Limits of Preprocessing

Yuval Filmus[1], Yuval Ishai[1], Avi Kaplan[1], and Guy Kindler[2]

[1]Technion — Israel Institute of Technology
[2]Hebrew University of Jerusalem, Israel

April 9, 2020

Abstract

It is a classical result that the parity function cannot be computed by an AC^0 circuit [FSS81, Ajt83, Hås86]. It is conjectured that this holds even if we allow arbitrary preprocessing of each of the two inputs separately. We prove this conjecture when the preprocessing of the first input is limited to $n+n/(\log^{\omega(1)} n)$ bits. Our methods extend to many other functions, including pseudorandom functions, and imply a (weak but nontrivial) limitation on the power of encoding inputs in low-complexity cryptography. Finally, under cryptographic assumptions, we relate the question of proving variants of the main conjecture with the question of learning AC^0 under simple input distributions.

1 Introduction

Can preprocessing help in computation? This question, which arises in several areas of complexity theory, can be formalized in many ways. We consider the following version:

> Suppose that $f(x, y)\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a function hard for AC^0. Are there functions $\alpha, \beta\colon \{0,1\}^n \to \{0,1\}^{\mathsf{poly}(n)}$ such that $f(x, y)$ can be computed from $\alpha(x), \beta(y)$ using an AC^0 circuit?

We think of α, β as functions that preprocess the inputs x, y in order to make the computation of f easier. Alternatively, one can think of $\alpha(x)$ and $\beta(y)$ as messages sent simultaneously by two parties to an AC^0 referee, whose goal is to compute $f(x, y)$. An alternative formulation is:

> Let $\mathcal{F}$ be a collection of hard functions $f_y\colon \{0,1\}^n \to \{0,1\}$. Does there exist a function $\beta\colon \{0,1\}^n \to \{0,1\}^{\mathsf{poly}(n)}$ such that each $f_x(y) \in \mathcal{F}$ can be computed from $\beta(y)$ using an AC^0 circuit?

The two formulations are equivalent due to the completeness of circuit evaluation: if f_x can be computed efficiently from $\beta(y)$, then the function $f(x, y) = f_x(y)$ can be computed efficiently from $\beta(y)$ and the description of the circuit for f_x.

A simple example where preprocessing does help is when the function $f(x, y)$ depends only on the Hamming weights of x and y (e.g., $|x| > |y|$). Another simple example is any equivalence relation (e.g., graph isomorphism), where the two parties can send to the referee canonical representatives of the equivalence class of their respective inputs.

In contrast to the above examples, it is widely believed that for $f(x, y) = \sum_{i=1}^n x_i y_i \mod 2$ (also known as *mod-2 inner product*) the answer to the above questions is negative. Following Rothblum [Rot12], we refer to this as the *inner product with preprocessing* (IPPP) conjecture. Our main result proves a weak version of the IPPP conjecture, ruling out the utility of preprocessing when the output of α is short:

Theorem 1.1 (Main theorem, informal)**.** *Let f be the mod-2 inner product function, or alternatively any exponentially-secure cryptographic pseudorandom function, and let $m = n+n/(\log^{\omega(1)} n)$. There are no functions $\alpha\colon \{0,1\}^n \to \{0,1\}^m$ and $\beta\colon \{0,1\}^n \to \{0,1\}^{\mathsf{poly}(n)}$ for which $f(x,y)$ can be computed from $\alpha(x), \beta(y)$ using an AC^0 circuit.*

Our result is in fact more general, applying to a broad class of other functions, and ruling out not only AC^0 circuits, but also bounded depth circuits of subexponential size.

Our main theorem implies a modest but meaningful limitation on the power of preprocessing in low-complexity cryptography. There is a large body of work on minimizing the complexity of pseudorandom functions (PRFs) [GGM86]; see [BR17] for a survey. A recent work of Boneh et al. [BIP$^+$18] proposed a relaxed notion of PRF, dubbed "encoded-input PRF", that allows an arbitrary polynomial-time encoding of the input. This is motivated by several applications of low-complexity PRFs for which the relaxed notion suffices. The result of Linial et al. [LMN89] rules out the existence of PRFs (with better than quasipolynomial security) in the complexity class AC^0. A natural question is whether one can circumvent this impossibility by encoding the input. We show that such an encoding (if it exists) must have a nontrivial stretch.

As a final contribution, we relate the question of fully settling variants of the IPPP conjecture to another wide-open question: learning AC^0 under "simple" input distributions, such as polynomial-time samplable distributions, or uniform distributions over linear subspaces of $\mathbb{F}_2^n$. Under cryptographic assumptions from [BPR12, BIP$^+$18], we show that either (1) the known quasipolynomial time learning algorithm for AC^0 under the uniform distribution [LMN89] cannot be extended to other simple distributions, even with subexponential time; or (2) IPPP-style hardness conjectures are true. Put differently, progress on learning AC^0 (even under simple distributions and in subexponential time) would lead to proving IPPP-style conjectures under cryptographic assumptions. The latter currently seems difficult. The idea behind this connection is that the functions α and β corresponding to a refutation of an IPPP-style conjecture define a *reduction* from breaking "rounded inner-product" style (weak) PRF candidates to learning AC^0 under simple distributions.

1.1 Related Work

The power of preprocessing is relevant to many problems in computer science. For instance, the broad goal of *data structures* is to preprocess x into a polynomially longer $\hat{x} = \alpha(x)$, such that queries of the form $f(x,y)$ can be answered by reading few bits of $\hat{x}$. In our case, we replace "reading few bits of $\hat{x}$" by "computing an AC^0 function of $\hat{x}$". Below we survey several settings in complexity theory and cryptography that motivate this kind of questions.

Communication complexity — Polynomial hierarchy Communication complexity contains analogs of the familiar complexity classes of computational complexity. For example, P^{cc} consists of all two-party functions which can be computed using polylogarithmically many bits, and $\mathsf{NP}^{\mathsf{cc}}$ consists of all two-party functions which can be *verified* using polylogarithmically many bits.

An $\mathsf{NP}^{\mathsf{cc}}$ protocol for a function $f\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ proceeds as follows: an oracle sends the two parties the index of a combinatorial rectangle $X \times Y \subseteq \{0,1\}^n \times \{0,1\}^n$ on which $f = 1$, and the two parties verify that their inputs x, y belong to the rectangle: $x \in X$ and $y \in Y$; the complexity of the protocol is the length of the index. Equivalently, $f \in \mathsf{NP}^{\mathsf{cc}}$ if it can be written as a disjunction of $2^{\mathsf{polylog}(n)}$ combinatorial rectangles, that is, functions of the form "$x \in A$ and $y \in B$".

Babai, Frankl and Simon [BFS86] extended this by defining an analog of the polynomial hierarchy, $\mathsf{PH}^{\mathsf{cc}}$. A function $f\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ belongs to $\mathsf{PH}^{\mathsf{cc}}$ if it can be expressed as a constant depth circuit of quasipolynomial size $2^{\mathsf{polylog}(n)}$ whose leaves are combinatorial rectangles. Equivalently, $f \in \mathsf{PH}^{\mathsf{cc}}$ if it can be expressed as a constant depth circuits of quasipolynomial size whose leaves are arbitrary functions depending arbitrarily on one of the inputs. This is an instance of our main question, with a slight difference: $\mathsf{PH}^{\mathsf{cc}}$ allows the circuits to have quasipolynomial size.

Existing lower bound methods in communication complexity only go as far as $\mathsf{P}^{\mathsf{NP}^{\mathsf{cc}}}$ [GPW16] Nevertheless, it is a folklore conjecture that the inner product function IP lies outside $\mathsf{PH}^{\mathsf{cc}}$. This is considered as one of the most important outstanding open problems in the field.

Razborov [Raz89] showed that a function whose matrix representation is rigid doesn't belong to $\mathsf{PH}^{\mathsf{cc}}$ (see also [Wun12]), thus giving one potential avenue to prove lower bounds against $\mathsf{PH}^{\mathsf{cc}}$. Recently, in a surprising result, Alman and Williams [AW17] (see also [DE17]) showed that this method cannot show that $\mathsf{IP} \notin \mathsf{PH}^{\mathsf{cc}}$, since the inner product (or Hadamard) matrix isn't rigid enough! This shows that matrix rigidity cannot be used to prove the IPPP conjecture.

Communication complexity — Simultaneous messages and compression As noted above, the question we study can be naturally cast as a computationally bounded variant of the *simultaneous messages* (SM) model in communication complexity [Yao79, BGKL03]. In this model, $k \geq 2$ parties send their messages to a referee, who should immediately output the value of the function. In our case, $k = 2$ and the referee is limited to be an AC^0 circuit. On the other hand, the two parties are computationally unbounded, and the message sent by each party can be longer than its input.

A different communication complexity model that considers AC^0-bounded parties is the compression model from [DI06, CS12]. In this model, there is an AC^0 party whose goal is to compute the parity of its n-bit input x using the help of a computationally unbounded party, while minimizing the communication. This model is very different from ours; in particular, the model is trivialized if one allows n bits of communication.

Circuit complexity — Graph complexity Pudlák, Rödl and Savický [PRS88] developed the concept of *graph complexity* as a new approach to circuit lower bounds. Given a graph, we attempt to build it up from "axioms" using union, intersection, and complementation. In the particular case of *bipartite complexity*, the graph to be constructed is bipartite, and the axioms are complete bipartite graphs respecting the bipartition of the target graph.

A bipartite graph naturally defines a Boolean function with two inputs: the inputs are one vertex from each side, and the output is whether the edge exists. This correspondence shows that bipartite complexity is the same as a circuit whose leaves are combinatorial rectangles. Alternatively, we can allow each leaf to depend arbitrarily on one of the inputs, thus recovering our model of study.

Bipartite complexity can be studied for various circuit classes. One recent highlight is the work of Tal [Tal16], in which he shows that bipartite formulas computing IP must have quadratic size.

Circuit complexity — $\mathsf{AC}^0 \circ \mathsf{MOD}_2$ Our understanding of $\mathsf{AC}^0[p]$ circuits lacks compared to AC^0 circuits. While we have strong lower bounds against $\mathsf{AC}^0[p]$ circuits, the existing correlation bounds are significantly weaker, and this is a barrier for constructing pseudorandom generators for $\mathsf{AC}^0[p]$. Observing all of this, Servedio and Viola [SV12] suggest considering a weakening of $\mathsf{AC}^0[2]$, in which all parity gates appear in the bottom layer. They conjecture that inner product cannot be computed by such circuits, and prove their conjecture for depth-3 circuits. Akavia et al. [ABG$^+$14] give cryptographic applications for lower bounds against this class, and Cheragchi et al. [CGJ$^+$16] give superlinear lower bounds for inner product.

AC^0 circuits with parity gates at the bottom are the same as AC^0 circuits with *linear* preprocessing, namely where the preprocessing functions α, β are linear over $\mathbb{F}_2$. In other words, the conjecture of Serverdio and Viola is a special case of our conjecture, in which it suffices to rule out linear α, β.

Cryptography Our formulation of the IPPP conjecture is a close variant of the IPPP conjecture made by Rothblum [Rot12], where it was used to construct circuits resilient to AC^0 leakage. (The flavor of IPPP from [Rot12] is different from ours in that it restricts α, β to be polynomial-time computable and assumes hardness of approximation as opposed to just worst-case hardness.) In a recent work of Bogdanov et al. [BIS19], a similar result was obtained unconditionally.

As discussed above, our main question is strongly relevant to the goal of implementing cryptographic primitives in AC^0. The work of Boneh et al. [BIP$^+$18] poses the question of implementing an "encoded-input pseudorandom function" in AC^0, namely a pseudorandom function family $f_k(x)$, where each function f_k can be computed in AC^0 given an encoding of the input x. This is essentially the same as asking whether our main question can be answered affirmatively for some $f(x,y)$ such that $f_y(x)$ is a pseudorandom function family.

Extractors As part of his study of extractors for NC^0 and AC^0 sources, Viola [Vio11] constructed a function $f\colon \{0,1\}^n \to \{0,1\}$ such that the distribution $(\mathbf{x}, f(\mathbf{x}))$ (with $\mathbf{x}$ uniform) is hard for AC^0 to sample, even approximately. In particular, his results imply that the function $F\colon [n+1] \times \{0,1\}^n \to \{0,1\}$ given by

$$F(i,y) = \begin{cases} y_i & i \in [n], \\ f(y) & i = n+1, \end{cases}$$

cannot be computed by an AC^0 circuit from $\alpha(x), \beta(y)$, where $\alpha\colon [n+1] \to \{0,1\}^{\mathsf{poly}(n)}$ and $\beta\colon \{0,1\}^n \to \{0,1\}^n$. Therefore $F(x,y)$ requires exactly $n+1$ bits of preprocessing of y.

1.2 Overview of techniques

We outline the technique used for proving Theorem 1.1. The main tool we use is the LMN inequality [LMN89, Tal17], which states that AC^0 functions can be approximated by low degree functions. Let us illustrate the main idea behind the proof by sketching the proof of the following special case.

Theorem 1.2. *Let $\alpha\colon \{0,1\}^n \to \{0,1\}^*$ and let $\beta\colon \{0,1\}^n \to \{0,1\}^n$, and suppose that C is a bounded-depth circuit satisfying $C(\alpha(x), \beta(y)) = \mathsf{IP}(x,y)$ for all $x, y \in \{0,1\}^n$. Then C has exponential size.*

Since the inner product function is injective in each of its inputs, the preprocessing function α must be bijective.

For each $x \in \{0,1\}^n$, we can plug in the values $\alpha(x)$ to obtain constant-depth circuit C_x, of size at most that of C, satisfying $C_x(y) = \mathsf{IP}(x, \beta^{-1}(y))$ for all $y \in \{0,1\}^n$.

For any two $x \neq z$, the functions $f_x(y) = \mathsf{IP}(x,y)$ and $f_z(y) = \mathsf{IP}(z,y)$ are orthogonal (this is the well-known orthogonality of the Fourier characters). This property is crucially maintained by the functions C_x, C_z, which are also orthogonal.

Suppose that C has small size. We are thus in the following situation: we have 2^n orthogonal functions C_x, each of which can be approximated by a low degree function (by LMN), and so close to a low-dimensional subspace x of $\mathbb{R}[\{0,1\}^n]$. This is, however, impossible.

The argument works in much the same way for any function $f(x,y)$ which is injective in its first input and whose "slices" $f_x(y) = f(x,y)$ are approximately orthogonal on average. A short hybrid argument shows that PRFs fit the bill.

It is more challenging to extend the argument to functions β with larger output $\{0,1\}^m$. The basic idea is to complete the functions C_x, which are a priori defined only on 2^n of the 2^m possible inputs, to total functions which are still approximately orthogonal. Therefore if C has small size then one of the functions C_x will be far from V. On the other hand, since C_x agrees with a function computed by an AC^0 circuit on a 2^{n-m} fraction of the input, and that function is close to V. These two properties contradict each other.

This sketch explains why we can only expect to handle this way $m = n + o(n)$: if m is any larger, then the correlation of C_x with the output of the circuit is too small, and so we cannot reach any contradiction.

Organization After brief preliminaries (Section 2), we state our main results in Section 3, including Theorem 1.1 above. The connection to learning AC^0 functions under simple input distributions appears in Section 4. We prove our main technical theorem in Section 5, which is followed by applications to encoded-input PRFs (Section 6) and rounded inner product (Section 7).

2 Preliminaries

2.1 Definitions and notation

Simultaneous messages protocols A (two-party) simultaneous messages (SM) protocol consists of two players, which we refer to as Alice and Bob, and a referee, which we refer to as Carol, that together compute a function. Alice and Bob each send a message, which is based on the input, to Carol, who then computes a function of the two messages received. Formally, we have the following definitions:

Definition 2.1 (Simultaneous messages protocols)**.** Let $X, Y, \widehat{X}, \widehat{Y}, Z$ be finite nonempty sets. A *simultaneous messages protocol* (shortly, *SM protocol* or *SMP*) $\mathcal{P}$ is a triplet of functions (A, B, C), where $A\colon X \to \widehat{X}$, $B\colon Y \to \widehat{Y}$, and $C\colon \widehat{X} \times \widehat{Y} \to Z$. We call C the *referee function*.

Definition 2.2 (SM protocol admittance)**.** Let $f\colon X \times Y \to Z$ be a function. We say that f *admits an SM protocol* (A, B, C) if $f(x, y) = C(A(x), B(y))$ for every $(x, y) \in X \times Y$. In that case, we also say that (A, B, C) *computes* f.

Inner product space of Boolean functions For the purpose of utilizing Fourier analysis, we will consider the inner product space of all functions $\{-1, 1\}^n \to \mathbb{R}$ with the following inner product:

$$\langle f, g\rangle = \mathop{\mathrm{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}[f(\boldsymbol{x})g(\boldsymbol{x})] = \frac{1}{2^n}\sum_{x\in\{-1,1\}^n} f(x)\cdot g(x).$$

The Inner Product function The *inner product modulo* 2 function $\mathsf{IP}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined by

$$\mathsf{IP}(x, y) = \sum_{i=1}^{n} x_i y_i \pmod 2.$$

One-to-one condition Some functions have a certain property, formally defined below, rendering them harder to compute with preprocessing. Restricting attention to these functions seems to help in obtaining lower bounds.

Definition 2.3 (One-to-one condition)**.** Let $f\colon X \times Y \to Z$ be a function. We say that f *satisfies the left one-to-one condition* if for every $x \neq x' \in X$ there exists $y \in Y$ such that $f(x, y) \neq f(x', y)$. Similarly, we say that f *satisfies the right one-to-one condition* if for every $y \neq y' \in Y$ there exists $x \in X$ such that $f(x, y) \neq f(x, y')$. Finally, we say that f *satisfies the one-to-one condition* if f satisfies both the left and right one-to-one conditions.

Proposition 2.4 (One-to-one preprocessing)**.** *Let $f\colon X \times Y \to Z$ be a function. Then:*

- *If f satisfies the left one-to-one condition, then for every SM protocol (A, B, C) that f admits, A computes a one-to-one mapping.*
- *If f satisfies the right one-to-one condition, then for every SM protocol (A, B, C) that f admits, B computes a one-to-one mapping.*

2.2 Known facts

The following are known facts we will need later.

Theorem 2.5 (Tal's LMN improvement (LMNT) [LMN89, Tal17])**.** *Let f be a Boolean function with n variables computable by an unbounded fan-in circuit of depth h and size M, and let t be any integer. Then,*

$$\|f^{\geq t}\|^2 \leq 2 \cdot 2^{-t/O_h(\log M)^{h-1}}.$$

Lemma 2.6 (Lemma 3.6 in [Gra11]). *Let* $\mathrm{H}\colon [0,1] \to \mathbb{R}$ *be the* binary entropy function *defined by*

$$\mathrm{H}(p) = -p \log p - (1-p)\log(1-p).$$

Then, for any $0 < a \leq 1/2$ *and* $n \in \mathbb{N}$,

$$\binom{n}{\leq an} \leq 2^{\mathrm{H}(a)n}.$$

3 Main results

Our main technical result, proved in Section 5, states that a "large" collection of functions that are "close" to being orthonormal, is computationally hard for SM protocols in which the referee is an unbounded fan-in circuit of constant depth, and one player is limited to "short" preprocessing output length.

Theorem 3.1 (Main Theorem). *Let* $f\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *be a Boolean function, let* $0 \leq k \leq n/2 - 1$*, and let* $0 \leq t \leq n+k$ *be an integer. Denote* $f_x(y) \triangleq f(x,y)$*. Suppose the following hold:*

- f *satisfies the right one-to-one condition.*
- *There exists a subset* $X \subseteq \{-1,1\}^n$ *of size* $|X| \geq 13 \cdot 2^{2(k+1)} \cdot \binom{n+k}{\leq t}$ *such that*

$$\mathop{\mathrm{E}}_{\boldsymbol{x} \neq \boldsymbol{x}' \sim X}\left[\langle f_{\boldsymbol{x}}, f_{\boldsymbol{x}'}\rangle^2\right] \leq \frac{2^{2k}}{36|X|^2}.$$

- f *admits an SM protocol* $\mathcal{P} = (A,B,C)$ *such that* $B\colon \{-1,1\}^n \to \{-1,1\}^{n+k}$ *and* C *is an unbounded fan-in circuit of depth* h *and size* M*.*

Then:

$$M \geq 2^{\Omega_h\left(\left[\frac{t}{k}\right]^{1/(h-1)}\right)}.$$

A straightforward consequence of Theorem 3.1 is Theorem 1.1, which states that computing the inner product function with preprocessing remains as hard as without, given that one player is limited to output a string of length that stretches the input length by an additive sublinear factor:

Proposition 3.2 (Formal version of Theorem 1.1). *Let* $k \leq n/3$*, and suppose that* IP *admits an SM protocol* (A,B,C) *such that* $B\colon \{0,1\}^n \to \{0,1\}^{n+k}$ *and* C *is an unbounded fan-in circuit of depth* h *and size* M*. Then:*

$$M \geq 2^{\Omega_h\left(\left[\frac{n}{k}\right]^{1/(h-1)}\right)}.$$

Proof. We have:

- IP satisfies the right one-to-one condition.
- Choose $X = 2^{[n]}$, and let $0 < \alpha \leq 1/2$ be such that $\mathrm{H}(\alpha) \leq 1/8$. Then, if we set $t = \alpha(n+k)$, we get

$$13 \cdot 2^{2(k+1)} \cdot \binom{n+k}{\leq \alpha(n+k)} \underset{\text{Lemma 2.6}}{\leq} 2^{\mathrm{H}(\alpha)(n+k)+2(k+1)+4} \underset{k \leq n/3}{\leq} 2^n = |X|.$$

  Moreover, we have $\langle f_x, f_{x'}\rangle = \langle \chi_{S(x)}, \chi_{S(x')}\rangle = 0$ for any $x \neq x'$, which implies $\mathop{\mathrm{E}}_{\boldsymbol{x} \neq \boldsymbol{x}'}\left[\langle f_{\boldsymbol{x}}, f_{\boldsymbol{x}'}\rangle^2\right] = 0$.

Thus, by Theorem 3.1,

$$M \geq 2^{\Omega_h\left(\left[\frac{t}{k}\right]^{1/(h-1)}\right)} = 2^{\Omega_h\left(\left[\frac{\alpha(n+k)}{k}\right]^{1/(h-1)}\right)} = 2^{\Omega_h\left(\left[\frac{n}{k}\right]^{1/(h-1)}\right)}.$$ □

Corollary 3.3. *Let $k \leq n^\alpha$ for some $0 \leq \alpha < 1$, and suppose that* IP *admits an SM protocol (A, B, C) such that $B \colon \{0,1\}^n \to \{0,1\}^{n+k}$ and C is an unbounded fan-in circuit of depth h and size M. Then:*

$$M \geq 2^{\Omega_h\left(n^{\frac{1-\alpha}{h-1}}\right)}.$$

Corollary 3.4. *Let $k \leq \frac{n}{\log^\beta n}$ for every $\beta > 0$ (for large enough n), and suppose that* IP *admits an SM protocol (A, B, C) such that $B \colon \{0,1\}^n \to \{0,1\}^{n+k}$ and C is an unbounded fan-in circuit of depth h and size M. Then:*

$$M \geq 2^{\Omega_h(\log^c n)} \text{ for every } c > 0.$$

We now present an application of our main theorem to cryptography. We show that exponentially secure PRFs (in fact, even *weak* PRFs) are not computable in $\mathsf{AC^0}$, even if one allows an arbitrary sublinear-stretch encoding of the input. This implies a limitation on the power of encoded-input PRFs in $\mathsf{AC^0}$ [BIP+18].

Definition 3.5 (pseudorandom functions)**.** Let $\mathcal{K}$ be a keys domain, and let $F \colon \mathcal{K} \times \{0,1\}^n \to \{0,1\}$ be a family of functions; denote $F_k(x) \triangleq F(k, x)$. For integer m and $\epsilon \in [0,1]$, we say that F is a (strong) (ϵ, m)-*pseudorandom function function family* (shortly (ϵ, m)-*PRF*) if for every (non-uniform) circuit distinguisher D^f of size at most m, the following holds:

$$\left| \Pr_{\boldsymbol{k} \sim \mathcal{K}}\left[D^{F_{\boldsymbol{k}}}(1^n) = 1\right] - \Pr_{\boldsymbol{f}}\left[D^{\boldsymbol{f}}(1^n) = 1\right] \right| \leq \epsilon.$$

If the distinguisher is limited to querying the oracle on random and independent inputs, then we say that F is a *weak* (ϵ, m)-*PRF*.

For simplicity, we will consider the case in which $\mathcal{K} = \{0,1\}^n$ (under the uniform distribution).

We prove the following result in Section 6:

Theorem 3.6 (Lower bound for exponentially secure weak PRFs)**.** *Let $k \leq n^\alpha$ for some $0 \leq \alpha < 1$, and suppose that $F \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a weak $2^{\Omega(n)}$-PRF satisfying the right one-to-one condition, that admits an SM protocol (A, B, C) such that $B \colon \{0,1\}^n \to \{0,1\}^{n+k}$ and C is an unbounded fan-in circuit of depth h and size M. Then:*

$$M \geq 2^{\Omega_h\left(n^{\frac{1-\alpha}{h-1}}\right)}.$$

(Similar results hold for $2^{n^{\Omega(1)}}$-weak PRFs, with slightly worse bounds on M.)

The reason that we require the PRF to satisfy the right one-to-one condition is that the "effective" input size of the PRF could be much smaller than n. For example, imagine a PRF which ignores the right half of its input. A distinguisher would need 2^n random samples to notice this.

The right one-to-one condition is automatically satisfied by strong PRFs: if $f_k(x) = f_k(x')$ for all (or even most) keys k, then it is easy to distinguish f_k from a random function by querying the input function at x, x'.

We prove similar results for a class of functions obtained by applying a "rounding predicate" to inner-product modulo q.

Definition 3.7 (Rounded inner product)**.** For an integer $q \geq 2$ and a set $R \subseteq \{0, 1, \ldots, q-1\}$ we define the (q, R)-*rounded inner product function* $\mathsf{IP}^{[q,R]} \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ by

$$\mathsf{IP}^{[q,R]}(x, y) = \begin{cases} 0 & \sum_{i=1}^n x_i y_i \pmod{q} \in R, \\ 1 & \text{otherwise.} \end{cases}$$

One reason for our interest in this class is that some instances, such as rounded inner product modulo 6, are conjectured to be (weak) pseudorandom functions with near-exponential security [BIP+18]. Under such a conjecture, the desired negative result would follow from our results on (weak) PRFs. However, the results about rounded inner product functions are unconditional, and apply also to instances that are provably not (weak) PRFs.

We prove the following result in Section 7:

Theorem 3.8 (Lower bound for rounded inner product)**.** *Let $q \geq 2$ be even, and let $R \subseteq \{0, 1, \ldots, q-1\}$ such that $|R| = q/2$. Let $k \leq n^{\alpha}$ for some $0 \leq \alpha < 1$, and suppose that $\mathsf{IP}^{[q,R]}$ admits an SM protocol (A, B, C) such that $B\colon \{0,1\}^n \to \{0,1\}^{n+k}$ and C is an unbounded fan-in circuit of depth h and size M. Then:*

$$M \geq 2^{\Omega_h\left(n^{\frac{1-\alpha}{h-1}}\right)}.$$

4 Conditional Limits of Preprocessing and Learning AC^0

We were unable to settle the main IPPP conjecture or prove similar results on the limits of preprocessing for other explicit functions. Moreover, our current techniques seems insufficient. A second-best alternative is to settle such questions under widely believed conjectures from complexity theory or cryptography. While we are also unable to show such a conditional result (and view this as an interesting goal), we can relate this challenge to another intriguing question: learning AC^0 under "simple" distributions.

The work of Linial et al. [LMN89] shows that AC^0 can be learned in quasipolynomial time under the uniform distribution. It is open whether the same holds for PAC learning under arbitrary distributions. The question is still open even when restricted to simple input distributions, such as uniform distributions over linear subspaces of $\mathbb{F}_2^n$, and even if "quasipolynomial" is relaxed to "subexponential." In fact, we are not aware of hardness results that apply to any simple distributions or beyond quasipolynomial time. See [DLS14, Vad17] for weaker conditional hardness results, and [DRG17] for a survey of known learning algorithms for AC^0.

We observe that positive results on learning AC^0 under simple distributions can be used to base hardness of IPPP-style problems on cryptographic assumptions from [BPR12, BIP+18]. Equivalently, cryptographic assumptions imply that either (1) AC^0 cannot be learned under simple distributions in subexponential time, or (2) IPPP-style hardness conjectures are true. While both (1) and (2) seem highly plausible, strong versions of them may turn out to be false. Moreover, to the best of our knowledge, neither (1) nor (2) are known to be implied by standard conjectures in cryptography or complexity theory. A direct proof that either (1) or (2) hold also seems unlikely. For these reasons, we believe that the above connection is meaningful, and can potentially lead to future progress on either IPPP-style questions or learning AC^0 under simple distributions.

4.1 The conjectures

We will show connections between the following types of conjectures:

- Cryptographic assumptions:
  - (C1) Subexponential hardness of Learning With Rounding (LWR) [BPR12]: for some $\epsilon > 0$ and polynomials $p = p(n)$, $q = q(n)$, the function $f_k(x) = \mathsf{Round}(\langle k, x\rangle \pmod{2q})$ is a $2^{\Omega(n^{\epsilon})}$-secure weak PRF, where $k \in \{0, 1, \ldots, p-1\}^n$ and $x \in \{0,1\}^n$. Here, $\mathsf{Round}(y)$ returns 0 or 1 depending on whether y is closer to 0 or to the modulus $2q$.
  - (C2) Subexponential hardness of LWR mod 6 [BIP+18]: for some $\epsilon > 0$, $f_k(x) = \mathsf{Round}(\langle k, x\rangle \pmod 6)$ is a $2^{\Omega(n^{\epsilon})}$-secure weak PRF, where $k, x \in \{0,1\}^n$.
- Hardness of learning conjectures:
  - (L1) AC^0 cannot be learned in subexponential time under all polynomial-time samplable input distributions.
  - (L2) AC^0 cannot be learned in subexponential time under all $\mathbb{F}_2$-linear input distributions.

Here, learning in subexponential time refers to a $2^{n^{o(1)}}$-time learning algorithm in the standard PAC model [Val84].

- IPPP-style conjectures:
  - (P1) Integer-IP does not admit an SM protocol (A, B, C) where the referee C is in AC^0 and the parties A and B are polynomial-time. Here Integer-IP is the (non-boolean) inner product of two n-bit vectors over the integers.
  - (P2) Integer-IP is not in $\mathsf{AC}^0 \circ \mathsf{MOD}_2$.

Similarly, we define $(\mathsf{P1})^m$ and $(\mathsf{P2})^m$ as variants where Integer-IP is replaced by inner product modulo m. Note that $(\mathsf{P1})^2$ is the worst-case variant of Rothblum's IPPP conjecture [Rot12] and $(\mathsf{P2})^2$ is the IPPP with linear preprocessing conjecture made by Servedio and Viola [SV12].

4.2 The connections

We now establish simple connections between the previous conjectures.

Theorem 4.1. *The following implications hold:*

1. $(\mathsf{C1}) \Rightarrow (\mathsf{L1}) \vee (\mathsf{P1})$
2. $(\mathsf{C1}) \Rightarrow (\mathsf{L2}) \vee (\mathsf{P2})$
3. $(\mathsf{C2}) \Rightarrow (\mathsf{L1}) \vee (\mathsf{P1})^2 \vee (\mathsf{P1})^3$
4. $(\mathsf{C2}) \Rightarrow (\mathsf{L2}) \vee (\mathsf{P2})^2 \vee (\mathsf{P2})^3$

Proof. To prove (1), suppose that both (L1) and (P1) are false. We use the SM protocol implied by $\neg$(P1) to convert the learning algorithm implied by $\neg$(L1) into an attack against the LWR assumption in (C1). Let $f(a, b)$ be the Integer-IP function. By $\neg$(P1), there is an SM protocol (A, B, C) for f where C is in AC^0 and the parties A and B are polynomial-time. Letting $f'(k, x)$ be the rounded inner product function defined by polynomials p, q as in (C1), we get a similar SM protocol (A', B', C') for f' in the following natural way: A' expands each $k_i \in \{0, 1, \ldots, p-1\}$ to a binary length-p vector of weight k_i and invokes A; B' expands each $x_i \in \{0, 1\}$ to the length-p vector $(x_i, \ldots, x_i)$ and invokes B; and C' invokes C to compute the integer inner product $\langle k, x\rangle$, reduces the result modulo q, and rounds. Since p and q are polynomials, C' can indeed be implemented in AC^0. Now consider the message $\hat{k}$ sent by A' on a uniformly random input k, and let $C'_{\hat{k}}$ be the AC^0 circuit obtained by restricting C' to this first message. Let X be the (polynomial-time samplable) input distribution defined by the message sent by B' on a uniformly random input x. Using the subexponential time learning algorithm implied by $\neg$(L1) to learn $C'_{\hat{k}}$ on input distribution X, we get a subexponential time algorithm breaking (C1) as required.

The proofs of the other parts of the theorem follow similarly, noting that if neither $(\mathsf{P1})^2$ nor $(\mathsf{P1})^3$ hold (resp., neither $(\mathsf{P2})^2$ nor $(\mathsf{P2})^3$ hold), then f' computing rounded inner product modulo 6 admits an SM protocol with referee in AC^0 and polynomial-time parties (resp., parties computing an $\mathbb{F}_2$-linear function with polynomial stretch). □

5 Proof of Main Theorem

The following two results will be needed for proving the main theorem, Theorem 3.1.

Proposition 5.1 (High-degree spectral concentration bound)**.** *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, and let $0 \le \epsilon \le 1/2$. Then, for every integer $0 \le t \le n$ such that $\|f^{\le t}\| \le \epsilon$, if an unbounded fan-in circuit of depth h and size M agrees with f on at least $1/2 + \epsilon$ fraction of inputs, then*

$$M \ge 2^{\Omega_h\left(\left[\frac{t}{1-2\log\epsilon}\right]^{1/(h-1)}\right)}.$$

Proof. Let $0 \le t \le n$ be an integer such that $\|f^{\le t}\| \le \epsilon$, and suppose that an unbounded fan-in circuit of depth h and size M computes a function F that agrees with f on at least $1/2+\epsilon$ fraction of inputs.

On one hand, we have

$$\langle F, f\rangle = 2\Pr[F=f] - 1 \underset{\text{assumption}}{\ge} 2(1/2+\epsilon) - 1 = 2\epsilon.$$

On the other hand, we have

$$\langle F, f\rangle = \langle F^{\le t}, f^{\le t}\rangle + \langle F^{>t}, f^{>t}\rangle \underset{\text{Cauchy–Schwarz}}{\le} \|f^{\le t}\| + \|F^{>t}\| \underset{\text{LMNT}}{\le} \epsilon + \sqrt{2\cdot 2^{-t/O_h(\log M)^{h-1}}}.$$

Thus,

$$2\epsilon \le \epsilon + \sqrt{2\cdot 2^{-t/O_h(\log M)^{h-1}}} \implies M \ge 2^{\Omega_h\left(\left[\frac{t}{1-2\log\epsilon}\right]^{1/(h-1)}\right)}.$$

□

In what follows, we will use the following notation:

- For a set X, we write $\boldsymbol{i} \ne \boldsymbol{j} \sim X$ to mean that $(\boldsymbol{i}, \boldsymbol{j})$ is chosen uniformly at random from the set $\{(i,j) \in X \times X : i \ne j\}$.
- Given an inner product space V, a subspace $U \le V$, and a vector $v \in V$, we denote the projection of v onto U by $\mathrm{proj}_U(v)$.

Lemma 5.2 (The Projection Lemma)**.** *Let V be an inner product space over $\mathbb{R}$. Let $\{v_i\}_{i\in X} \subseteq V$ be a set of* unit *vectors indexed by X, and suppose that*

$$\mathop{\mathrm{E}}_{\boldsymbol{i}\ne\boldsymbol{j}\sim X}\left[\langle v_{\boldsymbol{i}}, v_{\boldsymbol{j}}\rangle^2\right] \le \frac{1}{36|X|^2}.$$

Then, for every subspace $U \le V$, there exists $i \in X$ such that $\|\mathrm{proj}_U(v_i)\|^2 = O\left(\frac{\dim U}{|X|}\right)$.

Proof. Let $U \le V$ be a subspace, and denote $D \triangleq \dim U$.

By Cauchy–Schwartz,

$$\mathop{\mathrm{E}}_{\boldsymbol{i}\ne\boldsymbol{j}\sim X}\left[|\langle v_{\boldsymbol{i}}, v_{\boldsymbol{j}}\rangle|\right] \le \mathop{\mathrm{E}}_{\boldsymbol{i}\ne\boldsymbol{j}\sim X}\left[\langle v_{\boldsymbol{i}}, v_{\boldsymbol{j}}\rangle^2\right]^{1/2} \le \frac{1}{6|X|}.$$

By Markov's inequality,

$$\Pr_{\boldsymbol{i}\sim X}\left[\mathop{\mathrm{E}}_{\boldsymbol{j}\sim X\setminus\{\boldsymbol{i}\}}\left[\langle v_{\boldsymbol{i}}, v_{\boldsymbol{j}}\rangle^2\right] > \frac{1}{12|X|^2}\right] \le 12|X|^2 \cdot \mathop{\mathrm{E}}_{\boldsymbol{i}\ne\boldsymbol{j}}\left[\langle v_{\boldsymbol{i}}, v_{\boldsymbol{j}}\rangle^2\right] \le \frac{1}{3}.$$

and similarly,

$$\Pr_{\boldsymbol{i}\sim X}\left[\mathop{\mathrm{E}}_{\boldsymbol{j}\sim X\setminus\{\boldsymbol{i}\}}\left[|\langle v_{\boldsymbol{i}}, v_{\boldsymbol{j}}\rangle|\right] > \frac{1}{2|X|}\right] \le 2|X| \cdot \mathop{\mathrm{E}}_{\boldsymbol{i}\ne\boldsymbol{j}}\left[|\langle v_{\boldsymbol{i}}, v_{\boldsymbol{j}}\rangle|\right] \le \frac{1}{3},$$

which implies that at least $1/3$ of the indices $i \in X$ satisfy

$$\mathop{\mathrm{E}}_{\boldsymbol{j}\sim X\setminus\{i\}}\left[\langle v_i, v_{\boldsymbol{j}}\rangle^2\right] \le \frac{1}{12|X|^2} \quad \text{and} \quad \mathop{\mathrm{E}}_{\boldsymbol{j}\sim X\setminus\{i\}}\left[|\langle v_i, v_{\boldsymbol{j}}\rangle|\right] \le \frac{1}{2|X|},$$

or equivalently,

$$\sum_{j\in X\setminus\{i\}} \langle v_i, v_j\rangle^2 \le \frac{1}{12|X|} \quad \text{and} \quad \sum_{j\in X\setminus\{i\}} |\langle v_i, v_j\rangle| \le \frac{1}{2}. \tag{1}$$

Put these indices in a set Y, and let $W \triangleq \mathrm{span}(\{v_i : i \in Y\})$.

Let $w \in W$ such that $\|w\| \leq 1$, and write $w = \sum_{i \in Y} c_i v_i$ with $c_i \in \mathbb{R}$. Then for $i \in Y$,

$$\langle w, v_i \rangle = \sum_{j \in Y} c_j \langle v_i, v_j \rangle = c_i + \sum_{j \in Y \setminus \{i\}} c_j \langle v_i, v_j \rangle.$$

Multiply this by c_i, and sum over all $i \in Y$ to obtain

$$1 \geq \|w\|^2 = \sum_{i \in Y} c_i^2 + \sum_{i \neq j} c_i c_j \langle v_i, v_j \rangle.$$

Since $2|c_i c_j| \leq c_i^2 + c_j^2$, it follows that

$$1 \geq \sum_{i \in Y} c_i^2 - \frac{1}{2} \sum_{i \neq j} (c_i^2 + c_j^2) |\langle v_i, v_j \rangle| = \sum_{i \in Y} c_i^2 \Big(1 - \sum_{j \in Y \setminus \{i\}} |\langle v_i, v_j \rangle|\Big) \underset{\text{Eq. (1)}}{\geq} \frac{1}{2} \sum_{i \in Y} c_i^2,$$

which implies $\sum_{i \in Y} c_i^2 \leq 2$.[1] Since $(a+b)^2 \leq 2a^2 + ab^2$, for every $i \in Y$ we have

$$\begin{aligned}
\langle w, v_i \rangle^2 &\leq 2c_i^2 + 2\Big(\sum_{j \in Y \setminus \{i\}} c_j \langle v_i, v_j \rangle\Big)^2 \underset{\text{Cauchy–Schwarz}}{\leq} 2c_i^2 + 2 \sum_{j \in Y \setminus \{i\}} c_j^2 \cdot \sum_{j \in Y \setminus \{i\}} \langle v_i, v_j \rangle^2 \\
&\leq 2c_i^2 + 4 \sum_{j \in Y \setminus \{i\}} \langle v_i, v_j \rangle^2 \underset{\text{Eq. (1)}}{\leq} 2c_i^2 + \frac{1}{3|X|}.
\end{aligned}$$

Taking expectation over $i \in Y$, we deduce

$$\begin{aligned}
\mathop{\mathrm{E}}_{\boldsymbol{i} \sim Y} \Big[\langle w, v_i \rangle^2\Big] &\leq \mathop{\mathrm{E}}_{\boldsymbol{i} \sim Y} \left[2c_i^2 + \frac{1}{3|X|}\right] = \frac{2}{|Y|} \sum_{i \in Y} c_i^2 + \frac{1}{3|X|} \leq \frac{4}{|Y|} + \frac{1}{3|X|} \\
&\underset{|Y| \geq |X|/3}{\leq} \frac{12}{|X|} + \frac{1}{3|X|} \leq \frac{13}{|X|}.
\end{aligned}$$

Now let $u_1, \ldots, u_D$ be an orthonormal basis for U, and for every $k \in [D]$, let $w_k \in W$ be the projection of u_k onto W (notice that $\|w_k\| \leq 1$). We have

$$\begin{aligned}
\mathop{\mathrm{E}}_{\boldsymbol{i} \sim Y} \Big[\|\mathrm{proj}_U(v_i)\|^2\Big] &= \mathop{\mathrm{E}}_{\boldsymbol{i} \sim Y} \left[\sum_{k \in [D]} \langle v_i, u_k \rangle^2\right] = \mathop{\mathrm{E}}_{\boldsymbol{i} \sim Y} \left[\sum_{k \in [D]} \langle v_i, w_k \rangle^2\right] \\
&= \sum_{k \in [D]} \mathop{\mathrm{E}}_{\boldsymbol{i} \sim Y} \Big[\langle v_i, w_k \rangle^2\Big] \leq \frac{13D}{|X|},
\end{aligned}$$

which implies there exists $i \in Y$ such that $\|\mathrm{proj}_U(v_i)\|^2 \leq \frac{13D}{|X|} = O\left(\frac{D}{|X|}\right)$, as desired. □

We can now prove our main theorem.

Proof of Theorem 3.1. The proof follows several steps.

<u>STEP 1</u>: Since f satisfies the left one-to-one condition, by Proposition 2.4, B computes a one-to-one mapping; hence, we can extend it to a permutation $\tau \colon \{-1,1\}^{n+k} \to \{-1,1\}^{n+k}$ as follows:

$$\tau(y_1, \ldots, y_{n+k}) = \begin{cases} B(y_1, \ldots, y_n) & \text{if } y_{n+1} = \cdots = y_{n+k} = 1, \\ \text{arbitrary choice} & \text{otherwise,} \end{cases}$$

where by arbitrary choice we mean one of the $(2^{n+k} - 2^n)!$ possible ways of completing the definition so as to yield a permutation. Define $\sigma = \tau^{-1}$ and note that σ is a permutation as well.

[1] Note that the argument implies that the vectors $\{v_i\}_{i \in Y}$ are linearly independent; otherwise, we can find representations of w for which $\sum_{i \in Y} c_i^2$ is arbitrary large.

STEP 2: Let $S\colon \{-1,1\}^n \to 2^{[n]}$ be a bijection (e.g., map a vector to the set it characterizes). For every $x \in \{-1,1\}^n$ and $R \subseteq \{n+1,\ldots,n+k\}$, define $f_x^R\colon \{-1,1\}^{n+k} \to \{-1,1\}^{n+k}$ by

$$f_x^R(y_1,\ldots,y_{n+k}) = \begin{cases} f_x(y_1,\ldots,y_n) & \text{if } y_{n+1} = \cdots = y_{n+k} = 1, \\ \chi_{S(x)}(y_1,\ldots,y_n)\cdot \chi_R(y_{n+1},\ldots,y_{n+k}) & \text{otherwise.} \end{cases}$$

What can we say about these functions?

- Fix $x \in \{-1,1\}^n$, and denote by C_x the circuit obtained from C when Alice is given x as input. Now, consider $y \in \{-1,1\}^{n+k}$.
  - If $y_{n+1} = \cdots = y_{n+k} = 1$, then $f_x^R(y) = f_x(y)$ by definition; hence, by the correctness of $\mathcal{P}$ and the definition of σ, we have that f_x^R agrees with $C_x \circ \sigma^{-1}$ on all such y's.
  - Otherwise, let $i \in \{n+1,\ldots,n+k\}$ such that $y_i = -1$. For every $R \subseteq \{n+1,\ldots,n+k\}$ that contains i, we have that $C_x \circ \sigma^{-1}$ agrees with exactly one of $f_x^R, f_x^{R\setminus\{i\}}$ on the input $(y_1,\ldots,y_{n+i-1},-1,y_{n+i+1},\ldots,y_{n+k})$; thus, for exactly half the subsets $R \subseteq \{n+1,\ldots,n+k\}$, f_x^R agrees with $C_x \circ \sigma^{-1}$ on y. Therefore,

    $$\Pr_{\boldsymbol{R}\sim 2^{[n+k]\setminus[n]}}\left[f_x^{\boldsymbol{R}}(y) = C_x(\sigma^{-1}(y))\right] = \frac{1}{2}.$$

    This holds for any y such that $(y_{n+1},\ldots,y_{n+k}) \neq (1,\ldots,1)$; hence,

    $$\mathop{\mathrm{E}}_{\substack{\boldsymbol{y}\sim\{-1,1\}^{n+k}\\ \exists j\in[k]\colon \boldsymbol{y}_{n+j}=-1}}\left[\Pr_{\boldsymbol{R}\sim 2^{[n+k]\setminus[n]}}[f_x^{\boldsymbol{R}}(\boldsymbol{y}) = C_x(\sigma^{-1}(\boldsymbol{y}))]\right] = \frac{1}{2},$$

    which implies there exists $R(x) \subseteq [n+k]\setminus[n]$ such that

    $$\Pr_{\substack{\boldsymbol{y}\sim\{-1,1\}^{n+k}\\ \exists j\in[k]\colon \boldsymbol{y}_{n+j}=-1}}\left[f_x^{R(x)}(\boldsymbol{y}) = C_x(\sigma^{-1}(\boldsymbol{y}))\right] \geq \frac{1}{2},$$

  It follows that the fraction of inputs on which $f_x^{R(x)}$ and $C_x \circ \sigma^{-1}$ agree is at least

  $$\frac{2^n + (1/2)\cdot\left(2^{n+k} - 2^n\right)}{2^{n+k}} = \frac{1}{2} + \frac{1}{2^{k+1}},$$

  which is also the fraction of inputs on which $F_x^{R(x)} \triangleq f_x^{R(x)} \circ \sigma$ and C_x agree.

- The second thing we observe is that for any $x \neq x' \in X$,

  $$\begin{aligned}\left\langle F_x^{R(x)}, F_{x'}^{R(x')}\right\rangle &= \left\langle f_x^{R(x)}\circ\sigma, f_{x'}^{R(x')}\circ\sigma\right\rangle = \left\langle f_x^{R(x)}, f_{x'}^{R(x')}\right\rangle \\ &= \mathop{\mathrm{E}}_{\boldsymbol{y}\sim\{-1,1\}^{n+k}}\left[f_x^{R(x)}(\boldsymbol{y})\cdot f_{x'}^{R(x')}(\boldsymbol{y})\right] \\ &= \mathop{\mathrm{E}}_{\boldsymbol{y}\sim\{-1,1\}^{n}}\left[f_x(\boldsymbol{y})\cdot f_{x'}(\boldsymbol{y})\right]\cdot 2^{-k} \\ &\quad + \sum_{\substack{z\in\{-1,1\}^k\\ \exists j\in[k]\colon z_j=-1}} \chi_{R(x)}(z)\cdot\chi_{R(x')}(z)\cdot \mathop{\mathrm{E}}_{\boldsymbol{y}\sim\{-1,1\}^{n}}\left[\chi_{S(x)}(\boldsymbol{y})\cdot\chi_{S(x')}(\boldsymbol{y})\right]\cdot 2^{-k} \\ &= \langle f_x, f_{x'}\rangle\cdot 2^{-k} + \sum_{\substack{z\in\{-1,1\}^k\\ \exists j\in[k]\colon z_j=-1}} \chi_{R(x)}(z)\cdot\chi_{R(x')}(z)\cdot \underbrace{\langle\chi_{S(x)},\chi_{S(x')}\rangle}_{0}\cdot 2^{-k} \\ &= \langle f_x, f_{x'}\rangle\cdot 2^{-k},\end{aligned}$$

  which implies

  $$\mathop{\mathrm{E}}_{\boldsymbol{x}\neq\boldsymbol{x}'\sim X}\left[\left\langle F_{\boldsymbol{x}}^{R(\boldsymbol{x})}, F_{\boldsymbol{x}'}^{R(\boldsymbol{x}')}\right\rangle^2\right] \leq 2^{-2k}\cdot \mathop{\mathrm{E}}_{\boldsymbol{x}\neq\boldsymbol{x}'\sim X}\left[\langle f_x, f_{x'}\rangle^2\right] \underset{\text{assumption}}{\leq} 2^{-2k}\cdot\frac{2^{2k}}{36|X|^2} = \frac{1}{36|X|^2}.$$

STEP 3: Let V be the inner product space of all functions $\{-1,1\}^{n+k} \to \mathbb{R}$, and let $U \leq V$ be the subspace of all functions of degree up to t, which is spanned by $\{\chi_T\}_{|T|\leq t}$ and has dimension $\dim U = \binom{n+k}{\leq t}$. By the Projection Lemma, there exists $x^* \in X$ such that

$$\left\| F_{x^*}^{R(x^*)\leq t} \right\|^2 \leq \frac{13\binom{n+k}{\leq t}}{|X|} \underset{\text{assumption}}{\leq} \frac{1}{2^{2(k+1)}} \implies \left\| F_{x^*}^{R(x^*)\leq t} \right\| \leq \frac{1}{2^{k+1}}.$$

Since $\left\| F_{x^*}^{R(x^*)\leq t} \right\| \leq \frac{1}{2^{k+1}}$ and C_{x^*} is an unbounded fan-in circuit of depth h and size $\leq M$ that agrees with $F_{x^*}^{R(x^*)}$ on at least $\frac{1}{2} + \frac{1}{2^{k+1}}$ fraction of inputs, by Proposition 5.1,

$$M \geq 2^{\Omega_h\left(\left[\frac{t}{1-2\log\left(2^{-(k+1)}\right)}\right]^{1/(h-1)}\right)} = 2^{\Omega_h\left(\left[\frac{t}{2k+3}\right]^{1/(h-1)}\right)} = 2^{\Omega_h\left(\left[\frac{t}{k}\right]^{1/(h-1)}\right)}. \qquad \square$$

6 Encoded-input pseudorandom functions

The goal of this section is to prove Theorem 3.6, which shows that weak PRFs are hard for our model.

Proposition 6.1 (Expected inner product bound for weak PRFs)**.** *Let $\delta \in (0,1]$, and suppose that $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a weak $(m, \frac{1}{m})$-PRF for $m = \Omega\big((1/\delta)^2 \ln(4/\delta)\cdot n\big)$. Then:*

$$\mathop{\mathrm{E}}_{\boldsymbol{k},\boldsymbol{k'}\sim\mathcal{K}}\left[\langle F_{\boldsymbol{k}}, F_{\boldsymbol{k'}}\rangle^2\right] \leq 4\delta.$$

Proof. We switch notation to $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$. The proof follows a hybrid argument.

Consider the following algorithm $M(f,g)$ which is given access to a pair of functions f,g and operates as follows:

1. M chooses uniformly and independently $N = 32(1/\delta)^2\ln(4/\delta)$ random inputs $\vec{\boldsymbol{x}} = (\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(N)})$.
2. M estimates $\langle f,g\rangle$ with the following estimator:
$$\hat{\boldsymbol{\theta}} = \frac{1}{N}\sum_{i\in[N]} f(\boldsymbol{x}^{(i)})g(\boldsymbol{x}^{(i)}).$$
3. M outputs 1 if $\hat{\boldsymbol{\theta}}^2 > \delta/2$, and 0 otherwise.

Suppose that f,g are randomly chosen. Denote $\boldsymbol{Z}_i = \boldsymbol{f}(\boldsymbol{x}^{(i)})\boldsymbol{g}(\boldsymbol{x}^{(i)})$ for every $i \in [N]$, and $\boldsymbol{Z} = \sum_{i\in[N]}\boldsymbol{Z}_i$. We have $\mathrm{E}_{\boldsymbol{f},\boldsymbol{g}}[\boldsymbol{Z}_i] = 0$ for every $i \in [N]$, implying $\mathrm{E}_{\boldsymbol{f},\boldsymbol{g}}[\boldsymbol{Z}] = 0$.[2] Thus,

$$\Pr_{\boldsymbol{f},\boldsymbol{g}}[M(\boldsymbol{f},\boldsymbol{g})=1] = \Pr_{\boldsymbol{f},\boldsymbol{g}}\left[|\hat{\boldsymbol{\theta}}| > \sqrt{\delta/2}\right] = \Pr\left[|\boldsymbol{Z}-\mathrm{E}[\boldsymbol{Z}]| > N\sqrt{\delta/2}\right] \underset{\text{Hoeffding}}{\leq} 2e^{-N\delta/4}.$$

To establish the hybrid argument, we define two distinguishers:

- Algorithm $A^f(1^n)$: runs $M(f,\boldsymbol{g})$, where $\boldsymbol{g}$ is chosen uniformly at random by A. This means that whenever M wishes to access $\boldsymbol{g}$, A chooses a random answer and passes it to M; to be consistent, A records past answers.
- Algorithm $B^g(1^n)$: runs $M(F_{\boldsymbol{k'}},g)$, where $\boldsymbol{k'}$ is chosen uniformly at random by B. This means that B draws $\boldsymbol{k'}$ once at the beginning, and that F is accessible.

[2]To ease notation, we shall omit references to $\vec{\boldsymbol{x}}$ when writing probabilities and expectations, yet we should keep in mind that these are taken with respect to the random choice of $\vec{\boldsymbol{x}}$ as well.

Observe that $\Pr_{\boldsymbol{k}}[A^{F_{\boldsymbol{k}}}(1^n)=1] = \Pr_{\boldsymbol{g}}[B^{\boldsymbol{g}}(1^n)=1]$. Thus,

$$\begin{aligned}
&\left|\Pr_{\boldsymbol{f},\boldsymbol{g}}[M(\boldsymbol{f},\boldsymbol{g})=1] - \Pr_{\boldsymbol{k},\boldsymbol{k}'}[M(F_{\boldsymbol{k}},F_{\boldsymbol{k}'})=1]\right| \\
&= \left|\Pr_{\boldsymbol{f}}[A^{\boldsymbol{f}}(1^n)=1] - \Pr_{\boldsymbol{k}}[B^{F_{\boldsymbol{k}}}(1^n)=1]\right| \\
&\le \left|\Pr_{\boldsymbol{f}}[A^{\boldsymbol{f}}(1^n)=1] - \Pr_{\boldsymbol{k}}[A^{F_{\boldsymbol{k}}}(1^n)=1]\right| + \left|\Pr_{\boldsymbol{k}}[B^{F_{\boldsymbol{k}}}(1^n)=1] - \Pr_{\boldsymbol{g}}[B^{\boldsymbol{g}}(1^n)=1]\right|.
\end{aligned}$$

Both A^f and B^g require circuits of size $m = O(Nn) = O\big((1/\delta)^2 \ln(4/\delta)\cdot n\big)$. Thus, by definition,

$$\left|\Pr_{\boldsymbol{f},\boldsymbol{g}}[M(\boldsymbol{f},\boldsymbol{g})=1] - \Pr_{\boldsymbol{k},\boldsymbol{k}'}[M(F_{\boldsymbol{k}},F_{\boldsymbol{k}'})=1]\right| \le \tfrac{2}{m} \le \tfrac{2}{N},$$

which implies

$$\Pr_{\boldsymbol{k},\boldsymbol{k}'}[M(F_{\boldsymbol{k}},F_{\boldsymbol{k}'})=1] \le \Pr_{\boldsymbol{f},\boldsymbol{g}}[M(\boldsymbol{f},\boldsymbol{g})=1] + \tfrac{2}{N} \le 2e^{-N\delta/4} + \tfrac{2}{N}.$$

Consider now running $M(F_{\boldsymbol{k}},F_{\boldsymbol{k}'})$ with $\boldsymbol{k},\boldsymbol{k}'$ chosen uniformly at random.

- By the analysis above: $\Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\hat{\boldsymbol{\theta}}^2 > \delta/2\right] \le 2e^{-N\delta/4} + \frac{2}{N}$.
- Applying Hoeffding's inequality once more,

$$\begin{aligned}
\Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\left|\hat{\boldsymbol{\theta}}^2 - \langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2\right| > \delta/2\right] &= \Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\left|\hat{\boldsymbol{\theta}} - \langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle\right| > \frac{\delta/2}{\left|\hat{\boldsymbol{\theta}} + \langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle\right|}\right] \\
&\le \Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\left|\hat{\boldsymbol{\theta}} - \langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle\right| > \delta/4\right] \\
&\le 2e^{-N\delta^2/32}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 > \delta\right] &= \Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 - \hat{\boldsymbol{\theta}}^2 + \hat{\boldsymbol{\theta}}^2 > \delta\right] \\
&\le \Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 - \hat{\boldsymbol{\theta}}^2 > \delta/2\right] + \Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\hat{\boldsymbol{\theta}}^2 > \delta/2\right] \\
&\le 2e^{-N\delta^2/32} + 2e^{-N\delta/4} + \tfrac{2}{N} \\
&\underset{\delta\in(0,1]}{\le} 4e^{-N\delta^2/32} + \tfrac{2}{N}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathop{\mathrm{E}}_{\boldsymbol{k},\boldsymbol{k}'}\left[\langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2\right] &= \mathop{\mathrm{E}}_{\boldsymbol{k},\boldsymbol{k}'}\left[\langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 \,\middle|\, \langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 > \delta\right] \cdot \Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 > \delta\right] \\
&\quad + \mathop{\mathrm{E}}_{\boldsymbol{k},\boldsymbol{k}'}\left[\langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 \,\middle|\, \langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 \le \delta\right] \cdot \Pr_{\boldsymbol{k},\boldsymbol{k}'}\left[\langle F_{\boldsymbol{k}},F_{\boldsymbol{k}'}\rangle^2 \le \delta\right] \\
&\le 1 \cdot (4e^{-N\delta^2/32} + \tfrac{2}{N}) + \delta \cdot 1 \\
&= \delta + 4e^{-N\delta^2/32} + \tfrac{2}{N} \\
&\le 2\delta + 2\delta = 4\delta,
\end{aligned}$$

the last inequality holding since $N = 32(1/\delta)^2 \ln(4/\delta) \ge 1/\delta$. □

Proposition 6.2 (General lower bound for weak PRFs)**.** *There exist a constant $0 < a \le 1/2$ such that for every $0 \le r \le n/2 - 1$ and integer $0 \le t \le a(n+r)$ the following holds: If $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ satisfies the right one-to-one condition and is a weak $(m, \frac{1}{m})$-PRF for*

$$m = \Omega\left(n \cdot 2^{4r} \cdot \binom{n+r}{t}^4 \left[r + \log\binom{n+r}{t}\right]\right),$$

and F admits an SM protocol (A, B, C) such that $B\colon \{0,1\}^n \to \{0,1\}^{n+r}$ and C is an unbounded fan-in circuit of depth h and size M, then

$$M \ge 2^{\Omega_h\left(\left[\frac{t}{r}\right]^{1/(h-1)}\right)}.$$

Proof. Let us define:

$$s \triangleq 13 \cdot 2^{2(r+1)} \cdot \binom{n+r}{t} \quad , \quad \delta \triangleq \frac{2^{2r-2}}{36s^2} = \frac{1}{36 \cdot 169 \cdot 2^{2r+6} \cdot \binom{n+r}{t}^2}.$$

We have:

$$(1/\delta)^2 \ln(1/\delta) = \Theta\left(2^{4r} \cdot \binom{n+r}{t}^4 \left[r + \log\binom{n+r}{t}\right]\right).$$

Thus, assuming $m = \Omega\big((1/\delta)^2 \ln(4/\delta) \cdot n\big)$, by assumption and Proposition 6.1, we get

$$\mathop{\mathrm{E}}_{\boldsymbol{k} \neq \boldsymbol{k}' \sim \{0,1\}^n}\left[\langle F_{\boldsymbol{k}}, F_{\boldsymbol{k}'}\rangle^2\right] \le 4\delta = \frac{2^{2r}}{36s^2}.$$

In particular, there exists a set $X \subseteq \{0,1\}^n$ of size $|X| = s$ such that

$$\mathop{\mathrm{E}}_{\boldsymbol{k} \neq \boldsymbol{k}' \sim X}\left[\langle F_{\boldsymbol{k}}, F_{\boldsymbol{k}'}\rangle^2\right] \le \frac{2^{2r}}{36s^2}.$$

We need to justify why $s \le 2^n$. As shown in the proof of Proposition 3.2, there exists $0 < a \le 1/2$ such that

$$13 \cdot 2^{2(r+1)} \cdot \binom{n+r}{\le a(n+r)} \le 2^n;$$

hence, for $t = a(n+r)$ we have $s \le 2^n$.[3] Thus, by Theorem 3.1,

$$M \ge 2^{\Omega_h\left(\left[\frac{t}{r}\right]^{1/(h-1)}\right)}. \qquad \square$$

Theorem 3.6 is an immediate corollary.

7 Rounded inner product

In this section we prove Theorem 3.8, an IPPP-style theorem (with sublinear stretch) for a class of functions obtained by applying a "rounding predicate" to an inner product modulo q. We remind the reader that these functions are given in Definition 3.7.

The following proposition will be useful.

[3] Note that for $a \le 1/2$, the function $t \mapsto \binom{n+r}{\le t}$ is monotone on $[0, a(n+r)]$.

Proposition 7.1 (Inner product convergence)**.** *Let $q \geq 2$ be an integer. Then, for every $r \in \{0, \ldots, q-1\}$,*

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\{0,1\}^{2n}}\left[\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i \pmod q = r\right] \xrightarrow[n\to\infty]{} \frac{1}{q}.$$

Moreover, there exists $0 < c < 1$ such that for every $r \in \{0, \ldots, q-1\}$, for large enough n,

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\{0,1\}^{2n}}\left[\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i \pmod q = r\right] = \frac{1}{q} \pm O(c^n).$$

Proof. For the finite state space $Q = \{0, \ldots, q-1\}$ of remainders modulo q, we define a sequence of random variables $\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ by

$$\boldsymbol{Z}_i = \begin{cases} 0 & i = 0, \\ \boldsymbol{Z}_{i-1} + \boldsymbol{x}_i \boldsymbol{y}_i \pmod q & i \in [n], \end{cases}$$

where $(\boldsymbol{x}, \boldsymbol{y}) \sim \{0,1\}^{2n}$. We are interested in $\lim_{n\to\infty} \Pr[\boldsymbol{Z}_n = r \mid \boldsymbol{Z}_0 = 0]$. For every $i \in [n]$, we have

$$\Pr[\boldsymbol{Z}_i \mid \boldsymbol{Z}_{i-1}] = \Pr[\boldsymbol{Z}_i \mid \boldsymbol{Z}_{i-1}, \ldots, \boldsymbol{Z}_0],$$

and for every $i \in [n]$ and $u \in Q$, we have

$$\Pr[\boldsymbol{Z}_i = u \mid \boldsymbol{Z}_{i-1} = u] = 3/4,$$
$$\Pr[\boldsymbol{Z}_i = u + 1 \pmod q \mid \boldsymbol{Z}_{i-1} = u] = 1/4.$$

Thus, the sequence $(\boldsymbol{Z}_i)$ forms a Markov chain, which we claim is *ergodic*. To see that, consider a walk of q steps; then, for every $u, v \in Q$, we have $\Pr[\boldsymbol{Z}_{i+q} = u \mid \boldsymbol{Z}_i = v] \geq (\frac{1}{4})^q > 0$. Since $(\boldsymbol{Z}_i)$ is a finite ergodic Markov chain, it follows that there exists a unique stationary distribution π over Q such that for every $r \in Q$, we have $\lim_{n\to\infty} \Pr[\boldsymbol{Z}_n = r \mid \boldsymbol{Z}_0 = 0] = \pi(r)$. It is easy to verify that $\pi^* = (\frac{1}{q}, \ldots, \frac{1}{q})$ is a distribution over Q which satisfies $\pi^* = \pi^* P$, where P is the transition matrix of the Markov chain, given by

$$P = \begin{bmatrix} 3/4 & 1/4 & 0 & \ldots & 0 \\ 0 & 3/4 & 1/4 & \ldots & 0 \\ 0 & 0 & 3/4 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/4 & 0 & 0 & \ldots & 3/4 \end{bmatrix}.$$

It follows that $\pi = \pi^*$, hence

$$\lim_{n\to\infty} \Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\{0,1\}^{2n}}\left[\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i \pmod q = r\right] = \lim_{n\to\infty} \Pr[\boldsymbol{Z}_n = r \mid \boldsymbol{Z}_0 = 0] = \frac{1}{q}.$$

The second part of the claim follows from known properties of convergence to a stationary distribution. $\square$

We are now ready to prove the lower bound for computing rounded inner products in our setting.

Proof of Theorem 3.8. Let $y, z \in \{0,1\}^n$ be such that the Hamming distance between y and z is at least $n/3$, and let S_y and S_z be the subsets of $[n]$ characterized by y and z, respectively. Without loss of generality, we may assume that $|S_y \setminus S_z| \geq n/6$, and let us denote $J \triangleq S_y \setminus S_z$.

For $x \in \{0,1\}^n$, let us write $x = (u, v)$ with $u \in \{0,1\}^J$ and $v \in \{0,1\}^{[n]\setminus J}$. Fix a v now. Define

$$a_v \triangleq \sum_{i\in[n]\setminus J} v_i y_i \quad , \quad b_v \triangleq \sum_{i\in[n]\setminus J} v_i z_i,$$

and observe that a_v, b_v are also fixed. We have

$$\sum_{i=1}^{n} x_i y_i \pmod q = \left(\sum_{i\in J} u_i + \sum_{i\in[n]\setminus J} v_i y_i\right) \pmod q = \left(\sum_{i\in J} u_i + a_v\right) \pmod q,$$
$$\sum_{i=1}^{n} x_i z_i \pmod q = \sum_{i\in[n]\setminus J} v_i z_i \pmod q = b_v \pmod q.$$

It follows that there exists a subset $R_v \subseteq \{0,1,\dots,q-1\}$ of size $q/2$ such that

$$\mathsf{IP}^{[q,R]}((u,v),y) = \mathsf{IP}^{[q,R]}((u,v),z) \iff \sum_{i\in J} u_i \pmod q \in R_v.$$

Therefore, by Proposition 7.1, for any fixed v and large enough n,

$$\Pr_{\boldsymbol{u}\sim\{0,1\}^J}\left[\mathsf{IP}^{[q,R]}((\boldsymbol{u},v),y) = \mathsf{IP}^{[q,R]}((\boldsymbol{u},v),z)\right] = \Pr_{\boldsymbol{u}\sim\{0,1\}^J}\left[\sum_{i\in J} \boldsymbol{u}_i \pmod q \in R_v\right]$$
$$= |R_v| \cdot \left(\frac{1}{q} \pm O(c^n)\right) = \frac{1}{2} \pm O(c^n),$$

which implies

$$\Pr_{\boldsymbol{x}}\left[\mathsf{IP}^{[q,R]}(\boldsymbol{x},y) = \mathsf{IP}^{[q,R]}(\boldsymbol{x},z)\right] = \mathop{\mathrm{E}}_{\boldsymbol{v}}\left[\Pr_{\boldsymbol{u}}\left[\mathsf{IP}^{[q,R]}((\boldsymbol{u},v),y) = \mathsf{IP}^{[q,R]}((\boldsymbol{u},v),z)\right]\right] = \frac{1}{2} \pm O(c^n).$$

Considering $\mathsf{IP}^{[q,R]}(x,y)$ and $\mathsf{IP}^{[q,R]}(x,z)$ as functions of x, and switching to $\{0,1\}^n \to \{-1,1\}$ notation, we get

$$\left\langle \mathsf{IP}^{[q,R]}(\cdot,y), \mathsf{IP}^{[q,R]}(\cdot,z)\right\rangle = 2\Pr_{\boldsymbol{x}}\left[\mathsf{IP}^{[q,R]}(\boldsymbol{x},y) = \mathsf{IP}^{[q,R]}(\boldsymbol{x},z)\right] - 1 = \pm O(c^n).$$

Finally, we have:

- $\mathsf{IP}^{[q,R]}$ satisfies the right one-to-one condition.
- The Gilbert–Varshamov bound [Gil52, Var57] tells us there exists $\mathcal{C} \subseteq \{0,1\}^n$ of size $2^{\Omega(n)}$ and minimal Hamming distance $n/3$. By the analysis above, there exists a constant $K > 0$ such that (for large enough n) $|\langle f_x, f_{x'}\rangle| \le Kc^n$ for every $x \ne x' \in \mathcal{C}$. Define $s = \min\left\{|\mathcal{C}|, \frac{1}{6K}2^{\log(1/c)n}\right\}$, and let $0 < \alpha \le 1/2$ be such that $\mathrm{H}(\alpha)$ is small enough so setting $t = \alpha(n+k)$ gives us

  $$13 \cdot 2^{2(k+1)} \cdot \binom{n+k}{\le \alpha(n+k)} \underset{\text{Lemma 2.6}}{\le} 2^{\mathrm{H}(\alpha)(n+k)+2(k+1)+4} \le s.$$

  It follows that any set $X \subseteq \mathcal{C}$ of size s satisfies both $13 \cdot 2^{2(k+1)} \cdot \binom{n+k}{\le\alpha(n+k)} \le |X|$ and

  $$|X| \le \frac{1}{6K} 2^{\log(1/c)n} \implies Kc^n \le \frac{1}{6|X|},$$

  which implies

  $$\mathop{\mathrm{E}}_{\boldsymbol{x}\ne\boldsymbol{x}'\sim X}\left[\langle f_{\boldsymbol{x}}, f_{\boldsymbol{x}'}\rangle^2\right] \le K^2 c^{2n} \le \frac{1}{36|X|^2} \le \frac{2^{2k}}{36|X|^2}.$$

Thus, by Theorem 3.1,

$$M \ge 2^{\Omega_h\left(\left[\frac{t}{k}\right]^{1/(h-1)}\right)} = 2^{\Omega_h\left(n^{\frac{1-\alpha}{h-1}}\right)}. \qquad \square$$

Acknowledgements We thank Andrej Bogdanov, Sajin Koroth, Dinesh Krishnamoorthy, and Srikanth Srinivasan for helpful discussions, comments, and pointers.

References

[ABG$^+$14] Adi Akavia, Andrej Bogdanov, Siyao Guo, Akshay Kamath, and Alon Rosen. Candidate weak pseudorandom functions in $\mathsf{AC}^0 \circ \mathsf{MOD}_2$. In *Innovations in Theoretical Computer Science, ITCS'14, Princeton, NJ, USA, January 12-14, 2014*, pages 251–260, 2014.

[Ajt83] Miklós Ajtai. Σ^1_1-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983.

[AW17] Josh Alman and R. Ryan Williams. Probabilistic rank and matrix rigidity. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 641–652, 2017.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 337–347, 1986.

[BGKL03] László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication complexity of simultaneous messages. *SIAM J. Comput.*, 33(1):137–166, 2003.

[BIP$^+$18] Dan Boneh, Yuval Ishai, Alain Passelègue, Amit Sahai, and David J. Wu. Exploring crypto dark matter: - new simple PRF candidates and their applications. In *Theory of Cryptography - 16th International Conference, TCC 2018, Panaji, India, November 11-14, 2018, Proceedings, Part II*, volume 11240 of *Lecture Notes in Computer Science*, pages 699–729, 2018.

[BIS19] Andrej Bogdanov, Yuval Ishai, and Akshayaram Srinivasan. Unconditionally secure computation against low-complexity leakage. In *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part II*, volume 11693 of *Lecture Notes in Computer Science*, pages 387–416, 2019.

[BPR12] Abhishek Banerjee, Chris Peikert, and Alon Rosen. Pseudorandom functions and lattices. In *Advances in Cryptology - EUROCRYPT 2012 - 31st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cambridge, UK, April 15-19, 2012. Proceedings*, pages 719–737, 2012.

[BR17] Andrej Bogdanov and Alon Rosen. Pseudorandom functions: Three decades later. In *Tutorials on the Foundations of Cryptography*, pages 79–158. Springer, Cham, 2017.

[CGJ$^+$16] Mahdi Cheraghchi, Elena Grigorescu, Brendan Juba, Karl Wimmer, and Ning Xie. $\mathsf{AC}^0 \circ \mathsf{MOD}_2$ lower bounds for the boolean inner product. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 35:1–35:14, 2016.

[CS12] Arkadev Chattopadhyay and Rahul Santhanam. Lower bounds on interactive compressibility by constant-depth circuits. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 619–628, 2012.

[DE17] Zeev Dvir and Benjamin Edelman. Matrix rigidity and the croot-lev-pach lemma. *CoRR*, abs/1708.01646, 2017.

[DI06] Bella Dubrov and Yuval Ishai. On the randomness complexity of efficient sampling. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 711–720, 2006.

[DLS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 441–448, 2014.

[DRG17] Ning Ding, Yanli Ren, and Dawu Gu. PAC learning depth-3 AC^0 circuits of bounded top fanin. In *International Conference on Algorithmic Learning Theory, ALT 2017, 15-17 October 2017, Kyoto University, Kyoto, Japan*, pages 667–680, 2017.

[FSS81] Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. In *22nd Annual Symposium on Foundations of Computer Science, Nashville, Tennessee, USA, 28-30 October 1981*, pages 260–270, 1981.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 33(4):792–807, 1986.

[Gil52] Edgar N. Gilbert. A comparison of signalling alphabets. *The Bell system technical journal*, 31(3):504–522, 1952.

[GPW16] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 86:1–86:15, 2016.

[Gra11] Robert M. Gray. *Entropy and information theory*. Springer Science & Business Media, 2011.

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing, May 28-30, 1986, Berkeley, California, USA*, pages 6–20, 1986.

[LMN89] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. In *30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, North Carolina, USA, 30 October - 1 November 1989*, pages 574–579, 1989.

[PRS88] Pavel Pudlák, Vojtech Rödl, and Petr Savický. Graph complexity. *Acta Inf.*, 25(5):515–535, 1988.

[Raz89] Alexander A. Razborov. On rigid matrices. *preprint*, 1989.

[Rot12] Guy N. Rothblum. How to compute under AC^0 leakage without secure hardware. In *Advances in Cryptology - CRYPTO 2012 - 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2012. Proceedings*, volume 7417 of *Lecture Notes in Computer Science*, pages 552–569, 2012.

[SV12] Rocco A. Servedio and Emanuele Viola. On a special case of rigidity. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:144, 2012.

[Tal16] Avishay Tal. The bipartite formula complexity of inner-product is quadratic. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:181, 2016.

[Tal17] Avishay Tal. Tight bounds on the fourier spectrum of AC^0. In *32nd Computational Complexity Conference, CCC 2017, July 6-9, 2017, Riga, Latvia*, volume 79 of *LIPIcs*, pages 15:1–15:31, 2017.

[Vad17] Salil P. Vadhan. On learning vs. refutation. In *Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017*, pages 1835–1848, 2017.

[Val84] Leslie G. Valiant. A theory of the learnable. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1984, Washington, DC, USA*, pages 436–445, 1984.

[Var57] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, SSSR*, 117:739–741, 1957.

[Vio11] Emanuele Viola. Extractors for circuit sources. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 220–229, 2011.

[Wun12] Henning Wunderlich. On a theorem of razborov. *Computational Complexity*, 21(3):431–477, 2012.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1979, Atlanta, Georgia, USA*, pages 209–213, 1979.