low degree.

We show that the  $\text{TRIBES}_{n/k,k}$  function with  $n = \Theta(k2^k)$  satisfies condition (2). We prove Theorem 3 which states that  $\text{TRIBES}_{n/k,k}$  is  $2^{-3d}$ -far from all polynomials of degree at most  $n - k$ . We remark that we didn't try to optimize the constant and the bound  $2^{-O(d)}$  may not be tight.

To meet condition (1) (which the  $\text{TRIBES}$  function family does not appear to satisfy), we give a simple general transformation that preserves correlation with degree  $d$  polynomials for every  $d$  but turns any function into an unbiased one. Let  $\text{Corr}_d(f)$  denote the maximum correlation between  $f: \{0,1\}^n \rightarrow \{-1,1\}$  and degree- $d$  polynomials:  $\text{Corr}_d(f) = \max_p \mathbb{E}_{x \sim \{0,1\}^n}[f(x) \cdot (-1)^{p(x)}]$  where  $p$  is  $GF(2)$  polynomial of degree  $d$ .

**Claim 11.** *Let  $f: \{0,1\}^n \rightarrow \{-1,1\}$  be a boolean function. Let  $g: \{0,1\}^{n+1} \rightarrow \{-1,1\}$  be given by  $g(y,z) = f(y) \cdot (-1)^z$ . Then for every  $d \geq 1$ ,  $\text{Corr}_d(g) = \text{Corr}_d(f)$ .*

Combining Theorem 3 and Claim 11 yields Corollary 4.

### 3.1 Proof of Theorem 3

Our proof relies on the following special case of Schwarz-Zippel Lemma.

**Lemma 12** (Schwarz-Zippel Lemma [Sch80, Zip79]). *If  $p$  is a nonzero multilinear polynomial over  $GF(2)$  of degree  $d$ , then  $\Pr_x[p(x) \neq 0] \geq 2^{-d}$ .*

For simplicity of notation, we denote  $\text{TRIBES}_{n/k,k}$  as  $T$ . Suppose that  $\Pr_x[p(x) \neq T(x)] < 2^{-3d}$  where  $0 < d < n - k$ . We decompose the tribes function  $T$  of interest as  $T(x_1, x_2) = T_1(x_1) \vee T_2(x_2)$ , where  $T_1$  and  $T_2$  are tribes on  $l$  and  $n - l$  variables respectively, where  $l = \lceil (d + 1)/k \rceil \cdot k$ . The size of each tribe is  $k$  in both  $T_1$  and  $T_2$ . We divide the analysis into two cases:  $d \geq k$  and  $d < k$ .

**$d$  is large.** If  $d \geq k$ , we show there exists an  $x_2$  such that  $T_2(x_2) = 0$  and for any  $x_1$ ,  $T(x_1, x_2) = p(x_1, x_2)$ . It follows  $p(x_1, x_2) = T_1(x_1)$ . However, the expansion of  $T_1(x)$  as a  $GF(2)$  polynomial is unique and can be easily checked to have degree  $l \geq d + 1$ . But  $p$  only has degree  $d$ . Contradiction!

By our choice of parameters, note that  $3d - l \geq 2d - k \geq d$  and  $\alpha < 2^{-k}n/k < (1 - \alpha)$ . We show the existence of  $x_2$  by the probabilistic method.

$$\begin{aligned} \Pr_{x_2}[T_2(x_2) \neq 0] &= 1 - (1 - 2^{-k})^{(n-l)/k} \\ &\leq 1 - (1 - (n-l)/k)2^{-k} \\ &= 2^{-k}(n-l)/k \\ &< 2^{-k}n/k < 1 - \alpha. \end{aligned}$$

$$\begin{aligned} \Pr_{x_2}[\exists x_1, T(x_1, x_2) \neq p(x_1, x_2)] &\leq \sum_{x_1} \Pr_{x_2}[T(x_1, x_2) \neq p(x_1, x_2)] \\ &\leq 2^l \Pr_{x_1, x_2}[T(x_1, x_2) \neq p(x_1, x_2)] \\ &\leq 2^{l-3d} \leq 2^{-d}. \end{aligned}$$

Thus the existence of  $x_2$  follows from

$$\Pr_{x_2}[T_2(x_2) = 0 \wedge \forall x_1, T(x_1, x_2) = p(x_1, x_2)] \geq 1 - (1 - \alpha) - 2^{-d} = \alpha - 2^{-d} = \Omega(1) > 0.$$

**$d$  is small.** Now we prove the case  $0 < d < k$ . By Markov's inequality, with probability at least  $1 - 2^{-d}$  over the choice of  $x_2$  we have  $\Pr_{x_1}[p(x_1, x_2) \neq T(x_1, x_2)] \leq 2^{-2d}$ . Notice that  $T_1$  is AND function over  $k$  bits. For those choices of  $x_2$ ,

$$\begin{aligned} \Pr_{x_1}[p(x_1, x_2) \neq T_2(x_2)] &\leq \Pr_{x_1}[p(x_1, x_2) \neq T_2(x_2) \vee T(x_1, x_2) \neq T_2(x_2)] \\ &\leq \Pr_{x_1}[p(x_1, x_2) \neq T(x_1, x_2)] + \Pr_{x_1}[T(x_1, x_2) \neq T_2(x_2)] \\ &\leq 2^{-2d} + 2^{-k} \\ &< 2^{-d}. \end{aligned}$$

By Schwarz-Zippel Lemma,  $\Pr_{x_1}[p(x_1, x_2) \neq T_2(x_2)] < 2^{-d}$  where  $p(x_1, x_2)$  has degree at most  $d$  implies  $p(x_1, x_2) - T_2(x_2)$  is the zero polynomial in  $x_1$ . Thus for at least  $1 - 2^{-d}$  choice of  $x_2$ ,

$p(x_1, x_2)$  is a constant function in  $x_1$ . We show  $p$  doesn't depend on  $x_1$ . For otherwise, we could write  $p$  as

$$p(x_1, x_2) = \sum_{T \subseteq [k]} f_T(x_2) \prod_{i \in T} x_{1i}$$

where  $x_{1i}$  denote the  $i$ th bit in  $x_1$  and there exists a non-empty  $T$  such that  $f_T$  is non-zero polynomial. Since  $f_T$  has degree at most  $d$ , by the Schwarz-Zippel Lemma, with probability at least  $1 - 2^{-d}$  over the choice of  $x_2$ ,  $f_T(x_2)$  is non-zero thus  $p(x_1, x_2)$  is not a constant function in  $x_1$ . Contradiction!

More generally, we can decompose  $T$  as  $T_1 \vee T_2$  where  $T_1$  is AND of the  $i$ th block of inputs and  $T_2$  is tribes over remaining inputs (or blocks). By the same reasoning, we can show  $p$  doesn't depend on  $i$ th block of inputs. Therefore  $p$  is a constant function which doesn't depend on any inputs which contradicts the assumption  $d \neq 0$ .

### 3.2 Proof of Claim 11

To prove Claim 11, notice that the correlation of  $f(y)$  with a polynomial  $p(y)$  equals the correlation of  $g(y, z)$  with the polynomial  $p(y) \oplus z$ . In the other direction, suppose  $g(y, z)$  and  $p(y, z)$  have correlation  $\gamma$ . By averaging  $z$  can be fixed to a constant  $c \in \{0, 1\}$  so that  $g(y, c)$  and  $p(y, c)$  have correlation at least  $\gamma$ . Then  $f(y)$  has correlation at least  $\gamma$  with the polynomial  $p(y, c) \oplus c$ .

Corollary 4 follows from Theorem 3 and Claim 11.

## 4 Lack of Correlation with Fixed Function Families

One main efficiency measure of statistical query learning algorithms is the number of statistical properties (also called statistical queries) the algorithms needs to estimate. In particular, the statistical queries lower bound the running time of statistical query learning algorithms. The characterization of the number of statistical queries was studied in [BFJ<sup>+</sup>94, BF02, Fel09]. In particular, Bshouty and Feldman [BF02] showed if a function family can be weakly learned by statistical query algorithm with  $s$  queries, then we find a collection of functions of size  $s + 1$  such that every function  $f \in \mathcal{F}$  correlates with  $\mathcal{H}$ .

**Theorem 13** ([BF02]). *Let  $\mathcal{F} = \{h: \{0, 1\}^n \rightarrow \{-1, 1\}\}$  be a function family. Suppose there exists an algorithm that for any  $f \in \mathcal{F}$  uses at most  $s$  statistical queries with error at most  $\delta$  to predict  $f$  on random inputs with probability  $1/2 + \varepsilon$ . There exists a collection of functions  $\mathcal{H} = \{h: \{0, 1\}^n \rightarrow \{-1, 1\}\}$  of size  $s + 1$  such that for any  $f \in \mathcal{F}$ , there exists  $h \in \mathcal{H}$ ,  $|\mathbb{E}_x[f(x)h(x)]| \geq \min(2/\delta, 1/\varepsilon)$ .*

Corollary 15 below shows that statistical query learning algorithms cannot work against our family  $F_A$  if  $g$  is unbiased and  $A$  is square matrix: for any fixed function family  $\mathcal{H}$  of size at most  $2^{n/4}$ , the correlation between  $F_A$  and any function in  $\mathcal{H}$  is exponentially small with overwhelming probability over the choice of the key, which implies, by Theorem 13, any statistical query learning algorithm needs an exponential number of statistical queries to learn noticeable fraction of functions in this family (which is required to break pseudorandomness of weak PRFs).

**Proposition 14.** *For any function  $h: \{0, 1\}^n \rightarrow \{-1, 1\}$ ,*

$$\Pr_A[|\mathbb{E}_x[g(Ax)h(x)] - \mathbb{E}_x[g(x)]\mathbb{E}_x[h(x)]| > \varepsilon + 2^{-n+1}] = O(2^{-n}/\varepsilon^2)$$

where  $A$  is a random  $n \times n$  matrix.

Our proof is essentially a second moment calculation which relies on the pairwise independence of the values  $Ax$ .

*Proof.* Consider the random variable  $Z(A) = \mathbb{E}_x[g(Ax)h(x)]$ . We will estimate the first and second moments of this random variable.

Conditioned on  $x \neq 0$ ,  $Ax$  is uniformly distributed in  $\{0, 1\}^m$  and independent of  $x$ . Therefore

$$\begin{aligned} \mathbb{E}_A[Z(A)] &= \frac{1}{2^n} \sum_{x \neq 0} \mathbb{E}_A[g(Ax)h(x)] + \frac{1}{2^n} g(0)h(0) = \frac{1}{2^n} \sum_{x \neq 0} h(x) \mathbb{E}_y[g(y)] + \frac{1}{2^n} g(0)h(0) \\ &= \mathbb{E}_x[h(x)] \mathbb{E}_y[g(y)] - \frac{1}{2^n} h(0) \mathbb{E}_y[g(y)] + \frac{1}{2^n} g(0)h(0). \end{aligned}$$

Since  $|h(0) \mathbb{E}_y[g(y)]| \leq 1$  and  $|g(0)h(0)| \leq 1$ , we can obtain

$$|\mathbb{E}_A[Z(A)] - \mathbb{E}_y[g(y)] \mathbb{E}_x[h(x)]| \leq \left| \frac{1}{2^n} h(0) \mathbb{E}_y[g(y)] \right| + \left| \frac{1}{2^n} g(0)h(0) \right| \leq 2^{-n+1}.$$

All probabilities are over the uniform distribution. For the second moment, we have

$$\begin{aligned} \mathbb{E}_A[Z(A)^2] &= \mathbb{E}_A[\mathbb{E}_x[g(Ax)h(x)]^2] \\ &= \mathbb{E}_A[\mathbb{E}_x[g(Ax)h(x)] \mathbb{E}_{x'}[g(Ax')h(x')]] \\ &= \mathbb{E}_{x,x'}[h(x)h(x') \mathbb{E}_A[g(Ax)g(Ax')]]. \end{aligned}$$

Fix  $x$  and  $x'$  satisfying  $x \neq x'$  and  $x, x' \neq 0$ . Then  $Ax$  and  $Ax'$  are independent and uniformly distributed in  $\{0, 1\}^n$  over the choice of  $A$ , so  $\mathbb{E}_A[g(Ax)g(Ax')] = \mathbb{E}_y[g(y)]^2$ . Since the event “ $x = x'$  or  $x' = 0$  or  $x = 0$ ” happens with probability at most  $3 \cdot 2^{-n}$ , it follows that

$$\mathbb{E}_A[Z(A)^2] \leq \mathbb{E}_x[h(x)]^2 \mathbb{E}_x[g(x)]^2 + 3 \cdot 2^{-n}$$

and

$$\text{Var}_A[Z(A)] = \mathbb{E}_A[Z(A)^2] - \mathbb{E}_A[Z(A)]^2 \leq 7 \cdot 2^{-n}.$$

The proposition follows by applying Chebyshev’s inequality to  $Z(A)$ .  $\square$

By applying a union bound and setting parameters appropriately, we obtain the following corollary:

**Corollary 15.** *Let  $\mathcal{H}$  be any collection of functions  $h: \{0, 1\}^n \rightarrow \{-1, 1\}$  of size at most  $2^{n/4}$ . With probability  $2^{-\Omega(n)}$  over the choice of  $A$ , for every  $h \in \mathcal{H}$ ,  $\mathbb{E}_x[g(Ax)h(x)] \leq \beta + 2^{-\Omega(n)}$ , where  $\beta = \mathbb{E}_x[g(x)]$ .*

The  $\beta$  term is necessary; if  $g$  is biased then  $g(Ax)$  correlates with the zero function.

## 5 Noisy parities, rounding, and weak pseudorandom functions

As observed in the introduction, assuming Conjecture 1, the hardness of learning noisy parities (LPN) is a necessary assumption for the existence of weak PRFs in  $\text{AC}^0 \circ \text{MOD}_2$ . In this section we investigate whether this assumption is also sufficient.

For this purpose we view our function  $f_A(x) = g(Ax)$  as applying a “rounding” function  $g$  that adds a “deterministic noise” to the samples  $Ax$ . This viewpoint has proved instrumental in the context of learning with errors (LWE), which is a generalization of LPN to larger modulus  $q$ . Specifically, Banerjee et al. [BPR12] construct a weak PRF  $f'_A(x) = g'(Ax)$  where  $A, x$  and all operations are over  $GF(q)$  for a sufficiently large modulus  $q$ ; they then prove that for a suitably chosen rounding function  $g'$ , their “deterministic noise” is statistically close to an LWE noise, implying that their function  $f'$  is as hard as LWE; finally, as  $g'$  has negligible bias, they conclude that  $f'_A(x) = g'(Ax)$  is computationally indistinguishable from uniform; namely,  $f'$  is a weak PRF. Can their proof techniques [BPR12] be transferred to the field  $GF(2)$  for basing the hardness of our candidate function on LPN? In Theorem 16 we give a negative answer to this question by showing that for every function  $g$ , if the “deterministic noise” incurred by  $g$  is statistically close to an LPN noise, then  $g$  is highly biased (ie,  $g$  is close to a constant function and thus  $f(x) = g(Ax)$  cannot be a weak PRF).

Elaborating on the above, Banerjee et al. [BPR12] observe that under the LWE assumption (and for a suitable  $g'$ ), samples of the form  $(x, g'(\langle a, x \rangle))$  are computationally indistinguishable from samples  $(x, g'(u))$ , where  $x \sim GF(q)^n$  and  $u \sim GF(q)$  is independent of  $x$ . This follows by looking at the auxiliary distribution  $(x, g'(\langle a, x \rangle + e))$ , where  $e$  follows the LWE noise distribution. On the one hand, for a suitable choice of  $g'$ ,  $(x, g'(\langle a, x \rangle))$  and  $(x, g'(\langle a, x \rangle + e))$  are statistically close. On the other hand, by the LWE assumption,  $(x, g'(\langle a, x \rangle + e))$  and  $(x, g'(u))$  are computationally close. Thus, as  $g'$  has negligible bias, they conclude that the samples  $(x, g'(\langle a, x \rangle))$  are computationally indistinguishable from uniform samples  $(x, u)$ ; namely,  $f'$  is a weak PRF.

Transferring the proof technique of [BPR12] to the case  $q = 2$  would look as follows. Suppose the adversary sees  $t$  samples

$$(x_1, g(Ax_1)), \dots, (x_t, g(Ax_t)), \quad (2)$$

and consider the auxiliary distribution

$$(x_1, g(Ax_1 + e_1)), \dots, (x_t, g(Ax_t + e_t)) \quad (3)$$

where  $e_1, \dots, e_t \in \{0, 1\}^m$  follow the LPN noise distribution with rate  $\eta$  (ie, their coordinates are i.i.d. random variables accepting 1 with probability  $\eta$ ); denote this distribution  $\{0, 1\}_\eta^m$ . On the one hand, argue for some choice of  $g$  that distributions (2) and (3) are statistically close. On the other hand, by the LPN assumption<sup>4</sup> distribution (3) is computationally close to

$$(x_1, g(u_1)), \dots, (x_t, g(u_t))$$

where  $u_1, \dots, u_t \sim \{0, 1\}^m$  are independent of  $x_1, \dots, x_t$ . Finally, using the assumption that  $g$  has negligible bias, it would follow that  $g(Ax)$  is pseudo-random.

The following theorem shows that this proof method cannot work: essentially if distributions (2) and (3) are statistically close, then  $g$  must be biased. Specifically, if the statistical distance is  $\varepsilon \leq \frac{\eta^2}{m^2} \cdot \frac{1}{n^{2c}}$  for  $c > 0$  a constant and sufficiently many samples say  $t = \frac{m^2}{\eta^2} \cdot mn \cdot n^{2c}$  (note that  $t = \text{poly}(n)$  for  $m = \text{poly}(n)$  and LPN noise rate  $\eta$  which is noticeable), then we obtain that  $\text{Bias}[g]^2 = 1 - O(n^{-c})$ .

---

<sup>4</sup>The values  $(x_1, Ax_1 + e_1), \dots, (x_t, Ax_t + e_t)$  can be viewed as noisy samples for multiple random secrets. Specifically, the random secrets are the rows  $a_j$  of the matrix  $A$ , and the corresponding noisy samples are  $(x_1, \langle a_j, x_1 \rangle + e_1), \dots, (x_t, \langle a_j, x_t \rangle + e_t)$ . Such samples are computationally indistinguishable from uniform samples under LPN assumption (this is straightforward to prove, due to the self-reducibility of LPN).

**Theorem 16.** For  $n, m, t, \eta$  as above, and a function  $g: \{0, 1\}^m \rightarrow \{-1, 1\}$ , if distributions (2) and (3) are within statistical distance  $\varepsilon$ , then  $\text{Bias}[g]^2 \geq 1 - \frac{m}{\eta} \cdot (5\sqrt{\delta} + \sqrt{\varepsilon} + 1/2^n)$  for  $\delta = \frac{1}{t} \cdot (nm \ln 2 + \ln(1 - \sqrt{\varepsilon}))$ .

We remark that Theorem 16 can be generalized to the case where  $f(x) = g_x(Ax)$  ie  $g$  depends on  $x$ . Specifically, we can conclude for most choice of  $x$ ,  $g_x$  is biased ( $g_x$  is close to constant conditioned on  $x$ ) thus  $f(x) = g_x(Ax)$  cannot be weak PRF.

## 5.1 Proof of Theorem 16

In this section we prove Theorem 16 by showing that the following holds for a random matrix  $A$  (with high probability over the choice of the input  $x$  and the noise  $e$ ): First we give a statistical test and show that if this test cannot distinguish distributions (2) and (3) then we can eliminate the noise  $e$  by replacing  $A$  with a related matrix  $A'$ ; namely,  $g(Ax + e) = g(A'x)$  (see Claim 17). Then we show that even with a slightly higher noise rate the above still holds, ie,  $g(Ax + e + e') = g(A'x)$ ; implying that  $g(Ax + e + e') = g(Ax + e)$  (see Claim 19). Next, we employ the above to upper bound the noise-sensitivity of  $g$ . Specifically, we observe that  $Ax + e$  is uniform (because  $A$  is uniform), thus replacing  $Ax + e$  by  $u$  in the above we conclude that  $g(u + e') = g(u)$  with high probability; namely,  $g$  has low noise-sensitivity (see Claim 20). Finally we employ a simple relation between the noise-sensitivity and the bias to conclude that  $g$  has high bias (see Proposition 21). The proof details follow.

First we show that for most matrices  $A$  there is a matrix  $A'$  s.t.  $g(Ax + e) = g(A'x)$  with high probability over the choice of  $x, e$ .

**Claim 17.** With probability at least  $1 - \sqrt{\varepsilon}$  over the choice of  $A \sim \{0, 1\}^{m \times n}$ , there exists  $A' \in \{0, 1\}^{m \times n}$  such that

$$\Pr_{x,e}[g(Ax + e) = g(A'x)] \geq 1 - \delta$$

where  $x \sim \{0, 1\}^n, e \sim \{0, 1\}_\eta^m$ .

*Proof.* Consider statistical test  $T$ : on input  $(x_1, b_1, \dots, x_t, b_t)$ , output 1 if and only if there exists  $A \in \{0, 1\}^{m \times n}$  such that  $g(Ax_i) = b_i$  for all  $1 \leq i \leq t$ . Since distributions (2) and (3) are  $\varepsilon$  statistically close and  $D$  always accepts distribution (2),

$$\Pr_{A, x_i, e_i}[T(x_1, g(Ax_1 + e_1), \dots, x_t, g(Ax_t + e_m)) = 1] \geq 1 - \varepsilon.$$

By Markov's inequality, with probability at least  $1 - \sqrt{\varepsilon}$  over the choice of  $A$ ,

$$\Pr_{x_i, e_i}[T(x_1, g(Ax_1 + e_1), \dots, x_t, g(Ax_t + e_m)) = 1] \geq 1 - \sqrt{\varepsilon}. \quad (4)$$

For any such  $A$ , by definition of  $T$ ,

$$\begin{aligned} \Pr_{x_1, e_1, \dots, x_t, e_t}[\exists A' \forall i, g(Ax_i + e_i) = g(A'x_i)] &\leq \sum_{A' \in \{0, 1\}^{m \times n}} \Pr_{x_1, e_1, \dots, x_t, e_t}[\forall i, g(Ax_i + e_i) = g(A'x_i)] \\ &\leq \sum_{A' \in \{0, 1\}^{m \times n}} (\Pr_{x,e}[g(Ax + e) = g(A'x)])^t \\ &\leq 2^{mn} \max_{A' \in \{0, 1\}^{m \times n}} (\Pr_{x,e}[g(Ax + e) = g(A'x)])^t. \end{aligned}$$

Assuming that  $\max_{A' \in \{0,1\}^{m \times n}} \Pr_{x,e}[g(Ax + e) = g(A'x)] < 1 - \delta$ , then

$$\max_{A' \in \{0,1\}^{m \times n}} 2^{mn} (\Pr_{x,e}[g(Ax + e) = g(A'x)])^t < 2^{mn} (1 - \delta)^t \leq 2^{mn} e^{-t\delta} = 1 - \sqrt{\varepsilon},$$

which contradicts inequality (4). Therefore,  $\max_{A' \in \{0,1\}^{m \times n}} \Pr_{x,e}[g(Ax + e) = g(A'x)] \geq 1 - \delta$ .  $\square$

Next we show that  $g(Ax + e + e') = g(Ax + e)$  (for most matrices  $A$ , and with high probability over  $x, e, e'$ ) in Claim 19. To prove this, we show how to eliminate noise of a slightly higher rate than considered in Claim 17, ie,  $g(Ax + e + e') = g(A'x)$  which relies on following lemma.

**Lemma 18.** *For any  $0 < \eta < 1/2$  and  $h: \{0,1\}^m \rightarrow \{0,1\}$ ,  $\Pr_{e,e'}[h(e + e') \neq 0] \leq 3 \Pr_e[h(e) \neq 0]$  where  $e \sim \{0,1\}_\eta^m, e' \sim \{0,1\}_{\eta/m}^m$ .*

*Proof.* For any  $z \in \{0,1\}^m$ ,  $\Pr_{e,e'}[e + e' = z] \leq (1 - \eta)^{m-|z|} (1 + 1/m)^m \eta^{|z|} \leq (1 + 1/m)^m \Pr_e[e = z] < 3 \Pr_e[e = z]$  so that  $\Pr_{e,e'}[h(e + e') = 0] = \sum_{z:h(z)=0} \Pr_{e,e'}[e + e' = z] \leq 3 \sum_{z:h(z)=0} \Pr_e[e = z] = 3 \Pr_e[h(e) = 0]$ .  $\square$

**Claim 19.** *With probability at least  $1 - \sqrt{\varepsilon}$  over the choice of  $A \sim \{0,1\}^{m \times n}$ , there exists  $A' \in \{0,1\}^{m \times n}$  such that*

$$\Pr_{x,e}[g(Ax + e) = g(Ax + e + e')] \geq 1 - 5\sqrt{\delta}$$

where  $x \sim \{0,1\}^n, e \sim \{0,1\}_\eta^m, e' \sim \{0,1\}_{\eta/m}^m$ .

*Proof.* For  $A, A', x$ , let  $h_{A,A',x}(z) = 0$  if and only if  $g(Ax + z) \neq g(A'x)$ . By union bound,

$$\begin{aligned} \Pr_{e,e'}[g(Ax + e) \neq g(Ax + e + e')] &\leq \Pr_{e,e'}[g(Ax + e + e') \neq g(A'x)] + \Pr_e[g(Ax + e) \neq g(A'x)] \\ &= \Pr_{e,e'}[h_{A,A',x}(e + e') = 0] + \Pr_e[h_{A,A',x}(e) = 0] \\ &\leq 4 \Pr_e[h_{A,A',x}(e) = 0] \end{aligned}$$

where the last inequality is due to Lemma 18. By Claim 17 and Markov's inequality, for at least  $1 - \sqrt{\varepsilon}$  choice of  $A$  such that at least  $1 - \sqrt{\delta}$  choice of  $x$ ,

$$\Pr_e[g(Ax + e) \neq g(A'x)] = \Pr_e[h_{A,A',x}(e) = 0] \leq \sqrt{\delta}.$$

Therefore for at least  $1 - \sqrt{\varepsilon}$  choice of  $A$  such that at least  $1 - \sqrt{\delta}$  choice of  $x$ ,

$$\Pr_{e,e'}[g(Ax + e) \neq g(Ax + e + e')] \leq 4 \Pr_e[h_{A,A',x}(e) = 0] \leq 4\sqrt{\delta},$$

which implies the desired conclusion.  $\square$

Next we bound the noise-sensitivity of  $g$ , where recall that the noise sensitivity of a Boolean function  $g$  is defined to be

$$\text{NS}_{e'}(g) = \Pr_{u,e'}[g(u + e') \neq g(u)],$$

where  $u \sim \{0,1\}^m$  and  $e' \in \{0,1\}_{\eta/m}^m$ .

**Claim 20.**  $\text{NS}_{e'}(g) \leq 5\sqrt{\delta} + \sqrt{\varepsilon} + 1/2^n$ .

*Proof.* Claim 19 implies  $\Pr_{A,x,e,e'}[g(Ax + e) = g(Ax + e + e')] \geq 1 - 5\sqrt{\delta} - \sqrt{\varepsilon}$ . Since for any fixed  $e, e'$  and non-zero  $x$ ,  $Ax$  is uniformly distributed, we can derive

$$\Pr_{u,e'}[g(u) = g(u + e')] \geq \Pr_{A,x,e,e'}[g(Ax + e) = g(Ax + e + e')] - 1/2^n = 1 - 5\sqrt{\delta} - \sqrt{\varepsilon} - 1/2^n.$$

Hence  $\text{NS}_{e'}(g) \leq 5\sqrt{\delta} + \sqrt{\varepsilon} + 1/2^n$ . □

To conclude the proof we employ the above bound on the noise-sensitivity to bound the bias.

**Proposition 21.** *For any  $g: \{0, 1\}^m \rightarrow \{-1, 1\}$  and  $0 < \eta' < 1/2$ ,  $\text{Bias}[g]^2 \geq 1 - \text{NS}_{e'}(g)/\eta'$ .*

*Proof.* Since  $\sum_{a \in \{0,1\}^m} \hat{g}^2(a) = 1$  and  $0 < \eta' < 1/2$ ,

$$\sum_{a \in \{0,1\}^m} \hat{g}^2(a)(1 - 2\eta')^{|a|} \leq \hat{g}^2(0^m) + \sum_{a \neq 0^m} \hat{g}^2(a)(1 - 2\eta') = 1 - 2\eta'(1 - \hat{g}^2(0^m)).$$

Recall the fact  $\text{NS}_{e'}(g) = \frac{1}{2} - \frac{1}{2} \sum_{a \in \{0,1\}^m} \hat{g}^2(a)(1 - 2\eta')^{|a|}$ . Therefore  $\text{NS}_{e'}(g) \geq \eta'(1 - \hat{g}^2(0^m))$  which implies  $\text{Bias}[g]^2 = \hat{g}^2(0^m) \geq 1 - \text{NS}_{e'}(g)/\eta'$ . □

## References

- [AKPW13] Joël Alwen, Stephan Krenn, Krzysztof Pietrzak, and Daniel Wichs. Learning with rounding, revisited: New reduction, properties and applications. In *Advances in Cryptology – CRYPTO 2013 – 33rd Annual Cryptology Conference*, pages 57–74, 2013.
- [BF02] Nader H Bshouty and Vitaly Feldman. On using extended statistical queries to avoid membership queries. *The Journal of Machine Learning Research*, 2:359–395, 2002.
- [BFJ<sup>+</sup>94] Avrim Blum, Merrick Furst, Jeffrey Jackson, Michael Kearns, Yishay Mansour, and Steven Rudich. Weakly learning dnf and characterizing statistical query learning using fourier analysis. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 253–262. ACM, 1994.
- [BFKL94] Avrim Blum, Merrick Furst, Michael Kearns, and Richard J Lipton. Cryptographic primitives based on hard learning problems. In *Advances in cryptology CRYPTO93*, pages 278–291. Springer, 1994.
- [BPR12] Abhishek Banerjee, Chris Peikert, and Alon Rosen. Pseudorandom functions and lattices. In *EUROCRYPT*, pages 719–737, 2012.
- [Bra11] Mark Braverman. Poly-logarithmic independence fools bounded-depth boolean circuits. *Communications of the ACM*, 54(4):108–115, 2011.
- [DKPW12] Yevgeniy Dodis, Eike Kiltz, Krzysztof Pietrzak, and Daniel Wichs. Message authentication, revisited. In *EUROCRYPT*, pages 355–374, 2012.
- [DN02] Ivan Damgård and Jesper Buus Nielsen. Expanding pseudorandom functions; or: From known-plaintext security to chosen-plaintext security. In *CRYPTO*, pages 449–464, 2002.

- [DR02] Joan Daemen and Vincent Rijmen. *The Design of Rijndael: AES - The Advanced Encryption Standard*. Information Security and Cryptography. Springer, 2002.
- [Fel09] Vitaly Feldman. A complete characterization of statistical query learning with applications to evolvability. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 375–384. IEEE, 2009.
- [FGKP09] Vitaly Feldman, Parikshit Gopalan, Subhash Khot, and Ashok Kumar Ponnuswami. On agnostic learning of parities, monomials, and halfspaces. *SIAM J. Comput.*, 39(2):606–645, 2009.
- [GGM84] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions (extended abstract). In *FOCS*, pages 464–479, 1984.
- [Gol00] Oded Goldreich. Candidate one-way functions based on expander graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 7(90), 2000.
- [Hås87] Johan Håstad. *Computational limitations of small-depth circuits*. MIT press, 1987.
- [Jac97] Jeffrey C Jackson. An efficient membership-query algorithm for learning dnf with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55(3):414–440, 1997.
- [Kea98] Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998.
- [LM13] Vadim Lyubashevsky and Daniel Masny. Man-in-the-middle secure authentication schemes from lpn and weak prfs. *IACR Cryptology ePrint Archive*, 2013:92, 2013.
- [LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.
- [MS07] Ueli M. Maurer and Johan Sjödin. A fast and key-efficient reduction of chosen-ciphertext to known-plaintext security. In *EUROCRYPT*, pages 498–516, 2007.
- [MV12] Eric Miles and Emanuele Viola. Substitution-permutation networks, pseudorandom functions, and natural proofs. In *CRYPTO*, pages 68–85, 2012.
- [NR97] Moni Naor and Omer Reingold. Number-theoretic constructions of efficient pseudorandom functions. In *FOCS*, pages 458–467, 1997.
- [Pie09] Krzysztof Pietrzak. A leakage-resilient mode of operation. In *EUROCRYPT*, pages 462–482, 2009.
- [Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Matematicheskie Zametki*, 41(4):598–607, 1987.
- [Reg10] Oded Regev. The learning with errors problem (invited survey). In *IEEE Conference on Computational Complexity*, pages 191–204, 2010.

- [RR94] Alexander A. Razborov and Steven Rudich. Natural proofs. In *STOC*, pages 204–213, 1994.
- [Sch80] Jacob T Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM (JACM)*, 27(4):701–717, 1980.
- [Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *STOC*, pages 77–82, 1987.
- [SV12] Rocco Servedio and Emanuele Viola. On a special case of rigidity. Technical Report TR12-144, Electronic Colloquium on Computational Complexity (ECCC), 2012.
- [Zip79] Richard Zippel. *Probabilistic algorithms for sparse polynomials*. Springer, 1979.