Bounded Indistinguishability for Simple Sources*

Andrej Bogdanov[1,2], Krishnamoorthy Dinesh[3], Yuval Filmus[4], Yuval Ishai[4], Avi Kaplan[4], and Akshayaram Srinivasan[5]

[1]Department of Computer Science and Engineering, The Chinese University of Hong Kong

[2]Institute of Theoretical Computer Science and Communications, The Chinese University of Hong Kong

[3]Computer Science and Engineering department, Indian Institute of Technology Palakkad

[4]The Henry and Marylin Taub Faculty of Computer Science, Technion, Israel

[5]School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India

November 21, 2021

Abstract

A pair of sources $\boldsymbol{X}, \boldsymbol{Y}$ over $\{0,1\}^n$ are *k-indistinguishable* if their projections to any k coordinates are identically distributed. Can some AC^0 function distinguish between two such sources when k is big, say $k = n^{0.1}$? Braverman's theorem (Commun. ACM 2011) implies a negative answer when $\boldsymbol{X}$ is uniform, whereas Bogdanov et al. (Crypto 2016) observe that this is not the case in general.

We initiate a systematic study of this question for natural classes of *low-complexity* sources, including ones that arise in cryptographic applications, obtaining positive results, negative results, and barriers. In particular:

- There exist $\Omega(\sqrt{n})$-indistinguishable $\boldsymbol{X}, \boldsymbol{Y}$, samplable by degree-$O(\log n)$ polynomial maps (over $\mathbb{F}_2$) and by $\mathsf{poly}(n)$-size decision trees, that are $\Omega(1)$-distinguishable by OR.
- There exists a function f such that all $f(d,\epsilon)$-indistinguishable $\boldsymbol{X}, \boldsymbol{Y}$ that are samplable by degree-d polynomial maps are ϵ-indistinguishable by OR for all sufficiently large n. Moreover, $f(1,\epsilon) = \lceil \log(1/\epsilon) \rceil + 1$ and $f(2,\epsilon) = O(\log^{10}(1/\epsilon))$.
- Extending (weaker versions of) the above negative results to AC^0 distinguishers would require settling a conjecture of Servedio and Viola (ECCC 2012). Concretely, if every pair of $n^{0.9}$-indistinguishable $\boldsymbol{X}, \boldsymbol{Y}$ that are samplable by linear maps is ϵ-indistinguishable by AC^0 circuits, then the binary inner product function can have at most an ϵ-correlation with $\mathsf{AC}^0 \circ \oplus$ circuits.

Finally, we motivate the question and our results by presenting applications of positive results to low-complexity secret sharing and applications of negative results to leakage-resilient cryptography.

*This is a full version of [BDF+22].

Contents

1 Introduction

A pair of sources $\boldsymbol{X}, \boldsymbol{Y}$ over $\{0,1\}^n$ are *k-indistinguishable* if their projections to any k coordinates are identically distributed. Can some AC^0 function distinguish between two such sources when k is big, say $k = n^{0.1}$? Braverman's theorem [Bra11, Tal17] implies a negative answer when $\boldsymbol{X}$ is uniform, or equivalently when $\boldsymbol{X}, \boldsymbol{Y}$ are *k-independent*. What about the general case?

The above question was posed by Bogdanov et al. [BIVW16], who observed a tight connection[1] (via LP duality) with the *approximate degree* of the distinguisher. Using this connection, positive answers can be derived from the literature on the approximate degree of AC^0 functions [NS92, Pat92, Shi00, BBC+01, AS04, She13, BT13, BT16, BT18, BT19, BT20a, BT20b, She20]. In particular, there exist $\sqrt{n}$-indistinguishable sources that can be $\Omega(1)$-distinguished by the OR function [NS94] and $n^{1-\delta}$-indistinguishable sources that can be $\Omega(1)$-distinguished by an AC^0 function for every $\delta > 0$ [BT20b]. On the other hand, upper bounds on approximate degree imply limitations on the indistinguishability threshold k. In particular, the $\sqrt{n}$ threshold for OR distinguishers is known to be asymptotically tight, whereas the $n^{1-\delta}$ threshold for AC^0 distinguishers is only conjectured to be tight.

The study of the bounded indistinguishability question in [BIVW16] was motivated by the following "win-win" connection with cryptography. If the answer to the question turns out to be positive, namely there exist k-indistinguishable $\boldsymbol{X}, \boldsymbol{Y}$ that can be distinguished in AC^0, this implies *secret-sharing schemes*[2] where the secret can be reconstructed in AC^0. This is surprising in light of the fact that standard secret-sharing schemes, such as Shamir's scheme [Sha79], use a *linear* function to reconstruct the secret. On the flip side, a negative answer is motivated by the goal of protecting cryptographic applications against leakage of partial information on their internal state. Concretely, in any application that was designed to protect against *local* leakage of k bits, a negative answer implies automatic protection against *global* AC^0 leakage. Such applications abound in the vast literature on secure multiparty computation (MPC), originating from [Yao86, GMW87, BGW88, CCD88], and leakage-resilient circuits, originating from [ISW03]. Braverman's theorem does not apply here because the process of computing on secret-shared data, while respecting k-indistinguishability by design, inevitably creates local dependencies. Obtaining provable resilience to AC^0 leakage turned out to be a challenging task that has led to more intricate constructions and analysis [FRR+14, Rot12, BIS19].

On the downside, both kinds of "win" come with a caveat. In the secret-sharing application, schemes arising from the approximate degree literature minimize reconstruction complexity at the expense of a high *sharing complexity*, of generating the shares. The question of simultaneously minimizing the complexity of sharing and reconstruction remained largely open. For the leakage-resilience application, a general protection even against benign leakage by an OR function (capturing so-called "selective failure" attacks, discussed below) requires $k \gg \sqrt{n}$. Viewing n as the total number of wires in a circuit, existing constructions of leakage-resilient circuits (such as [ISW03]) are far from achieving this k-local secrecy threshold, rendering the generic "security upgrade" guarantee essentially useless in the context of natural applications.

Towards tackling both of the above challenges, we take a more fine-grained view of bounded indistinguishablity, asking the following main question:

> Can some AC^0 function distinguish between *simple* k-indistinguishable sources?

To make the question precise, we need to specify a class $\mathcal{F}$ of samplers that define a "simple" source. We will also consider distinguisher classes $\mathcal{C}$ that are strict subclasses of AC^0, such as depth 1 (OR) or depth 2 (DNF) distinguishers. Given $\mathcal{F}$ and $\mathcal{C}$, the goal is to understand the achievable tradeoff between the threshold k and the distinguishing advantage ϵ.

Braverman's theorem resolves the analogous question for *k-independent* sources. As k-independent sources can be sampled both linearly and locally, the fooling ability of such sources does not depend on

[1] The connection with approximate degree breaks down over non-binary alphabets [BIVW16]. Here we restrict the attention to the binary case, which suffices for our motivating applications.

[2] Here we refer to a relaxation of standard threshold secret sharing that allows for a gap between the secrecy and the reconstruction thresholds and for a small error probability. Bogdanov et al. [BIVW16] present general techniques for narrowing the gap and making the error probability negligible by increasing the share size, while keeping reconstruction in AC^0.

their complexity. In contrast, in this work we demonstrate that the fooling power of k-indistinguishable sources is significantly affected by their complexity.

Useful classes of simple sources. We will be mainly interested in sources that can be sampled by *low-degree* polynomial maps over $\mathbb{F}_2$. Beyond the complexity-theoretic interest in such sources (see, e.g., [Rao09, DGW09, DGRV11, BG13, Li16]), they are also motivated by the two kinds of cryptographic applications discussed above. In the context of secret sharing, positive answers for *degree 1* sources (also referred to as linear or affine sources) would give rise to *linear* secret-sharing schemes with AC^0 reconstruction. Linear schemes have the useful feature of supporting local addition of shared secrets. Perhaps more surprisingly, *degree 2* (quadratic) sources are also naturally motivated by cryptographic applications. We observe that many existing MPC protocols from the literature (including the most efficient ones [DIK10]) can be brought to a form where, for every fixed input, the full transcript is a degree 2 function of the randomness. This holds regardless of the complexity of the function being computed. If for quadratic sources we can get negative answers for much smaller values of k than for general sources, this would enable strong leakage-resilience guarantees for natural applications.

We also consider the minimal *depth* and *locality* required for sampling the sources. A positive result from [BIVW16] shows that OR can distinguish between a pair of k-indistinguishable AC^0-*samplable* sources. However, a direct implementation of this sampler has depth 9. How low can the depth be? Considering *locality*, can AC^0 distinguish between NC^0-samplable sources? Positive answers to the above questions are motivated by the goal of simultaneously minimizing the complexity of sharing and reconstructing secrets.

Useful classes of distinguishers. As random parity-0 and parity-1 strings are $(n-1)$-wise indistinguishable but samplable by essentially the simplest possible closed-under-projection class $\mathcal{F}$ of linear 2-local sources,[3] it is sensible to restrict attention to distinguisher classes $\mathcal{C}$ that cannot compute parities, such as AC^0 or some subclass of it. The simplest subclasses are depth 1 OR distinguishers (disjunction of a subset of the source bits and their negations) and depth 2 DNF distinguishers. Positive results for OR give rise to *visual* secret-sharing schemes [NS94], where the secret can be reconstructed by overlaying transparencies. Negative results for OR and DNF are motivated by securing computations against *selective failure* attacks, where there are multiple events that can trigger failure and only the existence of failure is leaked to the attacker. Beyond this direct motivation, OR leakage comes up naturally in MPC protocols based on garbled circuits [LP07, IKO+11]. DNF leakage can capture stronger selective failure attacks. See [BIVW16, BIS19] for further discussion.

1.1 Overview of results

We now give a detailed account of our main results, for the classes of source samplers $\mathcal{F}$ and distinguishers $\mathcal{C}$ discussed above. The results can be classified into three types: positive (distinguishability), negative (indistinguishability), and barriers. They are summarized in Table 1.

Some of our results merely require that one of the sources $\boldsymbol{X}, \boldsymbol{Y}$ be simple and allow the other to be of arbitrary complexity. For given parameters k, ϵ, we say that

- $\mathcal{F}$ *weakly ϵ-fools* $\mathcal{C}$ if for every k-indistinguishable pair $\boldsymbol{X}, \boldsymbol{Y}$ with $\boldsymbol{X} \in \mathcal{F}$ *and* $\boldsymbol{Y} \in \mathcal{F}$ and every $C \in \mathcal{C}$, $|\Pr[C(\boldsymbol{X}) = 1] - \Pr[C(\boldsymbol{Y}) = 1]| \leq \epsilon$. We refer to this as $\mathsf{MAIN}(k, \epsilon)$.
- $\mathcal{F}$ *strongly ϵ-fools* $\mathcal{C}$ if for every k-indistinguishable pair $\boldsymbol{X}, \boldsymbol{Y}$ with $\boldsymbol{X} \in \mathcal{F}$ *or* $\boldsymbol{Y} \in \mathcal{F}$ and every $C \in \mathcal{C}$, $|\Pr[C(\boldsymbol{X}) = 1] - \Pr[C(\boldsymbol{Y}) = 1]| \leq \epsilon$. We refer to this as $\mathsf{GENERAL}(k, \epsilon)$.

In this terminology, Braverman's theorem states that for $k = \mathsf{polylog}(n)$, the uniform distribution strongly $o(1)$-fools AC^0. We say that $\mathcal{C}$ *distinguishes* $\mathcal{F}$ if $\mathcal{F}$ does not fool $\mathcal{C}$.

[3] The sampler for parity-b strings of length n is $r_1, r_1 \oplus r_2, \ldots, r_{n-2} \oplus r_{n-1}, r_{n-1} \oplus b$.

	Source ($\mathcal{F}$)	Distinguisher ($\mathcal{C}$)	Statement	
			Result	Ref.
Positive	Symmetric, AC^0	OR	$\neg\mathsf{MAIN}(\Theta_\epsilon(\sqrt{n}), 1-\epsilon)$	[BIVW16]
	Mixture of IID, Poly-size decision trees, Degree $O(\log n)$	OR	$\neg\mathsf{MAIN}(\Theta_\epsilon(\sqrt{n}), 1-\epsilon)$	Theorem 5.1
Negative	Linear	$O(1)$-local DNF	$\mathsf{GENERAL}(O(\log \frac{1}{\epsilon}), \epsilon)$	Corollary 6.44
	Degree $O(1)$	OR	$\mathsf{MAIN}(O_\epsilon(1), \epsilon)$	Corollary 6.6
	Quadratic	Unambiguous DNF	$\mathsf{GENERAL}(\mathsf{poly}(\log \frac{n}{\epsilon}), \epsilon)$	Lemma 6.8
	Quadratic	OR	$\mathsf{GENERAL}(\mathsf{poly}(\log \frac{1}{\epsilon}), \epsilon)$	Corollary 6.7
	Depth 1	Arbitrary	$\mathsf{MAIN}(O(\log\log(n/\epsilon)), \epsilon)$	Theorem 6.33
Barrier	Linear	AC^0	$\mathsf{MAIN}(n/\log n, \epsilon) \Rightarrow \mathsf{IPAP}(\epsilon)$	Proposition 4.7
	Linear (LDPC)	AC^0	No NC^0 reduction to k-independence	Claim 7.6
	NC^0	AC^0	$\mathsf{MAIN}(n^{\Omega(1)}, 1/3) \Rightarrow$ Conjecture 7	Claim 8.6

Table 1: Our main results for sources in class $\mathcal{F}$ and distinguishers of type $\mathcal{C}$. A positive result gives a value of k such that there exist $\mathcal{F}$-samplable, k-indistinguishable distributions that are ϵ-distinguished by $\mathcal{C}$. A negative result gives a value of k for which any $\mathcal{F}$-samplable, k-indistinguishable distributions ϵ-fool $\mathcal{C}$. A barrier typically shows that proving a (stronger) negative result would settle a natural conjecture, implying a conditional difficulty to do so. All distinguishers are $\mathsf{poly}(n)$ sized. LDPC refers to uniform distributions over two distinct cosets of a good (linear) low-density parity-check code.

Positive results. In Section 5 we show the existence of an $O_\epsilon(\sqrt{n})$-indistinguishable pair of sources that are $(1-\epsilon)$-distinguishable by OR and samplable by (a) decision trees of size polynomial in n, and (b) polynomials of degree $O(\log n)$ (Theorem 5.1) thereby showing that OR ϵ-distinguishes the sources described in (a) as well as in (b). Part (a) improves on the aforementioned result of Bogdanov et al., by weakening the circuit class from AC^0 to decision trees. Moreover, these sources implement an evolving visual secret sharing scheme [KNY16] of very low informational and computational complexities (see Section 5.5).

Our positive result for degree-$O(\log n)$ sources is obtained by applying a suitable randomized encoding technique [Raz87, Smo87, AIK06] to sources sampled by decision trees. In Section 8 we consider other applications of this technique, showing that a (hypothetical) positive result for $o(\log\log n)$-local sources implies a positive result for 4-local sources. We also put forward a natural conjecture (Conjecture 7) on the complexity of randomized encoding of AC^0 functions that may be viewed as a barrier to negative results.

Negative results. In contrast to Theorem 5.1, we show that constant-degree sources are indistinguishable by OR (see Figure 1):

1. $O(\log(n/\epsilon))$-indistinguishable linear sources strongly ϵ-fool polysize unambiguous DNFs and ORs of $O(1)$-local functions. (Lemma 6.2 + Lemma 6.8)
2. $O(\log^{10}(n/\epsilon))$-indistinguishable quadratic sources strongly ϵ-fool polysize unambiguous DNFs. (Theorem 6.16 + Lemma 6.8)
3. $O_{d,\epsilon}(1)$-indistinguishable degree-d sources weakly ϵ-fool OR. (Corollary 6.15 + Corollary 6.6)

In applications to leakage-resilient cryptography, it is desirable to make the adversary's advantage ϵ a negligible function of the instance size n. The first two negative results allow a low indistinguishability parameter k even when ϵ must vanish exponentially with n. In particular, the first result implies that all linear secret-sharing schemes are automatically immune to selective failure attacks (see [BIVW16, Section 3.3]). The second result implies the same kinds of immunity for efficient MPC protocols, as it turns out that the joint view of the parties in such protocols can be sampled by quadratic polynomial maps (see Section 9.3.2).

As decision trees can be expressed by depth 2 AND/OR formulas (both CNFs and DNFs) of the same size, our positive result leaves open the fooling power of depth 1 sources. We obtain a strong negative result for such sources (see Figure 2):

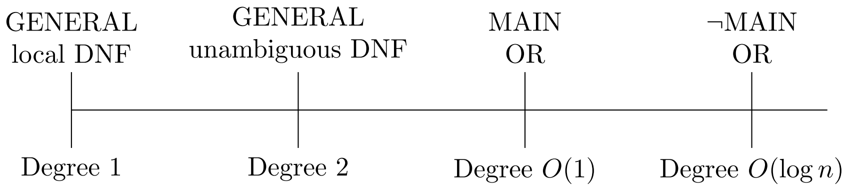


Figure 1: Main results in terms of degree for different classes of distinguishers.

4. $O(\log\log(n/\epsilon))$-indistinguishable depth 1 sources weakly ϵ-fool all functions. (Theorem 6.33)

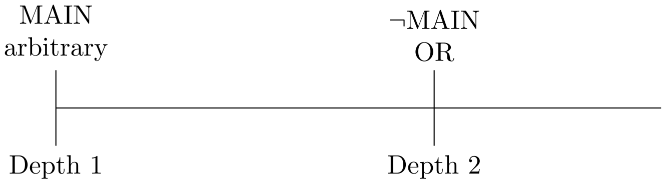


Figure 2: Main results in terms of depth for different classes of distinguishers.

This result is optimal not only in terms of the depth, but also in terms of the indistinguishability parameter, at least for constant ϵ (see a matching positive result in Lemma 6.39).

Barriers for linear sources. The basic building block of MPC protocols and other cryptographic applications is *linear* secret sharing. It is thus especially important to understand the consequences of bounded indistinguishability for linear sources. We believe that it is plausible to conjecture the following:

Conjecture 1. k-indistinguishable linear pairs of sources on n bits $o(1)$-fool AC^0 when $k = \mathsf{polylog}(n)$.

When one of the sources is uniform, this is implied by Braverman's theorem [Bra11, Tal17]. When the distinguisher is the OR function, it follows from our first negative result. In Section 4.2 we show, however, that proving Conjecture 1 for any $k = o(n/\log n)$ requires first proving the "IPAP conjecture" (Inner Product by AC^0 over Parities) of Servedio and Viola [SV12], which states that the binary inner product function on n inputs (IP) cannot be computed by $\mathsf{AC}^0 \circ \oplus$ circuits, i.e. bounded-depth AND/OR circuits with a bottom layer of PARITY gates. While a number of partial results have been obtained in support of IPAP [CS16, CGJ$^+$16, BKT19], it currently remains out of reach.

While IP is known not to be computable by the subclass $\mathsf{DNF} \circ \oplus$ of $\mathsf{AC}^0 \circ \oplus$ [SV12, ABG$^+$14], its approximability on a constant fraction of inputs remains open [CS16]. Proving even the special case of Conjecture 1 when the class of distinguishers is restricted to DNFs requires resolving this problem.

One possible approach for making progress on Conjecture 1 (and therefore also IPAP) is to find, for every pair of k-indistinguishable linear sources, an AC^0 reduction that maps them to some pair of k'-*independent* sources. Claim 7.6 in Section 7.2 rules out the existence of NC^0 reductions of this type in general. However, in Section 7.1 we give examples of linear NC^0 reductions to bounded independence for specific k-indistinguishable pairs of sources that describe the views of MPC protocols. The results of [BIS19] are also proved via reductions of this type.

The examples in Section 7.1 are related to the study of the complexity of distributions [ASTS$^+$98, GGN10, Vio10, LV11, BIL12, DW12, Vio12, Vio14, Vio16, Vio20, Vio21], intimately related to the study of extractors [VT00]. However, this line of study focuses on the complexity of sampling distributions given uniform sources, whereas we allow arbitrary k-independent sources.

On the gap between IPAP and Conjecture 1: predicting parity from parities. While a positive resolution of the IPAP conjecture is necessary to prove Conjecture 1, it is unclear if it is sufficient. Towards bridging this gap, in Section 4.2 we show that Conjecture 1 is implied by $\mathsf{PREDICTION}_{\oplus}(\mathsf{AC}^0, \Omega(1/n))$, where $\mathsf{PREDICTION}_{\oplus}(\mathcal{C}, \epsilon)$ is the following statement (see Conjecture 5):

> A class-$\mathcal{C}$ circuit on n inputs that is given as advice some set S of linear functions of its inputs, under the constraint that no $\mathsf{polylog}(n)$ of the functions in S XOR to the parity of all inputs, cannot predict parity on a $(1+\epsilon)/2$ fraction of inputs.

In the other direction, $\mathsf{PREDICTION}_{\oplus}(\mathsf{AC}^0, \Omega(1))$ implies the average-case IPAP conjecture (see Figure 3). As additional evidence towards Conjecture 1, we prove that $\mathsf{PREDICTION}_{\oplus}(\text{size-}s\ \mathsf{DNF}, 1-\Omega(1/s))$ holds for $s = \mathsf{poly}(n)$, thereby strengthening a result of Cohen and Shinkar [CS16] (see Corollary 4.6).

To give a bit more intuition on the distinction between Conjecture 1 and the IPAP conjecture: Refuting Conjecture 1 is equivalent to showing that some (polynomial-length) $\mathbb{F}_2$-linear encoding of n input bits can be used by an AC^0 circuit to nontrivially predict the parity of *some* subset of these bits. (Here "nontrivially" means that the target parity is not spanned by polylogarithmically many outputs of the encoding.) In contrast, refuting the IPAP conjecture requires proving the existence of a single encoding as above that enables AC^0 circuits to predict the parity of *every* subset. The equivalence between the two conjectures is open even if we replace "predict" by "exactly compute."

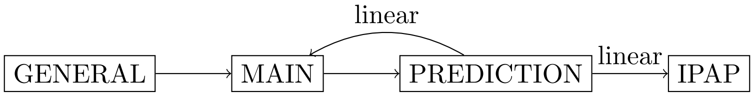


Figure 3: Relations between indistinguishability, prediction, and the IPAP conjecture

Applications to leakage-resilient cryptography. We already discussed applications to low-complexity secret sharing. In Section 9 we consider applications to *leakage-resilient circuit compilers* (LRCC) [ISW03], which protect sensitive computations against leakage from the internal wires of the computation. More concretely, an LRCC transforms a circuit C into a randomized circuit $\widehat{C}$ mapping an encoded input to an encoded output, such that revealing the output of a leakage function applied to wires of $\widehat{C}$ reveals essentially nothing about the input. Much of the work in this area focuses on obtaining efficient constructions for *local* leakage, confined to a small subset of k wires. Following [MR04], Faust et al. [FRR$^+$14] considered the global leakage model where the leakage function acts on all the wires but is restricted to a low complexity class such as AC^0. LRCC constructions in this model, such as those of Rothblum [Rot12] and Bogdanov et al. [BIS19], are complex to analyze and incur a significant overhead, compiling a circuit C to $\widehat{C}$ of size $\tilde{O}(\lambda^2|C|)$ for a security error parameter $2^{-\lambda}$. In contrast, the best known LRCC constructions in the local leakage model based on efficient MPC protocols [DN07, DIK10] can be quite efficient and only incur a polylogarithmic overhead in the local leakage parameter k. A natural question is whether this gap is inherent.

We show that one can bridge the efficiency gap between the local leakage and the global leakage models assuming our main conjecture holds for *quadratic* sources. Specifically, assuming this conjecture, we give a construction of LRCC against AC^0 circuits with $|\widehat{C}| = |C| \cdot \mathsf{polylog}(\lambda)$ (plus additive terms that only depend on the depth of C). As an additional application, we use the same conjecture for *linear* sources to show that a construction of LRCC from [ISW03, BIS19] for the class of circuits that only contain XOR gates satisfies a stronger security property. Namely, we show that security against AC^0 leakage is retained even when the output decoder is not implemented by a trusted hardware. We also show how to improve the efficiency of this construction by relying on a high-rate variant of Shamir's secret-sharing scheme [FY92].

Summary of unconditional applications. While several of the cryptographic applications presented in this work depend on unproven conjectures, others can be based on theorems we prove unconditionally. For convenience, we summarize applications of the latter kind below.

- Low-complexity secret sharing. Our positive results imply secret-sharing schemes with secrecy threshold $k = \Omega(\sqrt{n})$, reconstruction by OR[4] (with small constant error probability), and sharing by (depth-2) polynomial-size decision trees or degree-$O(\log n)$ $\mathbb{F}_2$-polynomials (Section 5.2 and Section 5.3 respectively). This improves over similar results in [BIVW16] in which sharing is done by higher depth AC^0 circuits. We show that our schemes are depth-optimal by ruling out similar schemes with *depth-1* sharing. Concretely, we show that the highest achievable secrecy threshold for schemes with depth-1 sharing is $k = \Theta(\log \log n)$ (see Section 6.5). Finally, our results imply the first *evolving* visual secret-sharing scheme in the sense of [KNY16] (see Section 5.5).

- Leakage-resilient cryptography. Our negative results imply that k-indistinguishability of degree-1 or degree-2 sources with $k \geq \mathsf{polylog}(n)$ suffices for protecting against low-depth leakage classes, including depth-1 AC^0 and unambiguous DNF. The latter capture natural kinds of selective failure attacks. We further show that degree-2 sources suffice in the context of efficient leakage-resilient circuit compilers. In particular, all of the applications discussed in Section 9 apply unconditionally to leakage by depth-1 AC^0 and unambiguous DNF.

1.2 Open questions

Our results suggest many open questions. We would like to single out the following.

Open Question 1. What is the smallest degree d for which there are $\Theta(\sqrt{n})$-indistinguishable degree d sources which OR can $\Omega(1)$-distinguish?

Our results show that $d = \omega(1)$ and $d = O(\log n)$.

Open Question 2. Are the GENERAL and MAIN conjectures equivalent? Is the PREDICTION conjecture for linear sources implied by IPAP?

We are mainly interested in the case of AC^0 distinguishers. GENERAL trivially implies MAIN, and PREDICTION for linear sources implies IPAP, so the open question is asking for the converse directions. We are able to show that MAIN and PREDICTION are equivalent for linear sources (for general sources, we only know that MAIN implies PREDICTION). A positive answer to the latter question roughly amounts to showing that if linear preprocessing can help AC^0 circuits nontrivially predict *some* parity of n bits then there is universal linear preprocessing that helps predict *all* parities. This implication is open even for exact computation.

Open Question 3. Is there a pair of $n^{\Omega(1)}$-indistinguishable sources, samplable in NC^0, which can be $\Omega(1)$-distinguished in AC^0?

A positive answer would imply an extreme form of low-complexity secret sharing, where secrets are shared by NC^0 circuits and reconstructed by AC^0 circuits. Our positive results imply weaker secret-sharing schemes with sharing by polynomial-size decision trees. In Section 8 we show that a negative answer to the question would imply a natural conjecture on low-complexity randomized encodings of functions. Another reason why settling Open Question 3 in the negative may be challenging is the difficulty of ruling out local sampling (up to a small statistical error) even for some simple and explicit distributions [Vio20].

Paper organization. We give an overview of our techniques in Section 2. After brief preliminaries in Section 3, we formally introduce our main conjectures in Section 4, where we also discuss some results on linear sources. Our constructions of low-complexity $\Omega(\sqrt{n})$-indistinguishable sources which OR can distinguish appear in Section 5. Our indistinguishability results appear in Section 6. The strategy of proving bounded indistinguishability results by reduction to bounded independence is considered in Section 7. Randomized encodings are considered in Section 8. Finally, applications to leakage-resilient cryptography are discussed in Section 9.

[4] Alternatively, allowing AC^0 reconstruction, an amplification technique from [BIVW16] can be used to obtain near-threshold schemes with negligible reconstruction error and the same sharing complexity.

Acknowledgements. We thank Chin Ho Lee, Igor Oliveira, Rahul Santhanam, Amir Shpilka, Justin Thaler, and Emanuele Viola for useful feedback. Special thanks go to Chin Ho Lee, who suggested Conjecture 2, and Justin Thaler, who suggested the construction in Section 5.1.

A. Bogdanov was supported by Hong Kong RGC GRF projects CUHK14207618 and CUHK14209419. The work of K. Dinesh was partially done while at the Institute of Theoretical Computer Science and Communications, The Chinese University of Hong Kong. Y. Filmus was supported by ERC Project HARMONIC (802020) and ISF grant 1337/16. Y. Ishai was supported by ERC Project NTSC (742754), BSF grant 2018393, and ISF grant 2774/20.

2 Overview of techniques

In this section we outline the proofs of some of our main results. In Section 2.1 we describe our construction of $\Omega(\sqrt{n})$-indistinguishable sources that are samplable by sources of degree $O(\log n)$ and are $\Omega(1)$-distinguished by OR, which appears in Section 5. In Section 2.2 we describe our various indistinguishability results, which comprise Section 6; we also cover the proof of Corollary 4.6. Finally, in Section 2.3 we outline the proof of the equivalence of MAIN and PREDICTION for linear sources, and the proof that LDPC sources cannot be reduced to bounded independence using local maps.

2.1 OR can distinguish logarithmic degree sources

Bogdanov et al. [BIVW16] showed that there exists a pair $\boldsymbol{X}, \boldsymbol{Y}$ of $\sqrt{n}$-indistinguishable sources over $\{0,1\}^n$ which OR distinguishes, by appealing to LP duality. Explicit constructions appear in other works, for example Špalek [Spa08] and Bun and Thaler [BT13]. However, except for a construction of AC^0-sampleable sources from [BIVW16], the corresponding distributions do not satisfy natural notions of computational simplicity.

We convert an arbitrary pair of $\sqrt{n}$-indistinguishable distributions which OR can distinguish into a similar pair samplable by simple sources using a sequence of reductions:

$$\textit{Arbitrary sources} \implies \textit{Mixtures of iid} \implies \textit{Decision trees} \implies O(\log n)\ \textit{degree}$$

Each of these reductions preserves indistinguishability (possibly modifying n) while having only a small effect on the distinguishing advantage of OR.

Mixtures of iid A distribution on $\{0,1\}^n$ is a *mixture of iid* if we can sample it using a two-step process:

1. Sample a bias $p \in [0,1]$ according to some distribution on $[0,1]$.
2. Sample n iid bits with bias p.

Given an arbitrary source $\boldsymbol{X}_0$ over $\{0,1\}^m$, we construct a mixture of iid $\boldsymbol{X}_1$ using *erase-all-subscripts symmetrization* [BT20a]: Sample $x \sim \boldsymbol{X}_0$, and then sample n uniform bits chosen from x.

If $\boldsymbol{X}_0, \boldsymbol{Y}_0$ are k-indistinguishable and we construct $\boldsymbol{X}_1, \boldsymbol{Y}_1$ in this fashion, then $\boldsymbol{X}_1, \boldsymbol{Y}_1$ are still k-indistinguishable. If $\boldsymbol{X}_0, \boldsymbol{Y}_0$ are ϵ-distinguished by OR then this means that $|\Pr[\boldsymbol{X}_0 = \mathbf{0}] - \Pr[\boldsymbol{Y}_0 = \mathbf{0}]| \geq \epsilon$. Since

$$\Pr[\boldsymbol{X}_0 = \mathbf{0}] \leq \Pr[\boldsymbol{X}_1 = \mathbf{0}] \leq \Pr[\boldsymbol{X}_0 = \mathbf{0}] + \left(1 - \frac{1}{m}\right)^n,$$

if we choose $n = \Theta(m \log(1/\epsilon))$ then $\boldsymbol{X}_1, \boldsymbol{Y}_1$ are $\Omega(\epsilon)$-distinguished by OR. We can choose $\boldsymbol{X}_0, \boldsymbol{Y}_0$ to be k-indistinguishable for $k = \Theta(\sqrt{m}) = \Theta(\sqrt{n})$.

Decision trees The next step is to show that we can approximately sample $\boldsymbol{X}_1, \boldsymbol{Y}_1$ using decision trees whose randomness derives from a supply of unbiased random bits. If we had access to biased random bits, then this would be immediate, and we can simulate biased random bits using unbiased random bits with some small failure probability. In order to maintain k-indistinguishability, in case of failure we output the constant vector $\mathbf{0}$. In this way we construct a pair of sources $\boldsymbol{X}_2, \boldsymbol{Y}_2$ which are k-indistinguishable and are $\Omega(\epsilon)$-distinguished by OR.

How large are the decision trees used to sample $\boldsymbol{X}_2, \boldsymbol{Y}_2$? This depends both on the failure probability and on the *complexity* of $\boldsymbol{X}_1, \boldsymbol{Y}_1$, as measured in the bit complexity of the probabilities used to define these mixtures of iid. Taking a close look at the construction of Bun and Thaler [BT13], we show that if we use it as our starting point $\boldsymbol{X}_0, \boldsymbol{Y}_0$ then the resulting $\boldsymbol{X}_1, \boldsymbol{Y}_1$ are low complexity, and so $\boldsymbol{X}_2, \boldsymbol{Y}_2$ are samplable using polynomial size decision trees for any constant failure probability.

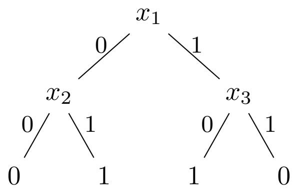
Logarithmic degree The final step is converting $\boldsymbol{X}_2, \boldsymbol{Y}_2$ to a pair of distributions $\boldsymbol{X}_3, \boldsymbol{Y}_3$ samplable by sources of degree $O(\log n)$. The idea is to used a *randomized encoding* inspired by the Razborov–Smolensky [Raz87, Smo87] lower bound technique. (See Section 8 for a more general perspective using the randomized encoding framework of [AIK06].)

Razborov and Smolensky approximate the AND function on ℓ bits to error 2^{-d} using the degree-d $\mathbb{F}_2$ polynomial

$$\prod_{i=1}^{d} \left(1 + \sum_{j=1}^{\ell} r_{i,j} (1 + x_j) \right).$$

Here $x_1, \ldots, x_\ell$ are the inputs, and $r_{i,j}$ are random bits. When $x_1 = \cdots = x_\ell = 1$, this expression always equals 1, and otherwise each factor is a random bit, and so the expression equals 0 with probability $1 - 2^{-d}$.

A decision tree can be written as an “unambiguous” sum of conjunctions, that is, at most one conjunction can be true. For example, the decision tree



can be expressed as

$$(1 - x_1)(1 - x_3) + x_1 x_2.$$

We have one conjunction per leaf labeled 1, and the conjunction corresponds to the path leading to the leaf.

We convert the decision tree into a polynomial by replacing each conjunction with its Razborov–Smolensky encoding. If the decision tree has size s then we need the error to be $O(\epsilon/s)$, and so the resulting degree is $\log(s/\epsilon)$. When s is polynomial, this is $O(\log(n/\epsilon))$.

We note that when attempting to apply the Razborov–Smolensky encoding to a general AC^0 circuit, rather than a decision tree or an unambiguous DNF, not only does the degree of the encoding grow to $\mathsf{polylog}(n)$, but there is also an encoding *privacy error*. The latter results in an approximate notion of k-indistinguishability in which the k-projections have $2^{-\mathsf{polylog}(n)}$ statistical distance. This relaxed notion, studied in [BW17], is qualitatively weaker than the perfect notion we consider in this work. In particular, it may totally break down when the projection set is chosen in an adaptive fashion. See Section 8 for more details.

2.2 Fooling OR and DNFs

In this section we describe our various negative results, as described in Table 1. Most of these results are proved via the notion of *predictability*, which we first explain. We then briefly outline the proofs of the remaining negative results.

2.2.1 Predictability

Let $\boldsymbol{X}$ be a source over $\{0,1\}^n$. We say that a subset S of coordinates ϵ-*predicts* $\boldsymbol{X}$ if

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] \leq \epsilon.$$

Roughly speaking, this means that in order to know the value of OR on $\boldsymbol{X}$, it suffices to peek at the coordinates in S.

If $\boldsymbol{X}, \boldsymbol{Y}$ are each ϵ-predicted by a subset of k coordinates, then the union of the two subsets ϵ-predicts both sources. Hence if $\boldsymbol{X}, \boldsymbol{Y}$ are $2k$-indistinguishable, then they ϵ-fool OR.

A more surprising observation is that if $\boldsymbol{Y}$ is ϵ/n-predicted by a subset S of k coordinates and $\boldsymbol{X}, \boldsymbol{Y}$ are $(k+1)$-indistinguishable, then S also ϵ-predicts $\boldsymbol{X}$; this is because for any coordinate $i \notin S$,

$$\Pr[\boldsymbol{Y}|_S = 0 \text{ and } \boldsymbol{Y}_i \neq 0] \leq \frac{\epsilon}{n}.$$

Accordingly, we define two notions of predictability for classes of sources:

- $\mathcal{F}$ is *weakly predictable* if for every $\epsilon > 0$, any source from $\mathcal{F}$ is ϵ-predicted by a subset of $C(\epsilon)$ coordinates.
  If $\mathcal{F}$ is weakly predictable and $\boldsymbol{X}, \boldsymbol{Y}$ are $C(\epsilon)$-indistinguishable sources from $\mathcal{F}$, then they ϵ-fool OR.
- $\mathcal{F}$ is *strongly predictable* if for every $\epsilon > 0$, any source from $\mathcal{F}$ is ϵ-predicted by a subset of $\mathsf{polylog}(1/\epsilon)$ coordinates.
  If $\mathcal{F}$ is strongly predictable and $\boldsymbol{X}, \boldsymbol{Y}$ are $\mathsf{polylog}(n/\epsilon)$-indistinguishable sources, where $\boldsymbol{Y} \in \mathcal{F}$, then they ϵ-fool OR.

Strongly predictable sources in fact fool not only OR, but also *unambiguous DNFs*. An unambiguous DNF is a disjunction of conjunctions, with the promise that no two conjunctions can be satisfied simultaneously. As explained in Section 2.1, a decision tree of size s can be converted to an unambiguous disjunction of at most s conjunctions. Writing the unambiguous DNF as a sum of ANDs (over the reals!), it suffices to (ϵ/s)-fool each AND in order to ϵ-fool the entire DNF. Consequently (since fooling ANDs and ORs is the same), $\mathsf{polylog}(ns/\epsilon)$-indistinguishable sources ϵ-fool unambiguous DNFs as long as one of the sources belongs to a strongly predictable class of sources which is closed under input negation.

2.2.2 Applying predictability

Our main results are:

- Constant degree sources are weakly predictable. This also includes sources of constant locality.
- Quadratic sources (i.e., degree 2 sources) are strongly predictable.

We also show that linear sources fool *local DNFs*, which are disjunctions of local functions. The proof is very similar to the proof that local sources fool OR, and so we do not describe it here.

Linear sources We prove predictability using the structure vs randomness paradigm. As an example, consider the class of linear sources, in which each output bit is an affine combination of input bits. For ease of exposition, we consider the special case in which each output bit is a *linear* combination of inputs bits (i.e., we disallow $x_1 = r_1 \oplus r_2 \oplus 1$). We will show that every linear source $\boldsymbol{X}$ is ϵ-predicted by a subset of $\log(1/\epsilon)$ coordinates.

The source $\boldsymbol{X}$ is *pseudorandom* if it has rank at least $\log(1/\epsilon)$. In this case, any subset S of $\log(1/\epsilon)$ linearly independent coordinates ϵ-predicts $\boldsymbol{X}$, since $\Pr[\boldsymbol{X}|_S = 0] \leq \epsilon$.

The source $\boldsymbol{X}$ is *structured* if it has rank at most $\log(1/\epsilon)$. In this case, we choose a subset S such that $\{\boldsymbol{X}_i\}_{i \in S}$ spans $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. This subset 0-predicts $\boldsymbol{X}$ since if $\boldsymbol{X}|_S = 0$ then $\boldsymbol{X} = 0$.

Local sources A more sophisticated example is that of s-local sources, that is, sources where every output bit $\boldsymbol{X}_i$ depends on at most s input bits, forming a set J_i. Suppose that we are given such a source $\boldsymbol{X}$.

The source $\boldsymbol{X}$ is *pseudorandom* if we can find $2^s \log(1/\epsilon)$ coordinates which depend on disjoint sets of inputs. A short calculation shows that the probability that all these coordinates equal zero is at most ϵ.

Otherwise, the source $\boldsymbol{X}$ is *structured*: we can find a "hitting set" T of size $s2^s \log(1/\epsilon)$ for $J_1, \ldots, J_n$. For each setting of the input bits in T, the source simplifies to an $(s-1)$-local source, and we can find an ϵ-predicting set by induction. Putting all of these sets together, we obtain an ϵ-predicting set for the original source.

A very similar argument appears in work of Trevisan [Tre04], in the context of deterministic approximate counting of solutions to k-CNFs, and in recent work of Akmal and Williams [AW21], in the context of threshold counting of solutions to k-CNFs. See Williams [Wil18] for deterministic approximate counting of solutions to systems of polynomial equations, a topic related to our next example, constant degree sources.

Constant degree sources We handle degree d sources using a similar argument. We need to find a pseudorandomness condition for a set S of coordinates which will guarantee that $\Pr[\boldsymbol{X}|_S = 0] \leq \epsilon$. Such a condition is supplied by higher-order Fourier analysis: if all linear combinations of $\{\boldsymbol{X}_i\}_{i \in S}$ have high *rank* (a notion we explain below) and S is large enough, then $\Pr[\boldsymbol{X}|_S = 0] \leq \epsilon$ (pseudorandom case).

Otherwise (structured case), we choose a maximal set T such that all linear combinations of $\{\boldsymbol{X}_i\}_{i \in T}$ have high rank. By the definition of rank, this implies that each $i \notin T$ simplifies, modulo $\{\boldsymbol{X}_i\}_{i \in T}$, to a function depending on a bounded number of degree $d-1$ polynomials, and we can complete the proof by induction.

Quadratic sources The arguments for local sources and for constant degree sources result in a very bad dependence between ϵ and the size $C(\epsilon)$ of the ϵ-fooling subset of coordinates. In the case of quadratic sources, we are able to use Dickson's structure theorem for quadratic polynomials, via a series of careful reductions, to obtain the much better dependence $C(\epsilon) = O(\log^{10}(1/\epsilon))$.

2.2.3 Other negative results

We prove two other negative results: the prediction variant holds for linear sources and DNF distinguishers, and depth 1 sources fool arbitrary distinguishers.

PREDICTION **holds for linear sources and DNF distinguishers** Given a DNF ϕ and a linear source $\boldsymbol{X}$, our goal is to show that if no k coordinates of $\boldsymbol{X}$ span some target parity π, then ϕ cannot compute π, even with a small error.

If T is any term of ϕ, then the probability that T is satisfied is $2^{-\operatorname{rank}(T)}$, where the rank of T is the rank of the span of the corresponding coordinates of $\boldsymbol{X}$. If T has large rank then it is unlikely to be satisfied, so we can drop all of these terms, obtaining a narrow DNF ψ.

We now apply Jackson's lemma [Jac94], according to which ψ must correlate with some Fourier character χ_S, where S is a subset of the set of variables appearing in some term of ψ. Since all terms in ψ are narrow and ψ computes π (with small error), this implies that π has nontrivial correlation with, and so is equal to, a linear combination of a small number of coordinates in $\boldsymbol{X}$, which contradicts our initial assumption.

Depth 1 sources fool arbitrary distinguishers Let $\boldsymbol{X}, \boldsymbol{Y}$ be k-indistinguishable depth 1 sources, that is, each coordinate is an AND or OR of literals. Since we allow arbitrary distinguishers, we can assume that each coordinate is an AND of literals.

Wide conjunctive coordinates are hardly ever 1, so allowing for a small error, we can replace them with constant 0 coordinates. We are left with only narrow coordinates, say of width at most $\log(n/\epsilon)$. Applying a result of Amano et al. [AIM+03], if $k = \log\log(n/\epsilon) + 2$ then the two truncated sources are identically distributed, completing the proof.

2.3 Other results

MAIN and PREDICTION are equivalent for linear sources To prove the equivalence between Conjecture 9 ($\mathsf{MAIN}_\oplus(\mathsf{AC}^0)$) and $\mathsf{PREDICTION}_\oplus(\mathsf{AC}^0)$, we consider an equivalent formulation of $\mathsf{PREDICTION}_\oplus(\mathsf{AC}^0)$, which we call $\mathsf{COSET}_\oplus(\mathsf{AC}^0)$. This is the special case of $\mathsf{MAIN}_\oplus(\mathsf{AC}^0)$ in which the two k-indistinguishable sources arise from a single source by fixing the first bit of the seed. The resulting sources are uniformly distributed on two cosets of the same linear subspace, hence the name. The equivalence of the two formulations is a simple exercise (see Section 4).

Two linear sources are k-indistinguishable if they satisfy the same affine constraints of width k or less. This suggests the following strategy for proving $\mathsf{MAIN}_\oplus$ (with parameters k, ϵ) given $\mathsf{COSET}_\oplus$ (with parameters k, δ): Given two k-indistinguishable linear sources $\boldsymbol{X}, \boldsymbol{Y}$, construct the "free k-indistinguishable source" $\boldsymbol{Z}$ given by all affine constraints of width at most k satisfied by $\boldsymbol{X}$. This is the most general linear source which is k-indistinguishable from $\boldsymbol{X}$. Moreover, we obtain exactly the same source if we apply the same construction to $\boldsymbol{Y}$. Therefore it suffices to show that $\boldsymbol{X}, \boldsymbol{Z}$ fool $\mathcal{C}$.

The idea is to construct a sequence of hybrids $\boldsymbol{Z}_0, \ldots, \boldsymbol{Z}_t$, where $\boldsymbol{Z}_0 = \boldsymbol{Z}$, $\boldsymbol{Z}_t = \boldsymbol{X}$, and $\boldsymbol{Z}_{i+1}$ is obtained from $\boldsymbol{Z}_i$ by imposing one more affine constraint. We can also define $\boldsymbol{W}_{i+1}$ in the same way, by imposing the opposite constraint (for example, $x_1 \oplus x_2 = 1$ rather than $x_1 \oplus x_2 = 0$). By construction, $\boldsymbol{Z}_{i+1}, \boldsymbol{W}_{i+1}$ are cosets, and so $\mathsf{COSET}_\oplus(\mathsf{AC}^0)$ shows that they δ-fool $\mathcal{C}$. On the other hand, $\boldsymbol{Z}_i$ is a $\frac{1}{2}$-$\frac{1}{2}$ mixture of $\boldsymbol{Z}_{i+1}, \boldsymbol{W}_{i+1}$, and so $\boldsymbol{Z}_i, \boldsymbol{Z}_{i+1}$ $\delta/2$-fool $\mathcal{C}$.

In total, $\boldsymbol{X}, \boldsymbol{Z}$ $t\delta/2$-fool $\mathcal{C}$, and so $\boldsymbol{X}, \boldsymbol{Y}$ $t\delta$-fool $\mathcal{C}$. Clearly $t \le n$, and so it suffices to take $\delta = \epsilon/n$.

LDPC codes cannot be reduced to bounded independence using local maps An LDPC code is a code whose parity-check matrix is sparse: every message bit appears in exactly D parity checks (this is one of several common definitions). If we choose a $\theta n \times n$ parity-check matrix at random, then the bipartite graph corresponding to the parity-check matrix will be an expander, and so the corresponding code will have linear minimum distance, say at least γn.

A simple sensitivity argument shows that for large n, such a code C cannot be generated using B-local maps from the uniform distribution over m bits: The $n \times m$ binary matrix describing which input bits each output bit depends on contains at most Bn ones, and so there must be some input bit affecting at most Bn/m output bits. Flipping this bit results in flipping at most Bn/m input bits. Since the minimum distance of C is at least γn, this shows that $m \le B/\gamma$. On the other hand, m must be at least the rate $(1-\theta)n$ of the code, and we obtain a contradiction for $n > B/\gamma(1-\theta)$.

Does the picture change if we are allowed to reduce to an arbitrary k-independent distribution $\boldsymbol{z}$? Let P be the parity-check matrix of C, and let F denote the B-local reduction. Thus $PF(z) = 0$ for all z in the support of $\boldsymbol{z}$. Since every column of P contains D many ones, the average row of P contains D/θ many ones, and so the typical entry of $PF(z)$ depends on at most BD/θ many bits of $\boldsymbol{z}$. If $BD/\theta \ll k$ then the projection of $\boldsymbol{z}$ to these coordinates will have full support due to k-independence, and so $PF(z) = 0$ for *all* z. Thus F also works as a reduction to the uniform distribution, allowing us to apply the earlier lower bound.

3 Preliminaries

In this section we provide definitions and notation that will be used throughout the paper.

Definition 3.1 (ϵ-approximation)**.** Let $f, g \colon \{0,1\}^m \to \{0,1\}$ be Boolean functions. We say that f *ϵ-approximates* g if $\Pr_{\boldsymbol{x}}[f(\boldsymbol{x}) \neq g(\boldsymbol{x})] \le \epsilon$.

Definition 3.2 (ϵ-fooling)**.** A pair of distributions $\boldsymbol{X}, \boldsymbol{Y}$ over $\{0,1\}^n$ is said to *ϵ-fool* a function $f \colon \{0,1\}^n \to \{0,1\}$ if $|\mathsf{E}[f(\boldsymbol{X})] - \mathsf{E}[f(\boldsymbol{Y})]| \le \epsilon$. We say that f *ϵ-distinguishes* between $\boldsymbol{X}, \boldsymbol{Y}$ if $|\mathsf{E}[f(\boldsymbol{X})] - \mathsf{E}[f(\boldsymbol{Y})]| \ge \epsilon$. Finally, we say that $\boldsymbol{X}$ ϵ-fools f if $\boldsymbol{X}, \boldsymbol{U}_n$ ϵ-fool f, where $\boldsymbol{U}_n$ is the uniform distribution over $\{0,1\}^n$.

Definition 3.3 (k-independence and k-indistinguishability)**.** We say that a distribution $\boldsymbol{X}$ over $\{0,1\}^n$ is *k-independent* if its marginal distribution on any subset of k coordinates is the uniform distribution.

We say that two distributions $\boldsymbol{X}, \boldsymbol{Y}$ over $\{0,1\}^n$ are *k-indistinguishable* if for any subset I of k coordinates, their marginal distributions on I are the same.

Definition 3.4 (Boolean circuits and distinguishers)**.** We consider Boolean circuits with AND/OR/NOT gates and unbounded fan-in. We refer to them interchangeably as *circuits* or *distinguishers*, where the latter emphasizes the usage of circuits to distinguish between distributions. The *depth* of a circuit is the longest path from any input gate to any output gate, and the *size* of a circuit is the number of wires it contains.

We are mainly interested in the following circuit subclasses: AC^0 *circuits*, which are circuits of constant depth and polynomial size; NC^0 *circuits*, which are circuits of constant size; circuits of depth 2 with a top OR gate and AND gates at the bottom layer, also known as DNFs; and the special case of a single OR/AND gate.

Definition 3.5 (Local functions)**.** A function is called *ℓ-local* if it depends on at most ℓ input bits.

Definition 3.6 (Sources and types of sources)**.** We call a *source* any distribution on $\{0,1\}^n$. Given a class of Boolean functions $\mathcal{F}$, we call a source $\boldsymbol{X}$ an *$\mathcal{F}$-samplable source* if there exists a *sampler* consisting of n functions $f_1, \ldots, f_n \colon \{0,1\}^m \to \{0,1\}$ in $\mathcal{F}$ such that $\boldsymbol{X}$ can be sampled by evaluating $(f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x}))$ on a uniformly random $\boldsymbol{x} \in \mathbb{F}_2^m$.[5] We shall also use the term *sampler* to refer to a single function taken from $\mathcal{F}$.

A sampler is called a *degree-d sampler* if it consists only of degree-d polynomials, and a source is called a *degree-d source* if it is samplable by a degree-d sampler; we refer to degree 1 sources as *linear* sources, and to degree 2 sources as *quadratic* sources. A sampler is called an *ℓ-local sampler* if it consists only of ℓ-local functions, and a source is called an *ℓ-local source* if it is samplable by an ℓ-local sampler.

Informally, we call a source *simple* if it is simple in some intuitive sense, such as being a low-degree, local, or low-complexity source.

4 Variants of bounded indistinguishability and prediction

In this section we investigate the effect of relaxing the complexity requirement on one of the two distributions, as well as the relation between bounded indistinguishability and the problem of predicting the first bit of the distributions from the others. While the connections we obtain are strongest for linear sources, we formulate them in full generality whenever possible.

We present statements that serve as *templates* for our conjectures. We start with the most general form of the statements and later instantiate them for specific sources and distinguishers, which arise in our applications, along with their parameters. The conjectures are parameterized by a collection of $\mathcal{F}$-samplable sources; a class of circuits (distinguishers) $\mathcal{C}$;

Conjecture 2 (General variant)**.** $\mathsf{GENERAL}(\mathcal{F}, \mathcal{C}, k, \epsilon)$: Let $\boldsymbol{X}$ be an $\mathcal{F}$-samplable source, and $\boldsymbol{Y}$ an arbitrary source. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are k-indistinguishable, then $\boldsymbol{X}, \boldsymbol{Y}$ ϵ-fool C for any circuit $C \in \mathcal{C}$.

Conjecture 3 (Main variant)**.** $\mathsf{MAIN}(\mathcal{F}, \mathcal{C}, k, \epsilon)$: Let $\boldsymbol{X}, \boldsymbol{Y}$ be $\mathcal{F}$-samplable sources. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are k-indistinguishable, then $\boldsymbol{X}, \boldsymbol{Y}$ ϵ-fool C for any circuit $C \in \mathcal{C}$.

Conjecture 4 (Coset variant)**.** $\mathsf{COSET}(\mathcal{F}, \mathcal{C}, k, \epsilon)$: Let $f_1, \ldots, f_n \in \mathcal{F}$ be a sampler, and for $b \in \{0,1\}$, let $\boldsymbol{X}_b$ be the distribution obtained from $(f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x}))$ by setting the first input bit of each f_i to b. If $(\boldsymbol{X}_0, \boldsymbol{X}_1)$ are k-indistinguishable, then $\boldsymbol{X}_0, \boldsymbol{X}_1$ ϵ-fool C for any circuit $C \in \mathcal{C}$.

Conjecture 5 (Prediction variant)**.** $\mathsf{PREDICTION}(\mathcal{F}, \mathcal{C}, k, \delta)$: Let $f_1, \ldots, f_n \in \mathcal{F}$ be a sampler such that for any $i_1, \ldots, i_k$, the marginal distribution of $f_{i_1}, \ldots, f_{i_k}$ is independent of $\boldsymbol{x}_1$, namely conditioned on $\boldsymbol{x}_1 = 0$ it is the same as conditioned on $\boldsymbol{x}_1 = 1$. Then, no circuit $C \in \mathcal{C}$ on top of $f_1, \ldots, f_n$ can $(\frac{1-\delta}{2})$-approximate x_1.

[5] Note that we have two parameters here—n, which is the dimension of the distribution, and m, which is the number of independent random bits used to sample the distribution. We think of m as being $\mathsf{poly}(n)$ when it comes to positive results; for negative results, we assume no relation between the two parameters.

The condition on $f_1, \dots, f_n$ in the prediction variant means that no information about x_1 can be obtained from any k bits of $(f_1(x), \dots, f_n(x))$. In the special case of linear sources, it turns out that either no information or complete information can be obtained, so we can equivalently require that no linear combination of k sampler bits equals x_1. However, this need not be true even for quadratic sources; indeed, consider the quadratic source given by $(x_1x_2, \dots, x_1x_n, x_1x_{n+1})$, where no linear combination of the coordinates equals x_1, but x_1x_2 alone 1/4-approximates x_1.

Another thing to note about linear sources is that the target function x_1 can be replaced by any other parity π over x, as we can obtain an equivalent condition with a suitable linear transform. This means that for linear sources, we can restate $\mathsf{PREDICTION}$ as follows: Let $f_1, \dots, f_n$ be a linear sampler, no k bits of which span π. Then, no circuit $C \in \mathcal{C}$ on top of $f_1, \dots, f_n$ can $(\frac{1-\delta}{2})$-approximate π. This observation will be useful in Section 4.2.

Equivalence of COSET and $\mathsf{PREDICTION}$. For any choice of $\mathcal{F}, \mathcal{C}, k, \epsilon$, we have that $\mathsf{COSET}(\mathcal{F}, \mathcal{C}, k, \epsilon)$ holds if and only if $\mathsf{PREDICTION}(\mathcal{F}, \mathcal{C}, k, \epsilon)$ holds.

To see this, suppose, for the sake of contradiction, that the coset variant is true, but there exists a sampler $f_1, \dots, f_n \in \mathcal{F}$ and a circuit $C \in \mathcal{C}$ such that C on top of the samplers $(\frac{1-\epsilon}{2})$-approximates x_1. For $b \in \{0, 1\}$, let $\boldsymbol{X}_b$ be the source obtained by sampling $(f_1(x), \dots, f_n(x))$ with $x_1 = b$. If $(\boldsymbol{X}_0, \boldsymbol{X}_1)$ are not k-indistinguishable, then we could deduce information about x_1 from some k bits of $(f_1(x), \dots, f_n(x))$, violating the premise of the prediction variant. Hence, $\boldsymbol{X}_0, \boldsymbol{X}_1$ are k-indistinguishable, yet C ϵ-distinguishes between $\boldsymbol{X}_0, \boldsymbol{X}_1$, thus contradicting the coset variant.

For the converse direction, suppose, for the sake of contradiction, that the prediction variant is true, but there exists a sampler $f_1, \dots, f_n \in \mathcal{F}$ such that the distributions $\boldsymbol{X}_0, \boldsymbol{X}_1$, where $\boldsymbol{X}_b$ is obtained from the sampler when setting the first coordinate to b, are k-indistinguishable, yet some circuit $C \in \mathcal{C}$ ϵ-distinguishes between X_0, X_1. By k-indistinguishability of $(\boldsymbol{X}_0, \boldsymbol{X}_1)$, the premise of the prediction variant holds with the sampler $f_1, \dots, f_n$, yet the circuit C on top of the sampler can $(\frac{1-\epsilon}{2})$-approximates x_1, contradicting the prediction variant.

Is COSET equivalent to MAIN? While it is unclear if COSET captures the hardness of the MAIN conjecture, we show that up to some loss in parameters they are equivalent in two cases of interest. First, in Section 4.1 we show that up to a factor n in distinguishing advantage, the two are equivalent for linear sources. Second, COSET is also equivalent to MAIN for sources sampled by polynomial maps up to a loss of one in the degree: If $\boldsymbol{X}, \boldsymbol{Y}$ are degree d sources falsifying $\mathsf{MAIN}(\text{degree-}d, \mathcal{C}, k, \epsilon)$ then $\boldsymbol{Z} = \boldsymbol{z} \cdot \boldsymbol{X} + (1 + \boldsymbol{z}) \cdot \boldsymbol{Y}$, with $\boldsymbol{z}$ being a new variable, falsifies $\mathsf{COSET}(\text{degree-}(d+1), \mathcal{C}, k, \epsilon)$.

Linear sources A linear source is an $\mathcal{F}$-samplable source, where $\mathcal{F}$ consists of all degree 1 polynomials. The main and coset variants are equivalent for linear sources, up to loss in parameters, as we show in Section 4.1.

We then show, in Section 4.2, that the prediction variant for linear sources holds for DNFs when the error is required to be very small. We also relate the prediction variant to the IPAP question, which constitutes a barrier to further progress on this front.

To indicate that we restrict our attention to linear sources, we use the notation $\mathsf{GENERAL}_{\oplus}$, $\mathsf{MAIN}_{\oplus}$, $\mathsf{COSET}_{\oplus}$, $\mathsf{PREDICTION}_{\oplus}$ for the corresponding conjectures stated in the preceding section.

4.1 Equivalence of COSET and MAIN for linear sources

We show that for linear sources the main variant (Conjecture 3) and the coset variant (Conjecture 4) are equivalent up to a loss of $\frac{1}{n}$ factor in the indistinguishability advantage. To show that the coset variant implies the main variant (the other direction is trivial), subject to some loss in parameters, we make use of a characterization of k-indistinguishability over affine spaces.

Definition 4.1 (k-relatedness)**.** We say that two affine spaces U, V over $\mathbb{F}_2$ are *k-related* if they satisfy the same affine constraints involving up to k coordinates.

Proposition 4.2 (Characterization of k-indistinguishability over affine spaces)**.** *Let $\boldsymbol{X}, \boldsymbol{Y}$ be distributions that are uniformly distributed over affine spaces $U, V \subseteq \mathbb{F}_2^n$, respectively. Then $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable if and only if U, V are k-related.*

Proof. The affine subspaces U, V are k-related iff the parity of any k coordinates of $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable. By the XOR lemma, the latter condition is equivalent to the k-indistinguishability of $\boldsymbol{X}, \boldsymbol{Y}$. □

We are ready now to show the target implication.

Theorem 4.3. $\mathsf{COSET}_{\oplus}(\mathcal{C}, k, \epsilon)$ *implies* $\mathsf{MAIN}_{\oplus}(\mathcal{C}, k, n\epsilon)$*.*

Proof. Let $\boldsymbol{X}, \boldsymbol{Y}$ be k-indistinguishable distributions distributed uniformly over affine spaces U, V, respectively. Define W to be the affine space of all vectors satisfying all the affine relations of width at most k that are satisfied by all vectors of U, and let $\boldsymbol{Z}$ be the uniform distribution over W.

Consider the following subspace chain $W = U_0 \supsetneq U_1 \supsetneq \ldots \supsetneq U_t = U$, where $t = \dim(W) - \dim(U)$, and for every $0 < i \leq t$, the subspace U_i is generated by taking the subspace U_{i-1} and adding an affine constraint involving more than k coordinates; and let $\boldsymbol{Z} = \boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_t = \boldsymbol{X}$ be distributions such that $\boldsymbol{Z}_i$ is uniform over U_i.

Let $C \in \mathcal{C}$ be an n-bit circuit. Fix $0 < i \leq r$, and let $\ell(u) = b$ be an affine constraint that when added to U_{i-1} gives U_i. Consider the two distributions obtained from $\boldsymbol{Z}_{i-1}$ by conditioning on $\ell(u) = b$ and on $\ell(u) = 1 - b$. These distributions are uniform over two cosets of the same linear space, and are k-indistinguishable by Proposition 4.2. Thus, we can apply our assumption, and get

$$\begin{aligned}
\big|\Pr[C(\boldsymbol{Z}_i) = 1] - \Pr[C(\boldsymbol{Z}_{i-1}) = 1]\big| &= \big|\Pr[C(\boldsymbol{Z}_{i-1}) = 1 \mid \ell(u) = b] - \Pr[C(\boldsymbol{Z}_{i-1}) = 1]\big| \\
&= \frac{1}{2} \cdot \big|\Pr[C(\boldsymbol{Z}_{i-1}) = 1 \mid \ell(u) = b] - \Pr[C(\boldsymbol{Z}_{i-1}) = 1 \mid \ell(u) = 1 - b]\big| \\
&\leq \epsilon/2.
\end{aligned}$$

Hence,

$$\big|\Pr[C(\boldsymbol{X}) = 1] - \Pr[C(\boldsymbol{Z}) = 1]\big| \leq \sum_{i=1}^{t} \big|\Pr[C(\boldsymbol{Z}_i) = 1] - \Pr[C(\boldsymbol{Z}_{i-1}) = 1]\big| \leq t\epsilon/2 \leq n\epsilon/2.$$

We can repeat the same analysis, this time with the affine space W' of all vectors satisfying all the affine relations of width at most k that are satisfied by all vectors of V, and consider the uniform distribution $\boldsymbol{Z}'$ over W'. Proposition 4.2 shows that $W' = W$, which implies $\boldsymbol{Z}' = \boldsymbol{Z}$; therefore,

$$\begin{aligned}
\big|\Pr[C(\boldsymbol{X}) = 1] - \Pr[C(\boldsymbol{Y}) = 1]\big| &\leq \big|\Pr[C(\boldsymbol{X}) = 1] - \Pr[C(\boldsymbol{Z}) = 1]\big| + \big|\Pr[C(\boldsymbol{Y}) = 1] - \Pr[C(\boldsymbol{Z}) = 1]\big| \\
&\leq n\epsilon.
\end{aligned}$$

□

4.2 DNFs cannot compute PARITY with locally random linear advice

A direct argument, outlined in Section 6.1, shows that $\mathsf{PREDICTION}_{\oplus}(\mathcal{C}, k, 2^{-\Omega(k)})$ holds for $\mathcal{C} = \{\mathsf{OR}, \mathsf{AND}\}$. Extending this even to DNFs is already difficult, and we are only able to prove $\mathsf{PREDICTION}_{\oplus}(\mathsf{DNF}, k, 1-\epsilon)$ when ϵ is negligible (the exact statement appears below). To help explain this difficulty, we relate the question the problem of computing parity from parities, related to various conjectures in complexity theory. Lack of progress on these conjectures forms a barrier for our conjecture.

We will need the following result to prove our result for DNFs.

Lemma 4.4 (Jackson's Lemma [Jac94])**.** *For every DNF ϕ with s terms and for every distribution $\mathcal{D}$ on the inputs to ϕ, there exist a term T and a subset S of the variables appearing in T such that*

$$\left|\mathsf{E}_{\boldsymbol{x}\sim\mathcal{D}}\left[(-1)^{\phi(\boldsymbol{x})}\chi_S(\boldsymbol{x})\right]\right| \geq \frac{1}{2s+1},$$

where $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$ is the Fourier character associated with S.

Proposition 4.5. *Suppose that $P = \{p_1, \ldots, p_m\}$ is a set of parities over $\{x_1, \ldots, x_n\}$, such that no k parities in P span a given parity π, and suppose that ϕ is a DNF of top fan-in s satisfying*

$$\Pr_{\boldsymbol{x} \sim \{0,1\}^n}[\phi(p_1(\boldsymbol{x}), \ldots, p_m(\boldsymbol{x})) \neq \pi(\boldsymbol{x})] \leq \epsilon,$$

for some $0 \leq \epsilon < 1/(4s+2)$. Then:

$$k \leq \log\left(\frac{s(2s+1)}{1 - 2\epsilon(2s+1)}\right).$$

Proof. Consider an AND term T in ϕ with r entries whose inputs are $p_{i_1}(x), \ldots, p_{i_r}(x)$. We can represent F using a matrix $M \in \mathbb{F}_2^{r \times n}$, whose jth row corresponds to the parity p_{i_j}, such that T evaluates to True iff $Mx = b$, for some $b \in \{0,1\}^r$ representing the parities of the variables in T.

Suppose now that we remove T from ϕ. The resulting DNF disagrees with ϕ only on inputs on which T is False; hence, the probability of disagreement between the two circuits is bounded by

$$\Pr_{\boldsymbol{x}}[M\boldsymbol{x} = b] = \frac{2^{|\ker(M)|}}{2^n} = 2^{-(n - |\ker(M)|)} = 2^{-\operatorname{rank}(M)}.$$

Thus, if we remove from ϕ all terms for which the representing matrix has rank more than k, we get a new DNF ψ of top fan-in $s' \leq s$, that, by the union bound, satisfies

$$\Pr_{\boldsymbol{u}}[\psi(\boldsymbol{u}) \neq \phi(\boldsymbol{u})] \leq s2^{-(k+1)},$$

where $u = (p_1(x), \ldots, p_m(x))$.

It follows from Jackson's lemma that there exists a term T in ψ, and a subset S of variables occurring in it, such that

$$\left|\mathbb{E}_{\boldsymbol{u}}\left[(-1)^{\psi(\boldsymbol{u})}\chi_S(\boldsymbol{u})\right]\right| \geq \frac{1}{2s'+1} \geq \frac{1}{2s+1}.$$

Since u itself consists of parities, the parity $\chi_S(u)$ can be written as a parity in terms of x, denote it $\chi_R(x)$. Furthermore, χ_R is a parity spanned by $|S|$ parities.

It is easy to verify the following fact (which we will use twice): $|\mathbb{E}[fh]| \geq |\mathbb{E}[gh]| - 2\Pr[f \neq g]$ for every $\{-1,1\}$-valued functions f, g, h. Thus, we have:

$$\left|\mathsf{E}_{\boldsymbol{u}}\left[(-1)^{\phi(\boldsymbol{u})}\chi_S(\boldsymbol{u})\right]\right| \geq \left|\mathbb{E}_{\boldsymbol{u}}\left[(-1)^{\psi(\boldsymbol{u})}\chi_S(\boldsymbol{u})\right]\right| - s2^{-k} \geq \frac{1}{2s+1} - s2^{-k},$$

by which it follows that

$$\begin{aligned}\left|\mathsf{E}_{\boldsymbol{x}}\left[(-1)^{\pi(\boldsymbol{x})}\chi_R(\boldsymbol{x})\right]\right| &\geq \left|\mathbb{E}_{\boldsymbol{x}}\left[(-1)^{\phi(p_1(\boldsymbol{x}),\ldots,p_m(\boldsymbol{x}))}\chi_R(\boldsymbol{x})\right]\right| - 2\epsilon \\ &= \left|\mathbb{E}_{\boldsymbol{u}}\left[(-1)^{\phi(\boldsymbol{u})}\chi_S(\boldsymbol{u})\right]\right| - 2\epsilon \\ &\geq \frac{1}{2s+1} - s2^{-k} - 2\epsilon.\end{aligned}$$

Now, if $k > \log\left(\frac{s(2s+1)}{1-2\epsilon(2s+1)}\right)$, it follows that $\mathsf{E}[(-1)^{\pi}\chi_R] \neq 0$, which implies that π and χ_R must correspond to the same parity; however, this contradicts our assumption, because $|S| \leq |T| \leq k$ and $|S|$ parities from P span χ_R. □

This gives the following corollary.

Corollary 4.6. *For the class $\mathcal{C}$ of DNFs with top fan-in at most s, $\mathsf{PREDICTION}_{\bigoplus}(\mathcal{C}, k, 1-\epsilon)$ holds for any $k > \log\left(\frac{s(2s+1)}{1-2\epsilon(2s+1)}\right)$ and $0 \leq \epsilon < 1/(2s+1)$. In particular, for the class $\mathcal{C}$ of poly-size DNFs, we have that $\mathsf{PREDICTION}_{\bigoplus}(\mathcal{C}, k, 1-\epsilon)$ holds for $k = \omega(\log n)$ and negligible ϵ, i.e., $\epsilon = n^{-\omega(1)}$.*

Can we do better? It is still open as of now, however we can provide a barrier for this task. To this end, let us consider the computational model of AC^0 circuits with parity gates at the bottom, denoted by $\mathsf{AC}^0 \circ \oplus$, along with the inner product modulo 2 function, denoted by IP. Formally, we write the following parameterized conjecture.

Conjecture 6 (Inner product computation from parities)**.** $\mathsf{IPAP}(s, d, \delta)$: There does not exist a circuit of size s and depth d, whose inputs are parities over x, y, that $(\frac{1-\delta}{2})$-approximates $\mathsf{IP}_n(x, y)$.

The conjecture that inner product is hard for $\mathsf{AC}^0 \circ \oplus$ circuits even on average, namely that Conjecture 6 holds with $\mathsf{IPAP}(\mathsf{poly}(n), O(1), 1/\mathsf{poly}(n))$, was first raised by Servedio and Viola [SV12], and was further motivated by cryptographic applications in [ABG+14]. Although this conjecture hasn't been resolved thus far, some progress has been made since. For the case of *exact* computation (i.e., $\delta = 1$) and depth 2, Cohen and Shinkar [CS16] give a tight exponential bound on computing IP by a $\mathsf{DNF} \circ \oplus$ circuit (also known as DNF of parities), and conjecture that such circuits must be exponential in size even to approximate IP (their result improves on other results already providing exponential lower bounds [Gro94, Juk06]). Further progress on the problem has been made in [CGJ+16], where they give superlinear lower bounds in the worst case, and their result was later improved by [BKT19] to superlinear bounds for the average case.

Motivated by applications to cryptography, Rothblum [Rot12] made the similar conjecture that IP cannot be computed using AC^0 circuits whose inputs are arbitrary functions of one of the inputs, a model of computation known as bipartite complexity [PRS88] (if we only allow linear functions, this is the same as IPAP). Essentially the same question appears in communication complexity, where bipartite circuits of quasipolynomial size $2^{\log^{O(1)} n}$ correspond to the complexity class $\mathsf{PH}^{\mathsf{cc}}$, the communication complexity analog of the polynomial hierarchy [BFS86].

The following result provides a barrier for proving the prediction variant, or any other variant from our web of conjectures.

Proposition 4.7. $\mathsf{PREDICTION}_{\oplus}(\mathcal{C}, k, \delta) \implies \mathsf{IPAP}(s, d, \delta)$ *for* $k = \Omega(n/\log s)$ *and* $\mathcal{C}$ *the collection of all unbounded fan-in circuits of size* s *and depth* d*.*

Proof. We argue the contrapositive. Suppose there exists a set of parities $P = \{p_1(x, y), \ldots, p_r(x, y)\}$ violating Conjecture 6 with parameters s, d, δ. Observe that every $y \in \{0, 1\}^{n/2}$ we fix corresponds to a collection of r parities in x, and the union over y of all these parities results in a set of the form $P' = \{p'_1(x), \ldots, p'_r(x), p'_1(x) \oplus 1, \ldots, p'_r(x) \oplus 1\}$, where the p'_i's are parities in x. For a given k, the number of possible (nonempty) parities spanned by a subset of k parities from P' is bounded by $\binom{2r}{\leq k} \leq (2r + 1)^k$. For $k < (n/2)/\log(2s + 1)$, since $r \leq s$ we get

$$\binom{2r}{\leq k} \leq (2r + 1)^k \leq (2s + 1)^k < 2^{n/2},$$

which implies that there exists a nonempty parity π over $\{0, 1\}^{n/2}$ not spanned by any subset of k parities from P. This violates the assumed prediction variant. □

As mentioned earlier, $\mathsf{IPAP}(\mathsf{poly}(n), 2, 1)$ is known to hold; yet the question whether $\mathsf{IPAP}(\mathsf{poly}(n), d, \delta)$ holds for $d = 2$ in the average case or for $d \geq 3$ in the worst case, remains open. Thus, Proposition 4.7 gives a barrier for proving $\mathsf{PREDICTION}_{\oplus}(\mathsf{DNF}, \Omega(n/\log n), \delta)$ for some $\delta \in (0, 1]$, and on proving even the exact version $\mathsf{PREDICTION}_{\oplus}(\mathsf{AC}^0, \Omega(n/\log n), 1)$.

5 OR can distinguish between logarithmic-degree distributions

While our main conjectures are about fooling, in this section we consider the opposite goal.

Nisan and Szegedy [NS92] showed that the approximate degree of OR is $\Theta(\sqrt{n})$. As noticed by Bogdanov et al. [BIVW16], this implies (via LP duality) that for any $\epsilon > 0$ there exists a pair $\boldsymbol{X}, \boldsymbol{Y}$ of $\Theta_\epsilon(\sqrt{n})$-indistinguishable distributions which OR can $(1 - \epsilon)$-distinguish. This shows that Conjecture 3 fails for arbitrary sources even when k is as large as $\Theta(\sqrt{n})$. However, the two distributions $\boldsymbol{X}, \boldsymbol{Y}$ are not guaranteed to be simple.

In this section, we show how to reduce $\boldsymbol{X}, \boldsymbol{Y}$ to sources samplable by polynomial size decision trees, as well as to sources of degree $O_\epsilon(\log n)$, proving the following result.

Theorem 5.1. *(a) For any $\epsilon > 0$ there exists a pair $\boldsymbol{X}, \boldsymbol{Y}$ of $\Theta_\epsilon(\sqrt{n})$-indistinguishable sources over $\{0,1\}^n$ samplable by decision trees of size $O_\epsilon(n^3 \log^2 n)$ that the OR function $\mathsf{OR}(x) = x_1 \vee \cdots \vee x_n$ can $(1-\epsilon)$-distinguish.*

(b) For any $\epsilon > 0$ there exists a pair $\boldsymbol{X}, \boldsymbol{Y}$ of $\Theta_\epsilon(\sqrt{n})$-indistinguishable sources over $\{0,1\}^n$ of degree $O_\epsilon(\log n)$ that the OR function $\mathsf{OR}(x) = x_1 \vee \cdots \vee x_n$ can $(1-\epsilon)$-distinguish.

We prove this theorem in several steps:

1. The starting point is a pair $\boldsymbol{X}, \boldsymbol{Y}$ of $\Theta_\epsilon(\sqrt{n})$-indistinguishable sources over $\{0,1\}^{n'}$ which OR can $1-\epsilon'$ distinguish, where $\epsilon' < \epsilon$ and n' is linear in n.

2. We convert $\boldsymbol{X}, \boldsymbol{Y}$ into *mixtures of iid* $\boldsymbol{X'}, \boldsymbol{Y'}$ with similar properties. (We define this class of sources below.)

3. We show how to sample $\boldsymbol{X'}, \boldsymbol{Y'}$ using polynomial size decision trees. The sampling introduces a small error, which is identical for both sources.

4. We use a randomized encoding based on the Razborov–Smolensky lower bound technique to convert the decision trees into low-degree polynomials, introducing another small error which is identical for both sources.

Definition 5.2 (mixture of iid)**.** A source $\boldsymbol{X}$ on $\{0,1\}^n$ is a *mixture of iid* if it can be sampled using the following two step process:

1. Sample a bias $p \in [0,1]$ according to some finitely supported distribution $\mathcal{D}$.

2. Sample n independent p-biased bits (that is, each bit equals 1 with probability p).

The size of the decision trees constructed from $\boldsymbol{X'}, \boldsymbol{Y'}$ is related to the complexity of the distribution $\mathcal{D}$, a notion we define formally below.

The pair $\boldsymbol{X'}, \boldsymbol{Y'}$ can be used to give a very simple protocol for *visual secret sharing* [NS94], as we explain in Section 5.5.

[BIVW16] exhibits a pair of $\Omega_\epsilon(\sqrt{n})$-indistinguishable distributions which $\mathsf{OR}(x)$ can $(1-\epsilon)$-distinguish and is samplable by an AC^0 circuits of $\mathsf{poly}(n)$ size and depth 9. Our result achieves the same with an OR distinguisher while improving the depth of the AC^0 sampler to 2.

The following sections follow the various steps of the construction, culminating in the proof of Theorem 5.1. We complement the construction with a lower bound on the complexity of the distribution $\mathcal{D}$ in Section 5.6.

5.1 Constructing mixtures of iid

Our starting point is a pair $\boldsymbol{X}, \boldsymbol{Y}$ of indistinguishable sources which OR can distinguish. We will structure the construction in such a way that it suffices to keep track of a single distribution in this pair.

In order to facilitate the construction of small decision trees (and so low-degree polynomials) later on, we will need $\boldsymbol{X}, \boldsymbol{Y}$ to have low *complexity*, in the following senses.

Definition 5.3 (weight-complexity of source)**.** A source $\boldsymbol{X}$ over $\{0,1\}^n$ has *weight-complexity* L if for all $i \in \{0, \ldots, n\}$, the probability that $\boldsymbol{X}$ is a vector of Hamming weight i is an integer multiple of $1/L$.

Definition 5.4 (complexity of mixture of iid)**.** A mixture of iid $\boldsymbol{X}$, defined via a distribution $\mathcal{D}$ over biases, has *complexity* L if every p in the support of $\mathcal{D}$ is an integer multiple of $1/L$, and the probability that $\mathcal{D}$ equals p is an integer multiple of $1/L$. Its *support size* is the size of the support of $\mathcal{D}$.

We convert $\boldsymbol{X}$ into a mixture of iid using a simple resampling procedure, which is similar to *t-biased symmetrization* or *erase-all-subscripts symmetrization* (see [BT20a]).

Lemma 5.5. *Suppose $\boldsymbol{X}$ is a source over $\{0,1\}^n$ of weight-complexity L.*

For every $C > 0$ there is a source $\boldsymbol{X'}$ over $\{0,1\}^{Cn}$, which is a mixture of iid of complexity nL and support size $n+1$, such that

$$(1 - e^{-C}) \Pr[\mathsf{OR}(\boldsymbol{X}) = 1] \leq \Pr[\mathsf{OR}(\boldsymbol{X'}) = 1] \leq \Pr[\mathsf{OR}(\boldsymbol{X}) = 1].$$

Furthermore, if $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable then for every $C > 0$, the two sources $\boldsymbol{X'}, \boldsymbol{Y'}$ constructed in this way are also k-indistinguishable.

Proof. The source $\boldsymbol{X'}$ is sampled as follows. First, we sample $x \sim \boldsymbol{X}$. Second, we sample $m = Cn$ independent indices $i_1, \ldots, i_m$ uniformly distributed on $\{1, \ldots, n\}$, and output $x' = x_{i_1}, \ldots, x_{i_m}$. This is clearly a mixture of iid of complexity nL.

If $x = 0$ then $x' = 0$ always. In contrast, if $x \neq 0$ then the probability that $x_{i_j} = 1$ is at least $1/n$, and so the probability that $x' = 0$ is at most $(1 - 1/n)^{Cn} \leq e^{-C}$. Therefore the probability that $x' \neq 0$ is at least $1 - e^{-C}$.

Finally, suppose that $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable. We can sample $\boldsymbol{X'}, \boldsymbol{Y'}$ in tandem by first sampling $i_1, \ldots, i_m$ and then letting $\boldsymbol{X'} = \boldsymbol{X}|_{i_1,\ldots,i_m}$ and $\boldsymbol{Y'} = \boldsymbol{Y}|_{i_1,\ldots,i_m}$.

Now let $j_1, \ldots, j_k$ be any k indices in $\{1, \ldots, m\}$. The two distributions $\boldsymbol{X'}|_{j_1,\ldots,j_k} = \boldsymbol{X}|_{i_{j_1},\ldots,i_{j_k}}$ and $\boldsymbol{Y'}|_{j_1,\ldots,j_k} = \boldsymbol{Y}|_{i_{j_1},\ldots,i_{j_k}}$ are identical since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are k-indistinguishable. □

5.2 Constructing decision trees

In this step, we show how to approximately sample a mixture of iid using small decision trees. The sampling procedure has an error probability which can be made as small as desired. In case of a sampling error, we output the zero vector.

The main difficulty is that using decision trees, we can only sample exactly from *dyadic* distributions, that is, distributions in which the probability of each value is of the form $a/2^b$. We can overcome this difficulty by approximating a low-complexity distribution by a dyadic distribution.

First, a technical lemma about implementing a selection procedure using decision trees.

Lemma 5.6. *Consider an arbitrary partition of $\{0, \ldots, 2^n - 1\}$ into m intervals. The partition defines a function $f \colon \{0,1\}^n \to \{1, \ldots, m\}$.*

There is a decision tree of size at most $n(m-1)+1$ computing f.

Proof. The proof is by induction on n. If $n = 0$ or $m = 1$ then the result is trivial, so suppose that $n \geq 1$ and $m \geq 2$.

After querying the most significant bit of the input, we know that the input lies in either the left half or the right half of $\{0, \ldots, 2^n - 1\}$. At most one of the m original intervals can be broken into two, and so if the left half contains m_0 intervals and the right half contains m_1 intervals, then $m_0 + m_1 \leq m + 1$. According to the induction hypothesis, there are decision trees computing f restricted to the two halves of sizes at most $(n-1)(m_0-1)+1$ and $(n-1)(m_1-1)+1$. Combining them, we obtain a decision tree for f of size at most

$$(n-1)(m_0-1)+1+(n-1)(m_1-1)+1 \leq (n-1)(m-1)+2 \leq n(m-1)+1,$$

since by assumption $m \geq 2$. □

We can now show how to sample arbitrary distributions, with an arbitrarily small error.

Lemma 5.7. *Let $\mathcal{D}$ be a distribution in which the probability of each element in the support is an integer multiple of $1/L$, and furthermore the support has size s.*

For every $\delta > 0$ we can construct a decision tree T of size $O(\log(L/\delta)s)$ over $r \in \{0,1\}^N$ (for some N) whose leaves are labeled by elements in the support of $\mathcal{D}$ or by $\perp$, satisfying the following two properties (where $\boldsymbol{r}$ is uniformly distributed over $\{0,1\}^N$):

- *For every x in the support of $\mathcal{D}$,*
$$\Pr[T(\boldsymbol{r}) = x \mid T(\boldsymbol{r}) \neq \perp] = \Pr[\mathcal{D} = x].$$
- $\Pr[T(\boldsymbol{r}) = \perp] = \gamma$, *where $\gamma \leq \delta$ depends only on δ and L.*

Proof. Denote the support of $\mathcal{D}$ by $x_1, \ldots, x_s$, and suppose that x_i has probability p_i/L.

Let $N = \lceil \log_2(L/\delta) \rceil$, and let $K = \lfloor 2^N/L \rfloor$. We will construct a decision tree T that outputs x_i with probability $Kp_i/2^N$ and $\perp$ with probability $1 - KL/2^N < L/2^N \leq \delta$.

The decision tree interprets $r_1, \ldots, r_N$ as encoding a value $R \in \{0, \ldots, 2^N - 1\}$, and computes the following function:

$$f(R) = \begin{cases} x_i & \text{if } K(r_1 + \cdots + r_{i-1}) \leq R < K(r_1 + \cdots + r_i), \\ \perp & \text{if } R \geq 2^n - KL. \end{cases}$$

According to Lemma 5.6, we can compute f using a decision tree of size $O(Ns) = O(\log(L/\delta)s)$. This decision tree satisfies all required properties. □

Using this construction, we can show how to convert a mixture of iid into a source samplable by decision trees.

Lemma 5.8. *Suppose $\boldsymbol{X}$ is a mixture of iid over $\{0,1\}^n$ of complexity L and support size s.*

For every $\delta > 0$ we can construct n decision trees $T_1, \ldots, T_n$ of size $O(\log^2(nL/\delta)s)$ over $r \in \{0,1\}^N$ (for some N) such that

$$(T_1(\boldsymbol{r}), \ldots, T_n(\boldsymbol{r})) \sim \boldsymbol{X'},$$

where $\boldsymbol{r}$ is the uniform distribution over $\{0,1\}^N$, and

$$\Pr[\mathsf{OR}(\boldsymbol{X}) = 1] - \delta \leq \Pr[\mathsf{OR}(\boldsymbol{X'}) = 1] \leq \Pr[\mathsf{OR}(\boldsymbol{X}) = 1].$$

Furthermore, if $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable then for any $\delta > 0$, the two sources $\boldsymbol{X'}, \boldsymbol{Y'}$ constructed in this way are also k-indistinguishable.

Proof. Let $\boldsymbol{X}$ be sampled using the distribution $\mathcal{D}$ over biases. We construct the decision tree T_i in the following way. We partition the bit vector r into $n+1$ parts $r^{(0)}, r^{(1)}, \ldots, r^{(n)}$. We will use $r^{(0)}$ to sample the bias, and $r^{(i)}$ to sample $\boldsymbol{X'_i}$ given the bias.

We invoke Lemma 5.7 with $\delta_{\text{Lemma 5.7}} = \delta/(n+1)$ to construct a decision tree over $r^{(0)}$ which samples a bias from $\mathcal{D}$ with failure probability $\gamma' \leq \delta/(n+1)$. If the sampling fails (that is, we reach a leaf labeled $\perp$) then we output 0. Otherwise, if we sampled the bias p, then we invoke Lemma 5.7 with $s = 2$ and $\delta_{\text{Lemma 5.7}} = \delta/(n+1)$ to construct a decision tree over $r^{(i)}$ which samples a random bit with bias p, with failure probability $\gamma' \leq \delta/(n+1)$; if the sampling fails, then we output 0.

The distribution of $(T_1(\boldsymbol{r}), \ldots, T_n(\boldsymbol{r}))$ can be described as follows:

1. Start with a sample from $\boldsymbol{X}$.
2. Zero the entire source with probability $\delta/(n+1)$.
3. Zero each bit individually with probability $\delta/(n+1)$.

This description explains the relation between $\Pr[\mathsf{OR}(\boldsymbol{X}) = 1]$ and $\Pr[\mathsf{OR}(\boldsymbol{X'}) = 1]$, and also makes it clear that this operation preserves k-indistinguishability.

It remains to estimate the size of the decision trees T_i. Each such decision tree is obtained by taking a decision tree of size $O(\log(nL/\delta)s)$ and replacing each leaf with a decision tree of size $O(\log(nL/\delta))$, for an overall size of $O(\log^2(nL/\delta)s)$ □

5.3 Constructing low-degree polynomials

The final piece of the puzzle is converting the decision trees constructed by Lemma 5.8 into low-degree polynomials. Just like the conversion from a mixture of iid into decision trees, this conversion introduces an arbitrarily small error. The idea is to use a randomized encoding based on the technique used by Razborov [Raz87] and Smolensky [Smo87] in their celebrated circuit lower bound.

We start with the special case of the AND function.

Lemma 5.9. *For every n, d there is an $\mathbb{F}_2$ polynomial f of degree $2d$, over variables $x_1, \dots, x_n, r_1, \dots, r_N$ for some N, such that if $x_1 x_2 \cdots x_n = 1$ then $f(x, r) = 1$, and otherwise $\Pr_r[f(x, r) = 1] = 2^{-d}$.*

Proof. The polynomial is

$$f(x, r) = \prod_{i=1}^{d} \left(1 + \sum_{j=1}^{n} r_{i,j}(1 + x_j) \right).$$

If $x_1 \cdots x_n = 1$ then all of the sums are identically 0, and so all of the factors are identically 1. If $x_i = 0$ then each factor is distributed uniformly over $\mathbb{F}_2$, and so the probability that all factors are 1 is 2^{-d}. □

We can extend this to arbitrary decision trees by representing a decision tree as a sum of ANDs. (The same idea works for any unambiguous DNF.)

Lemma 5.10. *Let T be a decision tree over $\{0, 1\}^m$ of size at most s. For every $\delta > 0$ there is a polynomial f of degree $O(\log(s/\delta))$ such that if $f(x) = b$ then*

$$\Pr_r[f(x, r) \neq b] = \gamma_b$$

for some $\gamma_b \leq \delta$, where γ_0, γ_1 depend only on s, δ.

Proof. Let Λ be the set of all leaves of T labelled 1. For each leaf $\ell \in \Lambda$, let T_ℓ be the indicator function of reaching that leaf. Thus T_ℓ is a product of literals, and as functions, $T = \sum_{\ell \in \Lambda} T_\ell$. Moreover, in this sum, at most one summand is 1 on any input.

Let $d = \lceil \log(s/\delta) \rceil$, so that $s/2^d \leq \delta$. Using Lemma 5.9, we construct a polynomial f_ℓ of degree $O(\log(s/\delta))$ such that $f_\ell(x, r) = 1$ whenever $T_\ell(x) = 1$, and otherwise $\Pr_r[f_\ell(x, r) = 1] = 2^{-d}$. For $t \in \{1, \dots, s - |\Lambda|\}$, let $g_t = \prod_{i=1}^{d} r_i$. We take

$$f(x, r) = \sum_{\ell \in \Lambda} f_\ell(x, r^{(\ell)}) + \sum_{t=1}^{s-|\Lambda|} g_t(x, r^{(t)}),$$

where r is partitioned into s parts $r^{(\ell)}, r^{(t)}$.

If $T(x) = 0$ then the probability that $f(x, r) = 1$ is exactly the probability that the sum of s many 2^{-d}-biased random bits is odd, which is at most $s/2^d$.

Similarly, if $T(x) = 1$ then the probability that $f(x, r) = 0$ is exactly the probability that the sum of $s - 1$ many 2^{-d}-biased random bits is odd, which is also at most $s/2^d$. □

5.4 Proof of main theorem

We now put everything together. As our starting point, we use a construction of $\Theta_\epsilon(\sqrt{n})$-indistinguishable sources which OR can $(1 - \epsilon)$-distinguish due to Bun and Thaler [BT13].

Theorem 5.11. *For any $\epsilon > 0$ there is a pair of $\Omega(\sqrt{\epsilon n})$-indistinguishable sources on $\{0, 1\}^n$, of weight-complexity $n^{O(n)}$, which OR can $(1 - \epsilon)$-distinguish.*

Proof. Bun and Thaler define the following twisted polynomial:

$$Q(x) = (-1)^{x+s} \frac{c^{2m}(m!)^2}{n!} \prod_{\substack{j \in \{2,\ldots,n\} \\ j \notin \{ck^2 : 1 \le k \le m\}}} (x - j),$$

where $s \in \{0, 1\}$, $c = \lceil 8/\epsilon \rceil$, and $m = \lfloor \sqrt{n/c} \rfloor$. They show (Proposition 14) that

$$\sum_{x=0}^{n} Q(x) x^d = 0$$

for $0 \le d \le k := \Omega(\sqrt{\epsilon n})$, and

$$\frac{Q(0) - \sum_{x=1}^{n} Q(x)}{\sum_{x=0}^{n} |Q(x)|} \ge 1 - \epsilon.$$

Since $\sum_{x=0}^{n} Q(x) = 0$, this implies that

$$\frac{2Q(0)}{\sum_{x=0}^{n} |Q(x)|} \ge 1 - \epsilon.$$

Let $S = \sum_{x=0}^{n} |Q(x)|/2$, and define

$$Q_+(x) = \begin{cases} Q(x)/S & \text{if } Q(x) > 0, \\ 0 & \text{otherwise,} \end{cases} \qquad Q_-(x) = \begin{cases} -Q(x)/S & \text{if } Q(x) < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\sum_{x=0}^{n} Q(x) = 0$ we clearly have $\sum_{Q(x)>0} Q(x) = \sum_{Q(x)<0} (-Q(x)) = S$, and so Q_+, Q_- are probability distributions. Moreover, if $0 \le d \le k$ then

$$\sum_{x=0}^{n} Q_+(x) x^d - \sum_{x=0}^{n} Q_-(x) x^d = \sum_{x=0}^{n} \frac{Q(x)}{S} x^d = 0.$$

We define two probability distributions P_+, P_- on $\{0,1\}^n$ as follows. To sample $P_\pm$, first sample x according to $Q_\pm$, and then sample a uniformly random vector of weight x. These will function as our $\boldsymbol{X}, \boldsymbol{Y}$.

We claim that P_+, P_- are k-indistinguishable. To see this, let $i_1, \ldots, i_k$ be any k coordinates, let $b_1, \ldots, b_k \in \{0, 1\}$, and let $h = b_1 + \cdots + b_k$. Then

$$\Pr[P_+|_{i_1,\ldots,i_k} = (b_1, \ldots, b_k)] = \sum_{x=0}^{n} Q_+(x) \frac{\binom{n-k}{x-h}}{\binom{n}{x}} = \frac{(n-k)!}{n!} \sum_{x=0}^{n} Q_+(x) x^{\underline{h}} (n-x)^{\underline{k-h}},$$

where $x^{\underline{h}} = x(x-1)\cdots(x-h+1)$. The coefficient next to $Q_+(x)$ is a polynomial of degree k, and so we get an identical probability if we replace P_+, Q_+ by P_-, Q_-.

We claim that OR can $(1-\epsilon)$-distinguish P_+, P_-. Indeed, assume without loss of generality that $Q(0) > 0$. Then $Q_+(0) = Q(0)/S \ge 1 - \epsilon$, whereas $Q_-(0) = 0$. Therefore $\Pr[\mathsf{OR}(P_+) = 0] \ge 1 - \epsilon$ while $\Pr[\mathsf{OR}(P_-) = 0] = 0$.

It remains to bound the weight-complexity of P_+, P_-. The probability that P_+ or P_- has Hamming weight x is either 0 or $|Q(x)|/S$. In the latter case, it is equal to

$$\frac{2 \prod_{j \in U} |x - j|}{\left| \sum_{y=0}^{n} (-1)^y \prod_{j \in U} (y - j) \right|},$$

where U is some subset of $\{2, \ldots, n\}$. Thus P_+, P_- have weight-complexity L, where

$$L = \left| \sum_{y=0}^{n} (-1)^y \prod_{j \in U} (y - j) \right|.$$

Each of the summands is at most $n!$ in magnitude, and so we can bound $L \le n \cdot n! = n^{O(n)}$. □

We now put everything together.

Proof of Theorem 5.1. Let $C = \lceil \ln(4/\epsilon) \rceil$, so that $1 - e^{-C} \le \eta/4$. Apply Theorem 5.11 to get a pair $\boldsymbol{X}, \boldsymbol{Y}$ of $\Theta_\epsilon(\sqrt{n})$-indistinguishable sources over $\{0,1\}^{n/C}$ of complexity $n^{O(n)}$ which OR can $(1-\epsilon/4)$-distinguish.

Apply Lemma 5.5 to get a pair $\boldsymbol{X'}, \boldsymbol{Y'}$ of mixtures of iid over $\{0,1\}^n$ of complexity $n^{O(n)}$ and support size $n/C + 1$ which OR can $(1 - 2\epsilon/4)$-distinguish.

Apply Lemma 5.7 (with $\delta = \eta/4$) to get a pair $\boldsymbol{X''}, \boldsymbol{Y''}$ of $\Theta_\epsilon(\sqrt{n})$-indistinguishable sources samplable by decision trees of size $O_\epsilon(n^3 \log^2 n)$ which OR can $(1 - 3\epsilon/4)$-distinguish.

Apply Lemma 5.10 (with $\delta = \epsilon/(8n)$) to get a pair $\boldsymbol{X'''}, \boldsymbol{Y'''}$ of $\Theta_\epsilon(\sqrt{n})$-indistinguishable sources of degree $O_\epsilon(\log n)$ which OR can $(1-\epsilon)$-distinguish. (We chose the parameter δ so that the randomized encoding embodied in Lemma 5.10 changes the output with probability at most $\epsilon/2$.) □

We outline an alternative construction of $\boldsymbol{X'}, \boldsymbol{Y'}$ in Appendix A. The advantage of this construction over Lemma 5.5 is that the distributions $\mathcal{D}$ are supported on only $k+1$ different biases, where k is the indistinguishability parameter. An appeal to linear programming duality allows strengthening Lemma 5.5 to give the same promise.

5.5 How to share an image, infinitely

Komargodski, Naor, and Yogev [KNY16] studied the information complexity of *evolving* threshold secret sharing. In this model, a stateful dealer that does not know the number of parties in advance assigns the shares in sequence. They showed the existence of an n-threshold scheme in which the t-th party is assigned $(n-1)\log t + o_n(\log t)$ bits, and showed that this is optimal for $n = 2$.

Our Theorem 5.11 and Lemma 5.5 give a simple evolving ramp secret sharing scheme with single bit shares and imperfect reconstruction with error ϵ. The state of the dealer consists of a single probability $p \in [0,1]$ with $O(k \log k)$ bits of precision for secrecy parameter k. Sampling a share consists of tossing a p-biased coin. Unlike in the proposals of Komargodski et al., the dealer's state does not need to be updated, so no synchronization is necessary in a distributed implementation. The reconstruction threshold is $n = Ck^2/\epsilon$ for some absolute constant C.

Moreover, owing to the one-sidedness of the error and the visual nature of our scheme, the "contrast" improves as more parties become involved in reconstruction. If the scheme is used to share an image described by a set B of black pixels, as the size n of the reconstruction set tends to infinity, the reconstructed image will contain all pixels in B with probability approaching one, while every pixel outside B will appear independently with probability approaching ϵ.

5.6 Lower bound on precision

One obstacle to a potential improvement of the degree of the sources in Theorem 5.1 is the exponential weight-complexity of the sources of Bun and Thaler in Theorem 5.11. While the existence of k-indistinguishable sources of weight complexity polylogarithmic in k that OR can distinguish would not immediately improve Theorem 5.1, it would at least obviate the need for the randomized encoding step in Lemma 5.10, which partially accounts for the logarithmic degree in our construction. We show that this exponential complexity is unavoidable.

Theorem 5.12. *The weight-complexity of any pair of distinct k-indistinguishable distributions over $\{0,1\}^n$ is at least* $\exp(\Omega(k))$.

The same conclusion holds for the complexity of any pair of distinct k-indistinguishable mixtures of iid by a similar proof.

Theorem 5.12 does not forbid the existence of low-complexity samplers for the two distributions. For example, the uniform distributions over n-bit strings of parity 0 and 1 are $(n-1)$-indistinguishable, have weight complexities 2^{n-1}, but are 2-locally samplable.[6] We interpret Theorem 5.12 as a limitation of the

[6] Vectors of parity b can be sampled as $r_1, r_1 \oplus r_2, \ldots, r_{n-2} \oplus r_{n-1}, r_{n-1} \oplus b$, a construction attributed to [Bab87, BL86].

proof method for Theorem 5.1, and not as evidence against the existence of samplers of sublogarithmic degree.

We now prove Theorem 5.12. Given two distributions $\boldsymbol{X}, \boldsymbol{Y}$ of weight-complexity L over $\{0,1\}^n$, let a_i be the coefficient

$$a_i = L\big(\Pr[\mathrm{wt}(\boldsymbol{X}) = i] - \Pr[\mathrm{wt}(\boldsymbol{Y}) = i]\big), \qquad 0 \le i \le n,$$

where wt stands for Hamming weight. By the distinctness and weight-complexity assumptions, $(a_0, \dots, a_n)$ is a nonzero vector in the integer lattice $\mathbb{Z}^{n+1}$.

Claim 5.13. *If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are k-indistinguishable then*

$$\sum_{i=0}^{n} a_i p(i) = 0 \quad \text{for every polynomial } p \text{ of degree at most } k. \tag{1}$$

Proof. The advantage of the distinguisher that checks whether all among a random set of t inputs evaluate to 1 is

$$\sum_{i=0}^{n} \frac{a_i}{L} \cdot \frac{i}{n} \cdot \frac{i-1}{n-1} \cdots \frac{i-t+1}{n-t+1}.$$

By k-indistinguishability,

$$\sum_{i=0}^{n} a_i \cdot i(i-1)\cdots(i-t+1) = 0 \qquad \text{for all } 1 \le t \le k.$$

This equality also holds for $t = 0$ as a_i/L is the difference of two distributions. The conclusion (1) follows because the functions $1, i, i(i-1), \dots, i(i-1)\cdots(i-k+1)$ span all degree-k polynomials. □

The main technical result of this section is the following:

Proposition 5.14. *If (1) holds then $\sum_{i=0}^{n} a_i^2 = \exp \Omega(k)$.*

Proof of Theorem 5.12. Since a_i/L represents the difference of two distributions, the 1-norm $\sum_{i=0}^{n} |a_i|$ can be at most $2L$. By Proposition 5.14, $\exp \Omega(k) \le \sum a_i^2 \le (\sum |a_i|)^2 \le 4L^2$, from which $L = \exp \Omega(k)$. □

We now prove Proposition 5.14. By the shift-invariance of (1) we may assume without loss of generality that $a_0 \ne 0$. In the unrealistic case $n \le 2(k+1)$, Proposition 5.14 easily follows from the fact that the $\exp(-\Omega(k))$-approximate degree of OR on n bits is at most k:

Fact 5.15 ([dW08]). *For every k there exists a univariate polynomial p such that $p(0) = 1$ and $|p(i)| \le \exp(-\Omega(k))$ for all $1 \le i \le 2(k+1)$.*

Plugging this p into (1) we get that

$$|a_0| = \left|\sum\nolimits_{i=1}^{n} a_i p(i)\right| \le n \max\{|a_1|, \dots, |a_n|\} \cdot \exp(-\Omega(k)),$$

so because a_0 is a nonzero integer, one of the other a_i must be at least $\exp \Omega(k)$.

The case of general n essentially reduces to this, but we do not know of a direct way of proving this without a detour into lattices. Let $A \subseteq \mathbb{Z}^{n+1}$ be the lattice of integer solutions $(a_0, \dots, a_n)$ to (1).

Claim 5.16. *The vectors $v_0, \dots, v_{n-k-1} \in \mathbb{Z}^{n+1}$ given by $v_{ij} = (-1)^{j-i}\binom{k+1}{j-i}$, where $0 \le j \le n$, are a basis of the lattice A. ($\binom{\star}{i}$ is zero for negative i.)*

Proof. First we argue that all v_i are in the lattice A. The polynomial $Dp(x) = p(x) - p(x+1)$ has strictly lower degree than p. If p has degree k then $D^{k+1}p(x) = \sum (-1)^j \binom{k+1}{j} p(x+j)$ vanishes. Taking $x = i$ we conclude from (1) that $v_0, \dots, v_{n-k-1}$ all belong to the lattice A.

Now we argue that any solution to (1) can be written as an integer combination of the v_i. First, the dimension of the solution space is precisely $n-k$. The matrix with rows $v_0, \dots, v_{n-k-1}$ is upper triangular with identity diagonal, so each integer vector in the solution space can be expressed as an integer combination of the v_i. □

Claim 5.17. *The length of the projection of v_i on the subspace orthogonal to $v_0, \dots, v_{i-1}$ is minimized when $i = k+1$.*

Proof. v_i and $v_{i'}$ are orthogonal whenever $|i - i'| > k+1$ as they have disjoint support. Therefore the projection of interest does not depend on $v_0, \dots, v_{i-k-2}$. Since v_i is a shift of v_0 ($v_{ij} = v_{0(j-i)}$), the length of the projection stays the same for all $i \geq k+1$. The minimum must therefore be attained at some i between 0 and $k+1$. Among these values of i, the projection of v_i on the span of $v_0, \dots, v_{i-1}$ is the same as the projection of v_{i-1} on the span of $v_0, \dots, v_{i-2}$ together with an additional vector v_{-1}, so the length of the projection of v_i onto the subspace spanned by $v_0, \dots, v_{i-1}$ does not decrease with i. Therefore projection on the orthogonal space does not increase with i and so must attain its minimum at $i = k+1$. □

The 2-norm of any nonzero vector in A, including $(a_0, \dots, a_n)$, is lower-bounded by the shortest vector in the Gram–Schmidt orthogonalization of the lattice basis $v_0, \dots, v_{n-k-1}$ (see e.g. [MG02, Theorem 1.1]), which by Claim 5.17 is precisely the projection of v_{k+1} onto the subspace orthogonal to $v_0, \dots, v_k$. For the purpose of lower bounding this projection we may and will assume without loss of generality that $n = 2(k+1)$.

Claim 5.18. *Assume $n = 2(k+1)$. The subspace of $\mathbb{R}^{n+1}$ dual to $v_0, \dots, v_k$ is spanned by all degree-k polynomials and the vector $\delta \in \mathbb{R}^{n+1}$ given by $\delta_j = 1$ when $j = n$ and 0 when $0 \leq j < n$.*

Proof. Since $v_0, \dots, v_k$ belong to A, the subspace of interest must contain all degree-k polynomials. Since $v_{i(n+1)} = 0$ for all $0 \leq i \leq k$, it must also contain δ. As δ has n zeroes, its degree is strictly larger than k, so it is linearly independent of the degree-k polynomials. By the rank-nullity theorem, the dimension of the subspace is $(n+1) - (k+1) = k+2$ so it must be spanned by the degree-k polynomials and δ. □

Proof of Proposition 5.14. By [MG02, Theorem 1.1] and Claim 5.17, it is sufficient to prove that the orthogonal projection of v_{k+1} to the subspace dual to $v_0, \dots, v_k$ has 2-norm at least $\exp \Omega(k)$. Since the length of this projection is the same for all $n \geq 2(k+1)$, we may assume that $n = 2(k+1)$. By Claim 5.18, it is then sufficient to lower bound the projection of v_{k+1} by the subspace spanned by δ and all degree-k polynomials.

By Fact 5.15 there exists a degree k polynomial p such that $\delta - p$ has infinity-norm at most $\exp(-\Omega(k))$ and therefore 2-norm at most $\sqrt{2(k+1)}\exp(-\Omega(k))$. The projection of v_{k+1} onto the line spanned by $\delta - p$ has then magnitude at least

$$\frac{|\langle v_{k+1}, \delta - p\rangle|}{\|\delta - p\|} = \frac{|\langle v_{k+1}, \delta\rangle - \langle v_{k+1}, p\rangle|}{\|\delta - p\|} \geq \frac{\left|(-1)^{k+1} - 0\right|}{\sqrt{2(k+1)}\exp(-\Omega(k))} = \exp \Omega(k). \qquad \square$$

6 Negative results

In this section we prove several special cases of Conjecture 2 and Conjecture 3, using the concept of *predictability*. We concentrate on sources of low degree or locality, and the circuit classes we consider are the OR function, decision trees (and more generally, unambiguous DNFs), and local DNFs (a generalization of narrow DNFs).

We describe predictability in Section 6.1, where we also work out the toy case of degree 1 sources. A more complicated example is local sources, worked out in Section 6.2. We show that constant degree sources are weakly predictable in Section 6.3, and that degree 2 sources are strongly predictable in Section 6.4. Contrasting these results, we show in Section 6.5 that depth 1 sources (sources computed by AND functions) are not predictable. Using an ad hoc argument, we show that indistinguishable depth 1 sources are *statistically* indistinguishable.

Finally, in Section 6.6 we use another notion of predictability to prove Conjecture 2 for linear sources and local DNFs (disjunctions of arbitrary functions of $O(1)$ input bits).

Our proof that quadratic sources are strongly predictable uses Dickson's theorem, a structure theorem for quadratic polynomials. Haramaty and Shpilka [HS10] proved a similar structure theorem for cubic polynomials, leaving the tantalizing possibility of extending our proof to cubic sources.

Source	Predictability	Reference	Section
Degree 1	$\log(1/\epsilon)$	Lemma 6.2	Section 6.1
Degree 1	$s2^s \log(1/\epsilon)$	Theorem 6.43	Section 6.6
Degree 2	$\log^{10}(1/\epsilon)$	Theorem 6.16	Section 6.4
Degree d	$O_{d,\epsilon}(1)$	Corollary 6.15	Section 6.3
s-local	$(1/\epsilon)^{O(s2^s)}$	Lemma 6.10	Section 6.2
Depth 1	$\log\log(n/\epsilon)^*$	Theorem 6.33	Section 6.5

Table 2: Summary of negative results. For each type of source, we give a value of k such that this class of sources is $(O(k), \epsilon)$-predictable (see Definition 6.1). This implies that $\mathsf{MAIN}(O(k), \epsilon)$ holds for this class of sources and OR distinguishers by Corollary 6.6.
If k involves s, then $\mathsf{MAIN}(O(k), \epsilon)$ holds for distinguishers which are ORs of s-local functions.
If $k = \mathsf{polylog}(1/\epsilon)$ then $\mathsf{GENERAL}(\mathsf{polylog}(1/\epsilon), \epsilon)$ holds for this class of sources and polynomial size unambiguous DNF distinguishers by Lemma 6.8.
For depth 1 sources, the stated bound is not predictability, but rather the parameter k for which $\mathsf{MAIN}(k, \epsilon)$ holds.

6.1 Predictability

Intuitively, a source is *predictable* if we can predict the value of the OR function by looking at a small number of bits. We formalize this in the following way.

Definition 6.1 (predictability)**.** Let $\boldsymbol{X}$ be a source over $\{0,1\}^n$. The source is (k,ϵ)-*predictable* if there exists a subset $S \subseteq \{1,\ldots,n\}$ of size at most k such that

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] \leq \epsilon.$$

(Here $\boldsymbol{X} = 0$ means that $\boldsymbol{X}$ is the zero vector.)

A set S with this property ϵ-*predicts* $\boldsymbol{X}$.

As a simple example, let us show that degree 1 sources are $(O(\log_2(1/\epsilon)), \epsilon)$-predictable.

Lemma 6.2. *Any degree* 1 *source is* $(\lceil \log_2(1/\epsilon) \rceil + 1, \epsilon)$*-predictable for any* $\epsilon > 0$*.*

Here and elsewhere in this section, we will omit floors and ceilings for brevity, and logarithms will always be base 2.

Proof. Let $\boldsymbol{X}$ be a degree 1 source. We consider several cases.

Case 1. The source has dimension at least $\log(1/\epsilon)$. In this case, let S be an arbitrary set of $\log(1/\epsilon)$ affinely independent coordinates of $\boldsymbol{X}$. Then S ϵ-predicts $\boldsymbol{X}$, since

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] \leq \Pr[\boldsymbol{X}|_S = 0] = 2^{-|S|} = \epsilon.$$

Case 2. The source has dimension at most $\log(1/\epsilon)$. In this case, let S be a set of at most $\log(1/\epsilon)$ coordinates such that every coordinate of $\boldsymbol{X}$ is an affine combination of $\boldsymbol{X}_i$ for $i \in S$.

Case 2a. Every coordinate of $\boldsymbol{X}$ is a *linear* combination of $\boldsymbol{X}_i$ for $i \in S$. In this case,

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] = 0.$$

Case 2b. We have $\boldsymbol{X}_j = \sum_{i\in T} \boldsymbol{X}_i + 1$, where $j \notin S$ and $T \subseteq S$. In this case,

$$\Pr[\boldsymbol{X}|_{S\cup\{j\}} = 0 \text{ and } \boldsymbol{X} \neq 0] \leq \Pr[\boldsymbol{X}|_{S\cup\{j\}} = 0] = 0. \qquad \square$$

We now give two ways of using predictability to show that indistinguishable sources fool OR: one which is useful for proving special cases of Conjecture 3, and one which is useful for proving special cases of Conjecture 2.

Lemma 6.3. *Suppose that $\boldsymbol{X}, \boldsymbol{Y}$ are $2k$-indistinguishable sources which are (k,ϵ)-predictable. Then $\boldsymbol{X}, \boldsymbol{Y}$ ϵ-fool OR.*

Proof. Let S, T be sets of at most k coordinates satisfying

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] \leq \epsilon,$$
$$\Pr[\boldsymbol{Y}|_T = 0 \text{ and } \boldsymbol{Y} \neq 0] \leq \epsilon.$$

Let $R = S \cup T$. It is easy to check that the two inequalities above still hold if we replace S, T with R. Therefore

$$\Pr[\boldsymbol{X}|_R = 0] \geq \Pr[\boldsymbol{X} = 0] = \Pr[\boldsymbol{X}|_R = 0] - \Pr[\boldsymbol{X}|_R = 0 \text{ and } \boldsymbol{X} \neq 0] \geq \Pr[\boldsymbol{X}|_R = 0] - \epsilon.$$

Since $\boldsymbol{X}, \boldsymbol{Y}$ are $2k$-indistinguishable and $|R| \leq 2k$, we have $\Pr[\boldsymbol{X}|_R = 0] = \Pr[\boldsymbol{Y}|_R = 0]$, and so the lemma follows. □

Lemma 6.4. *Suppose that $\boldsymbol{X}, \boldsymbol{Y}$ are $(k+1)$-indistinguishable sources, and $\boldsymbol{Y}$ is $(k,\epsilon/n)$-predictable. Then $\boldsymbol{X}, \boldsymbol{Y}$ ϵ-fool OR.*

Proof. Let S be a set of at most k coordinates satisfying

$$\Pr[\boldsymbol{Y}|_S = 0 \text{ and } Y \neq 0] \leq \epsilon.$$

As in the proof of the preceding lemma,

$$\Pr[\boldsymbol{Y}|_S = 0] \geq \Pr[\boldsymbol{Y} = 0] \geq \Pr[\boldsymbol{Y}|_S = 0] - \epsilon/n.$$

We will show that

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] \leq \epsilon,$$

and so

$$\Pr[\boldsymbol{X}|_S = 0] \geq \Pr[\boldsymbol{X} = 0] \geq \Pr[\boldsymbol{X}|_S = 0] - \epsilon.$$

This will complete the proof, since $\Pr[\boldsymbol{X}|_S = 0] = \Pr[\boldsymbol{Y}|_S = 0]$ by k-indistinguishability.

To prove the remaining inequality, we use a simple union bound:

$$\begin{aligned}\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] &\leq \sum_{i\notin S} \Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X}_i = 1] \\ &= \sum_{i\notin S} \Pr[\boldsymbol{Y}|_S = 0 \text{ and } \boldsymbol{Y}_i = 1] \leq n \Pr[\boldsymbol{Y}|_S = 0 \text{ and } \boldsymbol{Y} \neq 0] \leq \epsilon. \qquad \square\end{aligned}$$

We will be able to prove predictability results in two different regimes, corresponding to the following two definitions.

Definition 6.5 (weak and strong predictability)**.** A class of sources is *weakly predictable* if for every $\epsilon > 0$ there exists a constant k such that every source in the class is (k,ϵ)-predictable.

A class of sources is *strongly predictable* if there exists a polynomial p such that for every $\epsilon > 0$, every source in the class is $(p(\log(1/\epsilon)),\epsilon)$-predictable.

For example, Lemma 6.2 shows that degree 1 sources are strongly predictable. In Section 6.4 we show that degree 2 sources are also strongly predictable, and in Section 6.3 we show that constant degree sources are weakly predictable.

There are two meta-theorems relating predictable classes of sources and our two main conjectures.

Corollary 6.6. *If a class of sources is weakly predictable then* $\mathsf{MAIN}(O_\epsilon(1), \epsilon)$ *holds for this class of sources and OR distinguishers (on any subset of coordinates).*

Corollary 6.7. *If a class of sources is strongly predictable then* $\mathsf{GENERAL}(\mathsf{polylog}(n/\epsilon), \epsilon)$ *holds for this class of sources and OR distinguishers (on any subset of coordinates).*

Let us now show how to extend the latter corollary to unambiguous DNFs (and so, in particular, to decision trees).

Lemma 6.8. *If a class of sources is strongly predictable and closed under negations of coordinates then* $\mathsf{GENERAL}(\mathsf{polylog}(ns/\epsilon), \epsilon)$ *holds for this class of sources and distinguishers which are unambiguous DNFs with at most s clauses (for example, decision trees of size s).*

Proof. Let p be the polynomial associated with the class of sources, and let $\boldsymbol{X}, \boldsymbol{Y}$ be $(p(\log(s/\epsilon)) + 1)$-indistinguishable sources, where $\boldsymbol{Y}$ belongs to the class.

Denote the DNF by $f = f_1 + \cdots f_s$ (here addition is over the reals), where $f_1, \ldots, f_s$ are disjoint ANDs. Since the class of sources is closed under negations, Lemma 6.4 shows that $\boldsymbol{X}, \boldsymbol{Y}$ ϵ/s-fool each f_i. Therefore

$$|\mathsf{E}[f(\boldsymbol{X})] - \mathsf{E}[f(\boldsymbol{Y})]| \leq \sum_{i=1}^{s} |\mathsf{E}[f_i(\boldsymbol{X})] - \mathsf{E}[f_i(\boldsymbol{Y})]| \leq \epsilon. \qquad \square$$

6.2 Local sources

In this subsection we show that local sources are weakly predictable. Our argument actually works for *local-over-linear* sources, defined next.

Definition 6.9 (local-over-linear sources)**.** A source is *s-local-over-linear* if every bit is a function of s many degree 1 polynomials.

In the rest of this subsection, we prove the following result.

Lemma 6.10. *For every s, the class of s-local-over-linear sources is weakly predictable: every s-local-over-linear source is $((1/\epsilon)^{O(s2^s)}, \epsilon)$-predictable.*

The proof is by induction on s. For every s, ϵ, we will show that an s-local-over-linear source is $(c(s, \epsilon), \epsilon)$-predictable, for some constant $c(s, \epsilon) = (1/\epsilon)^{O(s2^s)}$.

When $s = 1$, an s-local-over-linear is just a degree 1 source, and so we can take $c(1, \epsilon) = \log(1/\epsilon) + 1$ by Lemma 6.2.

Suppose now that $s > 1$. Suppose that $\boldsymbol{X}_i$ depends on the degree 1 polynomials J_i, where $|J_i| \leq s$. Without loss of generality, the polynomials in J_i are affinely independent. Also, we can assume that no $\boldsymbol{X}_i$ is the constant zero, since such coordinates do not affect predictability.

Let I be a maximal set of coordinates such that the multiset $\bigcup_{i \in I} J_i$ is affinely independent. We consider two cases, according to the size of I.

Case 1. The set I contains at least $2^s \log(1/\epsilon)$ coordinates. Let $S \subseteq I$ be a subset of size exactly $2^s \log(1/\epsilon)$. Since $\boldsymbol{X}_i$ is not identically zero, $\Pr[\boldsymbol{X}_i = 0] \leq 1 - 2^{-|J_i|} \leq 1 - 2^{-s}$. Since the sets J_i are affinely independent,

$$\Pr[\boldsymbol{X}|_S = 0] = \prod_{i \in S} \Pr[\boldsymbol{X}_i = 0] \leq (1 - 2^{-s})^{|S|} \leq \epsilon.$$

Therefore $\boldsymbol{X}$ is $(2^s \log(1/\epsilon), \epsilon)$-predictable.

Case 2. The set I contains at most $2^s \log(1/\epsilon)$ coordinates. Let $J = \bigcup_{i \in I} J_i$, and let J' be the linear parts of the polynomials in J.

Let V be the span of all r_j appearing in all J_i. Decompose V into $\operatorname{span}(J') + U$. Thus every polynomial in every J_i can be written uniquely in the form $P + Q + b$, where $P \in \operatorname{span}(J')$, $Q \in U$, and $b \in \mathbb{F}_2$.

For every assignment α to the polynomials in J', we can consider the source $\boldsymbol{X}^\alpha$ which is obtained by replacing each $P + Q + b$ by $P(\alpha) + Q + b$ (using the terminology of the preceding paragraph).

The source $\boldsymbol{X}^\alpha$ is clearly an s-local-over-linear source. We claim that it is in fact $(s-1)$-local-over-linear. To see this, first note that if $i \in I$ then $\boldsymbol{X}^\alpha_i$ is constant. If $i \notin I$ then, by definition of I, there is some affine dependence in $J \cup J_i$, which must involve some polynomial in J_i, say $R_1 + \cdots + R_k + P = b$, where $R_1, \ldots, R_k \in J_i$, $P \in \operatorname{span}(J')$, and $b \in \mathbb{F}_2$. In $\boldsymbol{X}^\alpha$ this implies that $R_1 = R_2 + \cdots + R_k + P_k(\alpha) + b$, and so we can rewrite $\boldsymbol{X}^\alpha_i$ as depending on at most $s - 1$ polynomials.

By induction, each source $\boldsymbol{X}^\alpha$ is $(c(s-1,\epsilon),\epsilon)$-predictable, say witnessed by a set S_α. Take $S = \bigcup_\alpha S_\alpha$. Then

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] \le 2^{-J} \sum_\alpha \Pr[\boldsymbol{X}^\alpha|_{S_\alpha} \text{ and } \boldsymbol{X}^\alpha \neq 0] \le \epsilon.$$

The set S has size

$$2^{|J|} c(s-1,\epsilon) \le 2^{s2^s \log(1/\epsilon)} c(s-1,\epsilon) = (1/\epsilon)^{s2^s} c(s-1,\epsilon),$$

and so considering both cases together,

$$c(s,\epsilon) \le \max\left[2^s \log(1/\epsilon), (1/\epsilon)^{s2^s} c(s-1,\epsilon)\right].$$

Solving the recurrence, we obtain

$$c(s,\epsilon) = (1/\epsilon)^{O(s2^s)}.$$

Note If $\boldsymbol{X}$ is s-local then the argument showing that $\boldsymbol{X}^\alpha$ is $(s-1)$-local is simpler: the set I is a hitting set for the sets J_i, that is, $I \cap J_i \neq \emptyset$ for all i, and so it is clear that if we fix the values of the coordinates in I, then the resulting source is $(S-1)$-local.

6.3 Constant degree sources

In this subsection we show that sources of constant degree are weakly predictable, using higher-order Fourier analysis. The proof is very similar to the proof of the regularity lemma from higher-order Fourier analysis.

The only result we need from higher-order Fourier analysis is the following.

Definition 6.11 (regular factor)**.** A set S of $\mathbb{F}_2$ polynomials is *r-singular* if there exists a non-empty subset $T \subseteq S$ such that $\sum_{P \in T} P$ can be written as a function of r many polynomials of degree smaller than d, where d is the maximum degree of a polynomial in T.

A set of $\mathbb{F}_2$ polynomials is *r-regular* if it is not *r-singular*.

Theorem 6.12 ([HHL19, Lemma 7.24])**.** *For every $d, \epsilon > 0$ there is a constant $r = r(d,\epsilon)$ such that if $\{Q_1, \ldots, Q_m\}$ is an r-regular set of $\mathbb{F}_2$ polynomials of degree at most d then for all $b_1, \ldots, b_m \in \mathbb{F}_2$,*

$$2^{-m} - \epsilon \le \Pr[Q_1 = b_1, \ldots, Q_m = b_m] \le 2^{-m} + \epsilon.$$

The proof that degree d sources are weakly predictable proceeds by induction on a somewhat cumbersome well-ordered poset. We will need several definitions.

Definition 6.13 (local function)**.** Let $\lambda \in \mathbb{N}^d$. A function is *λ-local* if it is a function of $\lambda_1 + \cdots + \lambda_d$ many $\mathbb{F}_2$ polynomials, where the i'th group consists of λ_i many polynomials of degree at most i.

For a subset $\Lambda \subseteq \mathbb{N}^d$, a function is *$\Lambda$-local* if it is λ-local for some $\lambda \in \Lambda$.

A source is *λ-local* or *Λ-local* if each coordinate is.

We will prove the following result by induction.

Lemma 6.14. *For any non-empty finite Λ, the class of Λ-local sources is weakly predictable.*

Corollary 6.15. *For every d, the class of degree d sources is weakly predictable.*

Proof. Take $\Lambda = \{(0, \dots, 0, 1)\}$. □

In the rest of this subsection, we prove Lemma 6.14, in the following form: for every non-empty finite Λ and $\epsilon > 0$, every Λ-local source is $c(\Lambda, \epsilon)$-predictable for some constant $c(\Lambda, \epsilon)$. The proof is by induction on $\max \Lambda$, where the maximum is taken with respect to lexicographic ordering: $\lambda \prec \mu$ if $\lambda_j = \mu_j$ for all $j > i$, and $\lambda_i < \mu_i$. It is well-known that this is a well-ordering.

The base case is when $\max \Lambda = (0, \dots, 0)$, or equivalently, $\Lambda = \{(0, \dots, 0)\}$. In this case, the source is constant, and so it is trivially $(1, 0)$-predictable: either all coordinates are constant 0, or else, any constant 1 coordinate perfectly predicts the source.

Suppose now that $\max \Lambda \succ (0, \dots, 0)$. For $\lambda \in \Lambda$, let $|\lambda| = \lambda_1 + \cdots + \lambda_d$, and let $|\Lambda| = \max_{\lambda \in \Lambda} |\lambda|$.

Suppose that $\boldsymbol{X}_i$ is a λ_i-local function depending on the polynomials J_i. We can assume that $\boldsymbol{X}_i$ is not constant zero, since such coordinates do not affect predictability.

Let I be a maximal set of coordinates such that the multiset $\bigcup_{i \in I} J_i$ is $r(d, \delta)$-regular, where δ is defined as follows: $m = 2^{|\Lambda|} \log(2/\epsilon)$, and $\delta = 2^{-|\Lambda| m}(\epsilon/2)$. We consider two cases, according to the size of I.

Case 1. The set I contains at least m coordinates. Let $S \subseteq I$ be a subset of size exactly m. For each $i \in S$, since $\boldsymbol{X}_i$ is not identically zero, we can find some assignment α_i to the polynomials in J_i which causes $\boldsymbol{X}_i = 1$.

Let α be an assignment for $J = \bigcup_{i \in S} J_i$. According to Theorem 6.12, since J is $r(d, \delta)$-regular,

$$\Pr[J = \alpha] \leq 2^{-|J|} + \delta.$$

Call an assignment α *bad* if $\alpha|_{J_i} \neq \alpha_i$ for all $i \in S$. Thus

$$\Pr[\boldsymbol{X}|_S = 0] \leq \sum_{\alpha \text{ bad}} (2^{-|J|} + \delta) \leq \prod_{i \in S} (1 - 2^{-|J_i|}) + 2^{|J|}\delta \leq (1 - 2^{-|\Lambda|})^m + 2^{|\Lambda| m}\delta.$$

We chose m and δ so that each term is at most $\epsilon/2$, and so we conclude that S ϵ-predicts $\boldsymbol{X}$.

Case 2. The set I contains at most m coordinates. Let $J = \bigcup_{i \in I} J_i$. Let α be any assignment to the polynomials in J such that $\boldsymbol{X}|_I = 0$.

By definition of I, for each $i \notin I$ the multiset $J \cup J_i$ is not $r(d, \delta)$-regular. Therefore there must be a non-empty subset $A_i \subseteq J_i$ and a subset $B_i \subseteq J$ such that $\sum_{P \in A_i} P + \sum_{Q \in B_i} Q$ can be written as a function of at most $r(d, \delta)$ many polynomials of degree less than e, where e is the maximal degree of a polynomial in $A_i \cup B_i$. In particular, if $J = \alpha$ then we can rewrite $\boldsymbol{X}_i$ as a coordinate $\boldsymbol{X}_i^\alpha$ which is λ_i'-local, where $\lambda_i' \prec \lambda_i$ is obtained by subtracting 1 from the e'th index and (if $e > 1$) adding $r(d, \delta)$ to the $(e-1)$'th index. This creates a new source $\boldsymbol{X}^\alpha$ which is Λ'-local for some finite set Λ' satisfying $\max \Lambda' \prec \max \Lambda$.

By induction, every source $\boldsymbol{X}^\alpha$ is $2^{-|\Lambda| m}\epsilon$-predicted by a set S^α of size at most $c(\Lambda', 2^{-|\Lambda| m}\epsilon)$. Let $S = I \cup \bigcup_\alpha S^\alpha$. Then

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] \leq \sum_{\alpha \colon \boldsymbol{X}|_I = 0} \Pr[\boldsymbol{X}^\alpha|_{S^\alpha} = 0 \text{ and } \boldsymbol{X}^\alpha \neq 0] \leq \epsilon.$$

This shows that

$$c(\Lambda, \epsilon) \leq m + 2^{|\Lambda| m} c(\Lambda', 2^{-|\Lambda| m}\epsilon).$$

Note One might be tempted to use ϵ rather than $2^{-|\Lambda| m}\epsilon$, like in the preceding section. However, if we condition on $J = \alpha$, we no longer get a Λ'-source. Instead, we apply a union bound: if $\boldsymbol{X}|_S = 0$ and $\boldsymbol{X} \neq 0$ then $\boldsymbol{X}^\alpha|_S = 0$ and $\boldsymbol{X}^\alpha \neq 0$ for some α, namely the value of the polynomials in J.

6.4 Quadratic sources

In this subsection we prove that for quadratic sources, the conclusion of Corollary 6.15 can be improved to strong predictability.

Theorem 6.16. *The class of quadratic sources is $(O(\log^{10}(1/\epsilon)), \epsilon)$-predictable.*

Our argument is structured in two parts: A proof of the statement for a special subclass of quadratic sources that we call low-rank bilinear, and a reduction of the general case to this special case.

Definition 6.17. A set of quadratics $\mathcal{Q}$ is (r, d)-*bilinear* if there exists a partition of the common variables into two disjoint sets $\boldsymbol{x}$ and $\boldsymbol{x'}$ such that

- Every $q \in \mathcal{Q}$ can be written as $q(\boldsymbol{x}, \boldsymbol{x'}) = \sum_{i=1}^{r_q} y_{iq}(\boldsymbol{x}) y'_{iq}(\boldsymbol{x'})$, where $r_q \leq r$ and $y_{1q}, \dots, y_{r_q q}$ and $y'_{1q}, \dots, y'_{r_q q}$ are linear forms that are linearly independent.
- The span of all the linear forms $\{y_{iq} \colon 1 \leq i \leq r_q, q \in \mathcal{Q}\}$ has dimension at most d.

The linear independence condition can be assumed without loss of generality:

Fact 6.18. *If $q(\boldsymbol{x}, \boldsymbol{x'}) = \sum_{i=1}^{r} y_i(\boldsymbol{x}) \cdot y'_i(\boldsymbol{x'})$ for some linear forms y_i, y'_i then there exist linearly independent linear forms $z_i \in \operatorname{span}\{y_1, \dots, y_r\}$, $z'_i \in \operatorname{span}\{y'_1, \dots, y'_r\}$ such that $q(\boldsymbol{x}, \boldsymbol{x'}) = \sum_{i=1}^{s} z_i(\boldsymbol{x}) \cdot z'_i(\boldsymbol{x'})$, where $s \leq r$.*

Proposition 6.19. *All (r, d)-bilinear quadratic sources are $(rd^2 \log(d/\epsilon), d\epsilon)$-predictable.*

Proposition 6.20. *If all $(O(\log^5(1/\epsilon)), O(\log^2(1/\epsilon)))$-bilinear quadratic sources are $(k, \epsilon/16)$-predictable then all quadratic sources are $(k + O(\log^4(1/\epsilon)), \epsilon)$-predictable.*

Proof of Theorem 6.16. By Proposition 6.19, all $(O(\log^5(1/\epsilon)), O(\log^2(1/\epsilon)))$-bilinear quadratic sources are $(O(\log^{10}(1/\epsilon)), \epsilon/32)$-predictable. By Proposition 6.20 all quadratic sources are $(O(\log^{10}(1/\epsilon)), \epsilon)$-predictable. □

The proofs of both propositions rely on the following reducibility property of predictable sources:

Claim 6.21. *Let X, X' and Y be jointly distributed bit sequences. Assume X' is (k, ϵ)-predictable and $\Pr[X = 0$ and $Y \neq X'] \leq \delta$. Then (X, Y) is $(k + |X|, \epsilon + \delta)$-predictable.*

Proof. Let S be the set of indices of X and T be the set of indices that (k, ϵ)-predicts X'. Then

$$\begin{aligned}
\Pr[(X,Y)|_{S \cup T} = 0, (X,Y) \neq 0] &= \Pr[X = 0, Y|_T = 0, (X,Y) \neq 0] \\
&= \Pr[X = 0, Y|_T = 0, Y \neq 0] \\
&\leq \Pr[X = 0, Y|_T = 0, Y \neq 0, Y = X'] + \Pr[X = 0, Y \neq X'] \\
&\leq \Pr[X'|_T = 0, X' \neq 0] + \Pr[X = 0, Y \neq X'] \\
&= \epsilon + \delta.
\end{aligned}$$

□

Proof of Proposition 6.19

We first illustrate the proof in the special case $r = 1$ and $y'_{1q}(\boldsymbol{x'}) = x'_q$. Then all quadratics in this source have the form

$$q(\boldsymbol{x}, \boldsymbol{x'}) = y_q(\boldsymbol{x}) x'_q.$$

Let $\emptyset = A_0 \subset A_1 \subset A_2 \subset \cdots$ be an increasing sequence of subsets of $\mathcal{Q}$ such that the linear forms $Y_i = \{y_q \colon q \in A_i \setminus A_{i-1}\}$ are a basis of $\{y_q \colon q \in \mathcal{Q} \setminus A_{i-1}\}$. Then $|A_i \setminus A_{i-1}| \leq d$ for all i. If the sequence has length at most $k = \lfloor \log(d/\epsilon) \rfloor$ then $|\mathcal{Q}| \leq d \log(d/\epsilon)$ and predictability follows trivially. Otherwise we argue that A_k predicts $\mathcal{Q}$ except with probability ϵ.

Claim 6.22. *Let $Y_1, \dots, Y_k$ be (multi)sets of vectors and V be a subspace such that $\operatorname{span} Y_1 \supseteq \operatorname{span} Y_2 \supseteq \dots \supseteq \operatorname{span} Y_k \supseteq V$. Then for a uniformly random subset $\boldsymbol{S}$ of the disjoint union $Y_1 \uplus \dots \uplus Y_k$,*

$$\Pr\big[V \not\subseteq \operatorname{span} \boldsymbol{S}\big] \leq (\dim V) \cdot 2^{-k}.$$

We apply Claim 6.22 to $V = \operatorname{span}\{y_q \colon q \notin A_k\}$ and $\boldsymbol{S} = \{y_q \colon x'_q = 1\}$ to conclude that except with probability at most $d2^{-k} \leq \epsilon$, the forms $\{y_q \colon q \in A_k, x'_q = 1\}$ span all of V. When $\mathcal{Q}|_{A_k} = 0$ all these forms must vanish. Then V must also vanish and $\mathcal{Q}|_{\overline{A}_k} = 0$, so $\mathcal{Q}$ is (k, ϵ)-predictable.

Proof of Claim 6.22. Let $\boldsymbol{S}_i = Y_i \cap \boldsymbol{S}$ and $\boldsymbol{V}_i = V \cap \operatorname{span}(\boldsymbol{S}_1 \uplus \dots \uplus \boldsymbol{S}_i)$. Then $0 = \boldsymbol{V}_0 \subseteq \boldsymbol{V}_1 \subseteq \dots \subseteq \boldsymbol{V}_k \subseteq V$. Since Y_i spans V, for every fixing of $\boldsymbol{S}_1, \dots, \boldsymbol{S}_{i-1}$ there must exist a subset $\boldsymbol{Y}_i \subseteq Y_i$ of linearly independent vectors that satisfies the direct sum decomposition $V = \boldsymbol{V}_{i-1} \oplus \operatorname{span} \boldsymbol{Y}_i$, from which

$$\dim \operatorname{span} \boldsymbol{Y}_i = \operatorname{codim}_V \boldsymbol{V}_{i-1}. \tag{2}$$

Since $\boldsymbol{Y}_i \cap \boldsymbol{S}_i$ is a uniformly random subset of $\boldsymbol{Y}_i$ and the vectors in $\boldsymbol{Y}_i$ are linearly independent, we have

$$\mathsf{E}[\dim \operatorname{span}(\boldsymbol{Y}_i \cap \boldsymbol{S}_i) \mid \boldsymbol{S}_1, \dots, \boldsymbol{S}_{i-1}] = \frac{\dim \operatorname{span} \boldsymbol{Y}_i}{2}. \tag{3}$$

Since $\boldsymbol{V}_i$ contains $\boldsymbol{V}_{i-1} \oplus \operatorname{span}(\boldsymbol{Y}_i \cap \boldsymbol{S}_i)$,

$$\dim \boldsymbol{V}_i \geq \dim \boldsymbol{V}_{i-1} + \dim \operatorname{span}(\boldsymbol{Y}_i \cap \boldsymbol{S}_i).$$

Then

$$\begin{aligned}
\mathsf{E} \operatorname{codim}_V \boldsymbol{V}_i &\leq \mathsf{E} \operatorname{codim}_V \boldsymbol{V}_{i-1} - \mathsf{E} \dim \operatorname{span}(\boldsymbol{Y}_i \cap \boldsymbol{S}_i) && \text{by linearity of expectation} \\
&= \mathsf{E} \operatorname{codim}_V \boldsymbol{V}_{i-1} - (\mathsf{E} \dim \operatorname{span} \boldsymbol{Y}_i)/2 && \text{by (3)} \\
&= (\mathsf{E} \operatorname{codim}_V \boldsymbol{V}_{i-1})/2 && \text{by (2).}
\end{aligned}$$

Since $\boldsymbol{V}_0 = 0$ has codimension $\dim V$, iterating for k rounds and applying Markov's inequality gives

$$\Pr[\operatorname{codim}_V \boldsymbol{V}_k \neq 0] \leq \mathsf{E} \operatorname{codim}_V \boldsymbol{V}_k \leq (\dim V) \cdot 2^{-k}.$$

Therefore, except with probability $(\dim V) \cdot 2^{-k}$, $\boldsymbol{V}_k$ must equal V. □

In general the forms y'_{iq} may not be linearly independent. The proof strategy is to isolate sufficiently many linearly independent forms among them so that Claim 6.22 can be applied. If this fails the space of forms $\{y'_{iq} \colon 1 \leq i \leq r_q, q \in \mathcal{Q}\}$ must have small dimension. All of $\mathcal{Q}$ then has small dimension and is therefore predictable.

Fact 6.23. *If $v_1, \dots, v_m$ are linearly independent vectors and $w_1, \dots, w_m$ are arbitrary vectors, then $a_1 v_1 + w_1, \dots, a_m v_m + w_m$ are linearly independent for some $a \in \{0,1\}^m$.*

Proof. Since w_1 and $v_1 + w_1$ span v_1, there must be a choice for a_1 so that $a_1 v_1 + w_1, v_2, \dots, v_m$ are linearly independent. By the same reasoning, $a_1 v_1 + w_1, a_2 v_2 + w_2, v_3, \dots, v_m$ are linearly independent for some a_2. Continuing all the way to m proves the fact. □

Proof of Proposition 6.19. We prove the proposition by strong induction on d. When $d = 0$ the source is constant so it is $(0,0)$-predictable. We now assume the proposition is true for all $d' < d$ and show that it is true for d.

Let $Y = \{y_{iq} \colon i \leq r_q, q \in \mathcal{Q}\}$ and $Y' = \{y_{iq} \colon i \leq r_q, q \in \mathcal{Q}\}$. Let $B \subseteq \mathcal{Q}$ be a maximal set with the following property: For each $q \in B$ there exists an index $i_q \leq r_q$ such that $y'_{i_q q}$ is not in span of $Y'_{-q} = \{y'_{iq'} \colon i \leq r_{q'}, q' \in B \setminus \{q\}\}$.

For a given $q \in B$, the forms $\{y'_{iq} \colon i \le r_q\}$ may be linearly dependent modulo Y'_{-q}. To address that, we do a change of basis, moving from $Y_i = \{y'_{1q}, \dots, y'_{r_q q}\}$ to another basis $z'_{1q}, \dots, z'_{r_q q}$ formed by joining a basis C_1 of $\operatorname{span}(Y_i) \cap \operatorname{span}(Y'_{-q})$ and a basis C_2 of its dual with respect to $\operatorname{span}(Y_i)$, rewriting

$$q = \sum_{i=1}^{r_q} z_{iq} z'_{iq}.$$

We can construct the new basis so that $z'_{i_q q} = y'_{i_q q}$. Modulo Y'_{-q}, C_1 goes to zero, while C_2 remains linearly independent. Since $z'_{i_q q} \in C_2$, this shows that $z'_{i_q q}$ cannot be written as a linear combination of $Y'_{-q} \cup \{z'_{iq} \colon i \ne i_q\}$. Therefore the forms $\{z'_{i_q q} \colon q \in B\}$ are linearly independent modulo the span of

$$Y'_{-} = \{z'_{iq} \colon q \in B, i \ne i_q\}.$$

Let $\emptyset = A_0 \subset A_1 \subset A_2 \subset \cdots$ be an increasing sequence of subsets of B so that the linear forms $Z_i = \{z_{i_q q} \colon q \in A_i \setminus A_{i-1}\}$ are linearly independent and span all of $\{z_{i_q q} \colon q \in \mathcal{Q} \setminus A_{i-1}\}$.

If the length of the sequence is at most $k = \lfloor \log(d/\epsilon) \rfloor$, then $|B| \le \sum |A_i \setminus A_{i-1}| \le dk$. By the maximality of B, all the forms $\{y'_{iq} \colon q \notin B\}$ are in the span of $\{y'_{iq} \colon q \in B\}$, so span Y' can have dimension at most dkr, and $\mathcal{Q}$ viewed as a linear space of quadratic forms can have dimension at most d^2kr. Any basis of this space $(kd^2r, 0)$-predicts $\mathcal{Q}$.

Otherwise, let $V = \operatorname{span}\{z_{iq} \colon q \in \mathcal{Q} \setminus A_k\}$. For every $q \in B$, every assignment a'_{-} to span Y'_{-}, and every assignment a'_{+} to $\{z'_{i_q q} \colon q \in B\}$, if we substitute a'_{+} and a'_{-} then q becomes a *linear* form

$$z_{i_q q} a'_{+}(z'_{i_q q}) + \sum_{i \ne i_q} z_{iq} a'_{-}(z'_{iq}).$$

For every $i \le k$, $\operatorname{span}\{z_{i_q q} \colon q \in A_i \setminus A_{i-1}\}$ contains V. Fact 6.23 shows that for every assignment a'_{-} we can find an assignment b'_{+} so that the quadratic forms in $A_i \setminus A_{i-1}$ simplify under a'_{-}, b'_{+} to linear forms spanning V, for every $i \le k$.

Sample Y' in two stages. First, sample Y'_{-} to get an assignment a'_{-}, and let b'_{+} be the assignment constructed above. Then, sample an assignment a'_{+} to Y'_{+}. Let $\boldsymbol{S}$ be the random subset of A_k corresponding to indices where $a'_{+} = b'_{+}$. By Claim 6.22, except with probability $d2^{-k} = \epsilon$ over the assignment to Y', the forms $\{z_{i_q q} \colon q \in A_k\}$ span all of V. Therefore $\Pr[\mathcal{Q}|_{A_k} = 0 \text{ and } V \ne 0] \le \epsilon$. The source $\mathcal{Q} \bmod V$ is $(r, d - |V|)$-bilinear for $|V| \ne 0$, so by the inductive hypothesis it is $(r(d-1)^2 \log((d-1)/\epsilon), (d-1)\epsilon)$-predictable. By Claim 6.21 we get that $\mathcal{Q}$ is (k_d, ϵ)-predictable, where

$$k_d = \max\bigl\{rd^2 \log(d/\epsilon), |A_k| + r(d-1)^2 \log((d-1)/\epsilon)\bigr\} = rd^2 \log(d/\epsilon). \qquad \square$$

Proof of Proposition 6.20

To prove Proposition 6.20, we gradually refine the class of quadratic sources whose strong predictability is sufficient to establish Theorem 6.16.

Our starting point is a quantitative version of Theorem 6.12 for quadratic sources, for which it can be shown that $r(2, \epsilon) \le 2\log(1/\epsilon) + 1$. We do not rely on this bound explicitly but instead use an essentially equivalent characterization of the *rank* of quadratic maps, a notion that gives more structural information about the quadratic map.

Fact 6.24 (Dickson's Theorem [MS78]). *For every quadratic polynomial q there exist a unique number r and bits c, d such that*

$$q(\boldsymbol{x}) = y_1 y_2 + \cdots + y_{2r-1} y_{2r} + c y_{2r+1} + d, \tag{4}$$

where $\boldsymbol{y} = (y_1, \dots, y_{2r+c}) = A\boldsymbol{x} + b$ for some matrix A with independent rows.

The value r is called the *rank* of q. The *bias* of a random variable X is $\mathsf{E}[(-1)^X] = \Pr[X = 1] - \Pr[X = 0]$.

Claim 6.25. $|\text{bias}(q)| \le 2^{-\text{rank}(q)}$.

Proof. Since the transformation A preserves the output distribution of q, the rank determines the bias $\mathsf{E}[(-1)^{q(\boldsymbol{x})}]$ of the polynomial q:

$$\text{bias}(q) = \mathsf{E}[(-1)^{y_1y_2+\cdots+y_{2r-1}y_{2r}+cy_{2r+1}+d}] = (-1)^d\mathsf{E}[(-1)^{cy_{2r+1}}]\prod_{i=1}^{r}\mathsf{E}[(-1)^{y_{2i-1}y_{2i}}] = (1-c)(-1)^d 2^{-r}.$$

Therefore $|\text{bias}(q)| = (1-c)2^{-r} \le 2^{-r}$. □

We first show that it is sufficient to prove predictability for quadratic sources of small rank. The rank of a quadratic source is the maximum rank of all the quadratics in it.

Claim 6.26 (Rank reduction)**.** *If all quadratic sources of rank at most* $\log(2/\epsilon)$ *are* (k,ϵ)*-predictable then all quadratic sources are* $(k+\log(2/\epsilon),\epsilon)$*-predictable.*

We will need the following fact about pseudorandomness of small-biased random variables with respect to point functions:

Fact 6.27. *For every random variable* X *over* $\{0,1\}^n$, $|\Pr[X=0]-2^{-n}| \le \max_{\varnothing\neq T\subseteq[n]} \left|\text{bias}\sum_{i\in T} X_i\right|$.

Proof.
$$\Pr[X=0] = \mathsf{E}\prod_{i=1}^{n}\frac{1+(-1)^{X_i}}{2} = \sum_{T\subseteq[n]} 2^{-n}\mathsf{E}\prod_{i\in T}(-1)^{X_i} = 2^{-n} + 2^{-n}\sum_{T\neq\varnothing}\text{bias}\sum_{i\in T}X_i.$$

By the triangle inequality, $|\Pr[X=0]-2^{-n}| \le 2^{-n}(2^n-1)\max_{T\neq\varnothing}\left|\text{bias}\sum_{i\in T}X_i\right|$. □

Proof of Claim 6.26. Let $\mathcal{Q}$ be an arbitrary quadratic source. If $\mathcal{Q}$ contains a subset A of more than $\log(2/\epsilon)$ quadratics all of whose nontrivial sums have rank more than $\log(2/\epsilon)$, by Claim 6.25 bias $\sum_{i\subset T} q_i \le \epsilon/2$ for every nonempty subset T of A. By Fact 6.27,

$$\Pr[q=0 \text{ for all } q\in A] \le 2^{-|A|} + \max_{T\neq\varnothing}\left|\text{bias}\sum\nolimits_{q\in T} q\right| < \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon,$$

so $\mathcal{Q}$ is $(\log(2/\epsilon),\epsilon)$-predictable.

Otherwise, let A be a maximal subset of quadratics all of whose nontrivial sums have rank more than $\log(2/\epsilon)$. By maximality, for every quadratic $p\notin A$ there exists a nonempty subset T of A such that $p' = p + \sum_{q\in T} q$ has rank at most $\log(2/\epsilon)$. Let $\mathcal{P}$ be the set of all such p'. When $\mathcal{Q}|_A = 0$, $\mathcal{P} = \mathcal{Q}|_{\overline{A}}$. By Claim 6.21, $\mathcal{Q}$ is $(|A\cup B|,\epsilon)$-predictable. □

We next argue that with a small loss in parameters, the source can be "homogenized" in the sense that in the definition of rank the change of variable map is linear and the constants b and c are set to zero. We say a source $\mathcal{Q}$ has *homogeneous rank* at most R if every $q\in\mathcal{Q}$ can be written as $q(\boldsymbol{x}) = y_1(\boldsymbol{x})y_1'(\boldsymbol{x}) + \cdots + y_t(\boldsymbol{x})y_t'(\boldsymbol{x})$ for some linear and linearly independent $y_1, y_1', \dots, y_r, y_r'$ and $r\le R$.

Claim 6.28 (Homogenization)**.** *If all quadratic sources of homogeneous rank at most* $R+2$ *are* $(k,\epsilon/16)$*-predictable then all quadratic sources of rank at most* R *are* (k,ϵ)*-predictable.*

Proof. Given a rank r source $\mathcal{Q}$, we define a new source $\mathcal{Q}'$ whose seed consists of the seed of $\mathcal{Q}$ plus four new seed inputs u, s, t. By Fact 6.24, each $q\in\mathcal{Q}$ of rank $r\le R$ can be written in the form (4). We include in $\mathcal{Q}'$ the corresponding quadratic

$$q' = y_1'y_2' + \cdots + y_{2r-1}'y_{2r}' + cy_{2r+1}'v + dst,$$

where $\boldsymbol{y}' = (y_1',\dots,y_{2r+1}') = A\boldsymbol{x} + bu$. To check that q' has homogeneous rank at most $r+2$ we verify that the linear forms $y_1',\dots,y_{2r+1}', v, s, t$ are linearly independent. Independence of the y_i' follows from the full rank of A, and v, s, t are independent from the rest because these variables do not appear in the y'_i. Since $\mathcal{Q}'$ projects to $\mathcal{Q}$ when $u = v = s = t = 1$,

$$\Pr[\mathcal{Q}|_S = 0, \mathcal{Q} \neq 0] = \Pr[\mathcal{Q}'|_S = 0, \mathcal{Q}' \neq 0 \mid u = v = s = t = 1] \leq \frac{\Pr[\mathcal{Q}'|_S = 0, \mathcal{Q}' \neq 0]}{\Pr[u = v = s = t = 1]} = \frac{\epsilon/16}{1/16} = \epsilon$$

for the set S that witnesses the (k, ϵ)-predictability of $\mathcal{Q}'$. □

To complete the proof we reduce low-rank homogeneous sources to low-rank bilinear sources:

Claim 6.29 (Bilinearization). *If all $(O(R^3 \log^2 1/\epsilon), O(R \log 1/\epsilon))$-bilinear sources are (k, ϵ)-predictable then all homogeneous sources of rank R are $(k + O(R^2 \log^2 1/\epsilon), \epsilon)$-predictable.*

To prove Claim 6.29 we will use the following expansion of homogeneous quadratics. Given a direct sum decomposition $L \oplus M$ of the space of linear forms over the variables $\boldsymbol{x}$ we can uniquely expand every linear form y as $y^L + y^M$ with $y^L \in L$ and $y^M \in M$. A homogeneous quadratic form $q = \sum_{i=1}^r y_i y'_i$ can then be expanded as

$$q = \sum_{i=1}^r (y_i^L + y_i^M)(y'^L_i + y'^M_i) = \sum_{i=1}^r y_i^M y'^M_i + \sum_{i=1}^r (y_i^M y'^L_i + y_i^L y'^M_i) + \sum_{i=1}^r y_i^L y'^L_i. \tag{5}$$

We will say that q *linearizes modulo* L if $q \bmod L \equiv 0$, that is if the first summand $\sum_{i=1}^r y_i^M y'^M_i$ in this decomposition vanishes.

Claim 6.30. *If q does not linearize modulo L then for every fixing a of all the linear forms in L, the probability that $q \bmod (L + a)$ evaluates to zero is at most $3/4$.*

Proof. The degree of q modulo $L + a$ as a quadratic form over M is exactly two if and only if the first term in (5) does not vanish. Claim 6.25 shows that $q \bmod (L+a)$ has bias between $-1/2$ and $1/2$, and so evaluates to zero with probability at most $3/4$. □

Proof of Claim 6.29. Let $\mathcal{Q}$ be a rank R homogeneous source. Let A be a minimal subset of $\mathcal{Q}$ for which all the forms $q \in \mathcal{Q} \setminus A$ linearize modulo the span L of the linear forms in A.

If $|A| \geq k_0 = \log_{4/3}(1/\epsilon)$, take an arbitrary subset $\{q_1, \dots, q_{k_0}\}$ of A, and let L_i be the span of the linear forms in q_i. By the minimality of A, q_i does not linearize modulo $L_1 + \dots + L_{i-1}$ for any $i \leq k_0$. By Claim 6.30,

$$\Pr[q_i = 0 \mid q_1, \dots, q_{i-1}] \leq \max_a \Pr[q_i = 0 \mid L_1 + \dots + L_{i-1} = a] \leq \tfrac{3}{4},$$

so $\Pr[q_1 = \dots = q_{k_0} = 0] \leq (3/4)^{k_0} \leq \epsilon$. The set $\{q_1, \dots, q_{k_0}\}$ then witnesses the (k_0, ϵ)-predictability of A.

If $|A| < k_0$ then L has dimension at most $k_0 R$, so $\operatorname{span}(A)$, viewed as a linear space of quadratic forms, can have dimension at most $(k_0 R)^2$. For every $q \in \mathcal{Q}$, the leading term in the summation (5) vanishes, giving a simplified decomposition

$$q = q^L + q^{LM} \qquad \text{where} \qquad q^L = \sum_{i=1}^r y_i^L y'^L_i, \qquad q^{LM} = \sum_{i=1}^r (y_i^M y'^L_i + y_i^L y'^M_i).$$

Since $q^L \in \operatorname{span}(A)$, there must exist a subset B of $\mathcal{Q}$ of size at most $(k_0 R)^2$ for which the set of forms $\{q^L : q \in B\}$ form a basis of $\operatorname{span}(A)$. Then for every $p \in \mathcal{Q}$, p^L can be expressed as a linear combination $p^L = \sum_{q \in B'_p} q^L$ for some $B'_p \subseteq B$. For every $p \in \mathcal{Q} \setminus B$ define

$$p' = p^{LM} + \sum_{q \in B'_p} q^{LM},$$

and let $\mathcal{P}$ be the source consisting of all p' for $p \in \mathcal{Q} \setminus B$.

Under an invertible change of variables that identifies some bases of linear forms in L and M with new variables $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively, each p' becomes bilinear. The dimension of the span of all forms that depend on $\boldsymbol{x}$ is at most $\dim L \le k_0 R$ and the rank of each p' is at most $(|B'_p| + 1)R \le 2k_0^2 R^3$. By assumption, $\mathcal{P}$ is (k, ϵ)-predictable. If $\mathcal{Q}|_B = 0$ then

$$p' = p^{LM} + \sum_{q \in B'_p} q^{LM} = p^{LM} + \sum_{q \in B'_p} q^L = p^{LM} + p^L = p,$$

and therefore $\mathcal{Q}|_{\overline{B}} = \mathcal{P}$. By Claim 6.21, $\mathcal{Q}$ is $(k + |B|, \epsilon)$-predictable. □

Proof of Proposition 6.20. Set $R = \log(32/\epsilon) + 2$ and apply Claim 6.29, Claim 6.28, and Claim 6.26 in sequence. □

6.5 Depth 1 sources

In this subsection we give a simple example of a class of sources which is not predictable.

Definition 6.31 (depth 1 sources)**.** A source $\boldsymbol{X}$ on $\{0,1\}^n$ is a *depth* 1 *source* if $\boldsymbol{X}_i$ is a conjunction or disjunction of literals over the r_j.

Lemma 6.32. *Fix a constant $\epsilon > 0$. For each n there is a value m such that the following source on $\{0,1\}^n$ is (k, ϵ)-predictable only if $k = \Omega_\epsilon(n)$:*

$$\boldsymbol{X}_i = \bigwedge_{j=1}^{m} r_{i,j}.$$

Proof. Let S be an arbitrary set of size s. Then

$$\Pr[\boldsymbol{X}|_S = 0 \text{ and } \boldsymbol{X} \neq 0] = (1 - 2^{-m})^s [1 - (1 - 2^{-m})^{n-s}] \ge \left(\exp -O\left(\frac{s}{2^m}\right)\right) \cdot \left(1 - \exp -\frac{n-s}{2^m}\right).$$

If this is at most ϵ, then at least one of the factors is at most $\sqrt{\epsilon}$. Therefore either $s = O_\epsilon(2^m)$, or $n - s = \Omega_\epsilon(2^m)$. Choosing $2^m = \Theta_\epsilon(n)$, we obtain the lemma. □

In the rest of this subsection, we give an ad hoc argument showing that indistinguishable depth 1 sources fool *arbitrary* distinguishers.

Theorem 6.33. *If $\boldsymbol{X}, \boldsymbol{Y}$ are two $(\log\log(n/\epsilon) + 2)$-indistinguishable depth 1 sources then the statistical distance between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is at most ϵ.*

Let us start by classifying the coordinates of $\boldsymbol{X}$ or $\boldsymbol{Y}$ into four types:

- Constant coordinates.
- Balanced coordinates, i.e., $\Pr[\boldsymbol{X}_i = 1] = 1/2$.
- Conjunctive coordinates, i.e., $\Pr[\boldsymbol{X}_i = 1] < 1/2$.
- Disjunctive coordinates, i.e., $\Pr[\boldsymbol{X}_i = 1] > 1/2$.

For the sake of proving Theorem 6.33, we can ignore constant coordinates, since a coordinate is constant in $\boldsymbol{X}$ iff it is constant in $\boldsymbol{Y}$. By complementing conjunctive coordinates, we can assume that all coordinates are balanced or conjunctive, and so expressible as a conjunction. We call such a source a *conjunctive source*.

Definition 6.34 (conjunctive sources)**.** A source $\boldsymbol{X}$ on $\{0,1\}^n$ is *conjunctive* if each coordinate is a conjunction of literals over the r_j.

The *width* of a conjunctive source is the maximal number of literals in any conjunction.

We will show that every conjunctive source is statistically close to a narrow source, and that two indistinguishable narrow sources are equidistributed. We start with the latter, which follows almost directly from the following result of Amano, Iwama, Maruoka, Matsuo, and Matsuura [AIM$^+$03], which we translate from disjunctive sources to conjunctive sources.

Theorem 6.35 ([AIM$^+$03, Theorem 3.1])**.** *Let $k \geq 2$, and define $\lambda_k = \lfloor \log_2 k \rfloor + 2$.*
If $\boldsymbol{X}, \boldsymbol{Y}$ are λ_k-indistinguishable conjunctive sources of width at most k then $\Pr[\boldsymbol{X} = 0] = \Pr[\boldsymbol{Y} = 0]$*.*
Furthermore, the theorem doesn't hold for $\lambda_k = \lfloor \log_2 k \rfloor + 1$.

Corollary 6.36. *If $\boldsymbol{X}, \boldsymbol{Y}$ are λ_k-indistinguishable conjunctive sources of width at most k then $\boldsymbol{X}, \boldsymbol{Y}$ are equidistributed.*

Proof. Applying the theorem to all projections of $\boldsymbol{X}, \boldsymbol{Y}$, we obtain that for each $S \subseteq [n]$,

$$\Pr[\boldsymbol{X}|_S = 0] = \Pr[\boldsymbol{Y}|_S = 0].$$

The inclusion-exclusion principle shows that for each $T \subseteq [n]$,

$$\Pr[\boldsymbol{X}|_T = 0, \boldsymbol{X}|_{\overline{T}} = 1] = \sum_{R \subseteq \overline{T}} (-1)^{|R|} \Pr[\boldsymbol{X}|_{T \cup R} = 0],$$

and so these probabilities are identical for $\boldsymbol{X}$ and $\boldsymbol{Y}$. □

To complete the proof of the theorem, we show how to reduce to the case of narrow sources; Theorem 6.33 follows by taking $k = \log(n/\epsilon)$.

Lemma 6.37. *Let $\boldsymbol{X}, \boldsymbol{Y}$ be two λ_k-indistinguishable conjunctive sources. Then $\boldsymbol{X}, \boldsymbol{Y}$ have statistical distance at most $2^{-k} n$.*

Proof. Let $\boldsymbol{X}', \boldsymbol{Y}'$ be the sources obtained by replacing every coordinate of width more than k by zero. A coordinate i has width more than k iff $\Pr[\boldsymbol{X}_i = 1] < 2^{-k}$, showing that the replaced coordinates are found in identical positions in $\boldsymbol{X}$ and $\boldsymbol{Y}$, and so $\boldsymbol{X}', \boldsymbol{Y}'$ are also λ_k-indistinguishable. By construction, $\boldsymbol{X}', \boldsymbol{Y}'$ have width at most k, and so are identically distributed by Corollary 6.36.

To complete the proof of the lemma, we show that the statistical distance between $\boldsymbol{X}$ and $\boldsymbol{X'}$ is at most $2^{-k-1} n$. To this end, we consider a source $\boldsymbol{Z}$ obtained by zeroing a *single* coordinate i, and show that $\boldsymbol{X}$ and $\boldsymbol{Z}$ are at statistical distance at most 2^{-k-1}.

By definition, the statistical distance between $\boldsymbol{X}$ and $\boldsymbol{Z}$ is

$$\sum_{x\colon \Pr[\boldsymbol{X}=x] > \Pr[\boldsymbol{Z}=x]} (\Pr[\boldsymbol{X} = x] - \Pr[\boldsymbol{Z} = x]).$$

It is easy to see that if $\Pr[\boldsymbol{X} = x] > \Pr[\boldsymbol{Z} = z]$ then $x_i = 1$, and conversely if $x_i = 1$ then $\Pr[\boldsymbol{X} = x] \geq \Pr[\boldsymbol{Z} = x]$ (both could be zero). Therefore the statistical distance is exactly

$$\sum_{x\colon x_i = 1} (\Pr[\boldsymbol{X} = x] - \Pr[\boldsymbol{Z} = x]) = \Pr[\boldsymbol{X}_i = 1] - \Pr[\boldsymbol{Z}_i = 1] = 2^{-k_i},$$

where $k_i \geq k + 1$ is the width of $\boldsymbol{X}_i$. □

To complete the picture, we now give an example of two $\Theta(\log \log n)$-indistinguishable depth 1 sources which can be distinguished by a DNF, using a construction of Amano et al. [AIM$^+$03].

Theorem 6.38 ([AIM$^+$03, Section 3.2])**.** *Let ℓ be an integer. For $i \leq 1 \leq \ell$, let*

$$\boldsymbol{X}_i = \boldsymbol{Y}_i = \bigwedge_{\substack{S \subseteq \{1,\ldots,\ell\} \\ i \in S}} r_S,$$

and define

$$\boldsymbol{X}_{\ell+1} = \bigwedge_{\substack{S \subseteq \{1,\ldots,\ell\} \\ |S| \text{ odd}}} r_S, \qquad \boldsymbol{Y}_{\ell+1} = \bigwedge_{\substack{S \subseteq \{1,\ldots,\ell\} \\ |S| \text{ even}}} r_S.$$

The two sources $\boldsymbol{X}, \boldsymbol{Y}$ are ℓ-indistinguishable.

Proof sketch. By inclusion-exclusion, it suffices to show that for every $T \subsetneq \{1, \ldots, \ell+1\}$, $\Pr[\boldsymbol{X}|_T = 1] = \Pr[\boldsymbol{Y}|_T = 1]$. It suffices to consider T of the form $R \cup \{\ell+1\}$, where $R \subsetneq \{1, \ldots, \ell+1\}$.

For such R, T, $\Pr[\boldsymbol{X}|_T = 1] = 2^{-a}$, where a is the number of subsets of $\{1, \ldots, \ell\}$ containing R and having odd size, and $\Pr[\boldsymbol{Y}|_T = 1] = 2^{-b}$, where b is the number of subsets of $\{1, \ldots, \ell\}$ containing R and having even size. It is not hard to check that $a = b = 2^{\ell-1-|R|}$. □

As an example, if $\ell = 2$ then the sources are

$$\begin{aligned} \boldsymbol{X} &= r_1 \wedge r_{12}, r_2 \wedge r_{12}, r_1 \wedge r_2 \\ \boldsymbol{Y} &= r_1 \wedge r_{12}, r_2 \wedge r_{12}, r_\emptyset \wedge r_{12} \end{aligned}$$

We can now construct our example, following Amano et al.

Lemma 6.39. *For infinitely many n there exists a pair $\boldsymbol{Z}, \boldsymbol{W}$ of $\Theta(\log\log n)$-indistinguishable depth 1 sources over $\{0,1\}^n$ which can be $\Omega(1)$-distinguished by a DNF.*

Proof. Let $\ell = \Theta(\log\log n)$ be a parameter to be determined. We create $\boldsymbol{Z}, \boldsymbol{W}$ by taking $n/(\ell+1)$ independent copies $\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)}$ of the sources $\boldsymbol{X}, \boldsymbol{Y}$ given by Theorem 6.38. According to the theorem, the resulting sources are $(\ell+1)$-indistinguishable.

The source $\boldsymbol{X}$ mentions $2^\ell - 1$ variables (all but $r_\emptyset$), and the source $\boldsymbol{Y}$ mentions all 2^ℓ variables. Therefore

$$\begin{aligned} \Pr[\boldsymbol{X}^{(i)} \neq \mathbf{1} \text{ for all } i] &= \left(1 - (1/2)^{2^\ell - 1}\right)^{n/(\ell+1)} \approx \exp\left(-\frac{n}{(\ell+1)2^{2^\ell - 1}}\right), \\ \Pr[\boldsymbol{Y}^{(i)} \neq \mathbf{1} \text{ for all } i] &= \left(1 - (1/2)^{2^\ell}\right)^{n/(\ell+1)} \approx \exp\left(-\frac{n}{(\ell+1)2^{2^\ell}}\right). \end{aligned}$$

We choose ℓ so that $n \approx (\ell+1)2^{2^\ell}$ (this is possible for infinitely many n). Defining f to be the CNF corresponding to the event considered above, we obtain $\Pr[f(\boldsymbol{Z}) = 1] \approx e^{-2}$ while $\Pr[f(\boldsymbol{W}) = 1] \approx e^{-1}$. □

6.6 Local DNFs

In this subsection we adapt the definition of predictability to more general classes of functions, and prove that degree 1 sources are predictable in this sense for local DNFs.

Definition 6.40 (local DNF)**.** An *s-local DNF* is a disjunction of functions depending on at most s bits.

We will use the following notion of predictability. Since it does not generalize the earlier notion, we use a different term, *approximability*.

Definition 6.41 (approximability)**.** A source $\boldsymbol{X}$ on $\{0,1\}^n$ is (k, ϵ, f)-*approximable*, where $f\colon \{0,1\}^n \to \{0,1\}$, if there exists a decision tree of depth k, whose leaves are labelled $0, 1, \perp$, with the following properties:

- If an input x reaches a leaf labelled $b \in \{0,1\}$, then $f(x) = b$.
- The probability that $\boldsymbol{X}$ reaches a leaf labelled $\perp$ is at most ϵ.

Approximability is useful for the following reason.

Lemma 6.42. *Suppose that $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable, and that $\boldsymbol{Y}$ is (k, ϵ, f)-approximable. Then $\boldsymbol{X}, \boldsymbol{Y}$ ϵ-fool f.*

Proof. Let T be the decision tree promised by the definition of (k, ϵ, f)-approximability. Let Λ_b be the set of leaves of T labelled b. Thus

$$\Pr[T(\boldsymbol{X}) \in \Lambda_1] \leq \mathsf{E}[f(\boldsymbol{X})] \leq \Pr[T(\boldsymbol{X}) \notin \Lambda_0].$$

Since T has depth k, the probabilities on the left and on the right are the same for $\boldsymbol{X}$ and for $\boldsymbol{Y}$. Moreover, the difference between the two sides is

$$\Pr[T(\boldsymbol{X}) \notin \Lambda_0] - \Pr[T(\boldsymbol{X}) \in \Lambda_1] = \Pr[T(\boldsymbol{X}) \in \Lambda_\perp] \leq \epsilon,$$

implying the lemma. □

Our main result in this subsection states that degree 1 sources are (k, ϵ, f)-approximable for local DNFs (for appropriate choices of parameters).

Theorem 6.43. *If f is an s-local DNF and $\boldsymbol{X}$ is a degree 1 source then $\boldsymbol{X}$ is $(O(s2^s \log(1/\epsilon)), \epsilon, f)$-approximable.*

Proof. The proof is by induction on s. We will show that for every s there is a constant $c_s = O(s2^s \log(1/\epsilon))$ such that every degree 1 source is (c_s, ϵ, f)-approximable for every s-local DNF f.

The base case, $s = 0$, is trivial: we can take $c_0 = 0$.

Now suppose that $s > 0$. Let f be the disjunction of non-zero local functions f_i depending on coordinates J_i, where $|J_i| \leq s$. We can assume, without loss of generality, that $\boldsymbol{X}|_{J_i}$ are affinely independent.

Let I be an inclusion-maximal set of coordinates such that all coordinates in the multiset $\bigcup_{i \in I} J_i$ are affinely independent. We consider two cases, according to the size of I.

Case 1. The set I contains at least $2^s \log(1/\epsilon)$ coordinates. Let $B \subseteq I$ have exactly $2^s \log(1/\epsilon)$ coordinates. By assumption, $f_i \neq 0$ for all $i \in B$, and so $\Pr[f_i(\boldsymbol{X}) = 0] \leq 1 - 2^{-|J_i|} \leq 1 - 2^{-s}$. By construction,

$$\Pr[f_i(\boldsymbol{X}) = 0 \text{ for all } i \in B] = \prod_{i \in B} \Pr[f_i(\boldsymbol{X}) = 0] \leq (1 - 2^{-s})^{|B|} \leq \epsilon.$$

Accordingly, we construct a decision tree of depth $s|B| = s2^s \log(1/\epsilon)$ which queries all variables in $\bigcup_{i \in B} J_i$. After querying all variables, either we discover that $f_i(x) = 1$ for some $i \in B$ and so $f(x) = 1$, or else $f_i(x) = 0$ for all $i \in B$; but the latter case happens with probability at most ϵ. Hence $\boldsymbol{X}$ is $(s2^s \log(1/\epsilon), \epsilon, f)$-approximable.

Case 2. The set I contains at most $2^s \log(1/\epsilon)$ coordinates. Using a decision tree T of depth $s2^s \log(1/\epsilon)$, we can query all of these coordinates, either determining that $f(x) = 1$ or that $f_i(x) = 0$ for all $i \in I$. In the latter case, we also know the values of all coordinates in $J = \bigcup_{i \in I} J_i$. By definition of I, we can now write every f_j for $j \notin I$ as a function of at most $s - 1$ many input coordinates (since some linear combination in J_j is affinely dependent on J). In this way, we obtain an $(s-1)$-local DNF g_ℓ which agrees with f on the coset $\boldsymbol{X}_\ell$ corresponding to the leaf ℓ. Note that $\boldsymbol{X}_\ell$ is also a degree 1 source.

By induction, $\boldsymbol{X}_\ell$ is $(c_{s-1}, \epsilon, g_\ell)$-approximable. Attaching the corresponding decision trees to the leaves of T, we obtain a decision tree of depth $s2^s \log(1/\epsilon) + c_{s-1}$ which satisfies the definition of (c_s, ϵ, f)-approximability for

$$c_s = c_{s-1} + s2^s \log(1/\epsilon).$$

Finally, notice that

$$c_s = \sum_{t=1}^{s} t2^t \log(1/\epsilon) = O(s2^s \log(1/\epsilon)).$$

□

Corollary 6.44. $\mathsf{GENERAL}(s2^s \log(1/\epsilon), \epsilon)$ *holds for the class of degree 1 sources and the class of s-local DNF distinguishers.*

7 From bounded independence to bounded indistinguishability

A potential method for proving AC^0-indistinguishability of a specific pair of distributions $\boldsymbol{X}, \boldsymbol{Y}$ is by reduction to bounded independence.

Braverman's Theorem [Bra11, Tal17] *k-independence ϵ-fools AC^0 circuits of depth d and size s, where $k = \left(\log \frac{s}{\epsilon}\right)^{O(d)}$.*

If $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable and one of the distributions, say $\boldsymbol{X}$, happens to be k-independent, then so must $\boldsymbol{Y}$ be, and the pair 2ϵ-fools AC^0. More generally, it may be possible to reduce $\boldsymbol{X}, \boldsymbol{Y}$ to a pair of locally independent distributions $\boldsymbol{X}', \boldsymbol{Y}'$. Such reductions are modeled by the following definition:

Definition 7.1 (k-similarity)**.** Let $\mathcal{F}$ be a collection of samplers. We say that two distributions $\boldsymbol{X}, \boldsymbol{Y}$ are *k-similar with respect to $\mathcal{F}$* if there exists a sampler $F \in \mathcal{F}$ and a pair of k-independent distributions $\boldsymbol{X'}, \boldsymbol{Y'}$ such that $\boldsymbol{X} \sim F(\boldsymbol{X'})$ and $\boldsymbol{Y} \sim F(\boldsymbol{Y'})$.

In particular, if $\mathcal{F}$ consists of AC^0 circuits of size s' and depth d' then by Braverman's theorem, any $\log((s+s')/\epsilon)^{O(d+d')}$-similar pair 2ϵ-fools AC^0 circuits of size s and depth d.

In Section 7.1 we give two examples of pairs of distributions that are locally similar with respect to NC^0. In contrast, in Section 7.2, for every k we exhibit a distribution $\boldsymbol{X}$ such that the pair $(\boldsymbol{X}, \boldsymbol{X})$ is not k-similar with respect to NC^0 (that is, we cannot find an NC^0 sampler F and a k-independent distribution $\boldsymbol{X'}$ such that $\boldsymbol{X} \sim F(\boldsymbol{X'})$. We conjecture that the lack of similarity holds even with respect to AC^0. Section 9.2.2 discusses the relevance of this example for proving security of multiparty computation protocols against AC^0 adversaries.

7.1 Examples of locally similar distributions

In this subsection we give two examples of locally similar pairs of distributions. The first example is a natural secure multiparty protocol for the parity function. The second example, inspired by the degree-reduction step in MPC protocols such as [BGW88], concerns parity tree computations over certain pairs of distributions of disjoint support. This example illustrates the connection between k-indistinguishability and k-independence.

MPC for Computing Parity. Consider a setting where there are n parties, and each party $i \in [n]$ holds a bit x_i. The parties wish to compute the parity of their bits securely. A simple protocol for this task starts with each party secret-sharing its private input using an n-out-of-n secret sharing. The parties then send the corresponding shares to the other parties and they locally compute the parity of the shares obtained. They then send the resultant values to each other and compute the parity of these values to obtain the parity of their private inputs.

Once the parties have exchanged their shares with each other, the joint distribution of the values held by the parties can be described by a matrix $M \in \{0,1\}^{n \times n}$ such that the columns correspond to the secret shares of $x_1, \ldots, x_n$. The rows correspond to the shares held by each party. The local computation done by the parties corresponds to computing the parity of each row to obtain a column vector. The final output computation done by the parties corresponds to computing the parity of this column vector.

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be the distribution on the wires of the joint computations performed by the protocol on any pair of inputs $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ of identical parity.

Proposition 7.2. *$\boldsymbol{X}, \boldsymbol{Y}$ are $(n-1)$-similar with respect to 3-local samplers.*

The identical parity restriction is necessary, since otherwise $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be distinguished by the output wire.

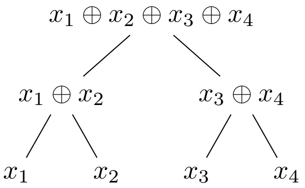
Proof of Proposition 7.2. The distribution $\boldsymbol{X}$ (resp., $\boldsymbol{Y}$) consists of:

- The entries $m_{i,j}$ of M, which are random bits conditioned on the column sums $m_{1,j} + \cdots + m_{n,j}$ being equal to x_j (resp., y_j), $1 \le i, j \le n$;
- The partial row sums $r_{i,j} = m_{1,j} + \cdots + m_{i,j}$, $1 \le i, j \le n$;
- The partial column sums $c_1 = r_{1,n}$, $c_2 = r_{1,n} + r_{2,n}, \ldots, c_n = r_{1,n} + \cdots + r_{n,n} = x_1 + \cdots + x_n = y_1 + \cdots + y_n$.

Consider the distribution $\boldsymbol{X}'$ (resp., $\boldsymbol{Y}'$) consisting of the $n(n-1)$ partial row sums $r_{i,j}$, $1 \le i \le n$, $1 \le j \le n-1$ and the $(n-1)$ partial column sums $c_1, c_2, \ldots, c_{n-1}$. Then $\boldsymbol{X}'$ (resp., $\boldsymbol{Y}'$) is $(n-1)$-independent: these values are random conditioned on $r_{1,j} + \cdots + r_{n,j} = x_j$ (resp., y_j), and so any $n-1$ of them are linearly, and therefore statistically, independent.

It remains to describe the sampler F. The bits $r_{i,n}$ can be computed 2-locally as $c_i + c_{i-1}$ for $i < n$ and 1-locally as $c_{n-1} + x_1 + \cdots + x_n = c_{n-1} + y_1 + \cdots + y_n$ for $i = n$. The bits $m_{i,j}$ can be computed as $r_{i,j} + r_{i-1,j}$, which is a 2-local function in the entries of $\boldsymbol{X}'$ (resp., $\boldsymbol{Y}'$) when $j < n$ and 3-local when $j = n$. □

Parity Tree Computations. A *parity tree* is a circuit consisting of a full binary tree where each gate is the XOR of its two children. Such a tree computes the XOR of $N = 2^m$ bits. For example, when $m = 2$, the tree is:



The *parity-tree vector* $\mathsf{PT}(x)$ in $\{0,1\}^{2N-1}$ is the vector that consists of all the wire values within the parity tree; e.g.,

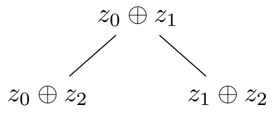
$$\mathsf{PT}(x_1, x_2, x_3, x_4) = (x_1,\, x_2,\, x_3,\, x_4,\, x_1 \oplus x_2,\, x_3 \oplus x_4,\, x_1 \oplus x_2 \oplus x_3 \oplus x_4).$$

Our next step is to introduce a pair of (linear) sources on $\{0,1\}^{2N-1}$. Suppose $m \ge 1$, and consider a parity tree on $x_1, x_2, \ldots, x_N$ and the corresponding parity-tree vector. Let π be a parity over $\{0,1\}^N$, let $\boldsymbol{X} = \mathsf{PT}(\boldsymbol{x})$ where $\boldsymbol{x}$ is chosen uniformly over all vectors in $\{0,1\}^N$ satisfying $\pi(\boldsymbol{x}) = 0$, and let $\boldsymbol{Y} = \mathsf{PT}(\boldsymbol{y})$ where $\boldsymbol{y}$ is chosen uniformly over all vectors in $\{0,1\}^N$ satisfying $\pi(\boldsymbol{y}) = 1$.

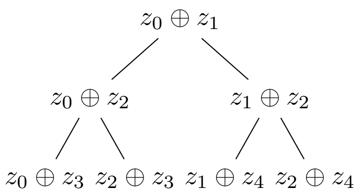
We would like to exhibit a parity π such $\boldsymbol{X}, \boldsymbol{Y}$ are $\Omega(N)$-indistinguishable and $\Omega(N)$-similar with respect to NC^0. The problem is that we cannot generate the parity tree even in AC^0 given the values of the leaves, since the output is a parity of N bits; however, we can generate it in NC^0 in a different way, as follows. Start with $m = 0$:

$$z_0 \oplus z_1$$

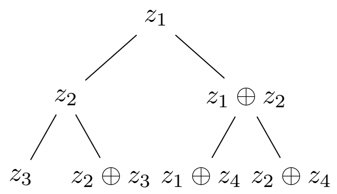
Associate the root with a new variables z_2, and use it to split $z_0 \oplus z_1$:



Associate the leaves with new variables z_3, z_4, and split again:



We can set $z_0 = 0$ and still get the same distribution, using the optimal number of input bits:



This idea can be easily generalized, and thus provides us with a way of generating the parity-tree vector with a 2-local sampler.

Having that in mind, we choose the parity to be $\pi(x) = x_1 + x_3 + x_5 + \cdots + x_{N-1}$. This parity translates to a parity ρ in terms of z, which denotes the randomness used to sample the parity tree, and following the above construction it is easy to observe that, although some cancellations may occur among the z_i's, the number of z_i's on which ρ depends is $\Omega(N)$. Thus, if we define $\boldsymbol{X'}$ and $\boldsymbol{Y'}$ to be the uniform distributions over $z \in \{0,1\}^N$ satisfying $\rho(z) = 0$ and $\rho(z) = 1$, respectively, we get that $\boldsymbol{X'}, \boldsymbol{Y'}$ are $\Omega(N)$-independent. Moreover, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the image of $\boldsymbol{X'}$ and $\boldsymbol{Y'}$, respectively, under our parity-tree sampler, which implies that $\boldsymbol{X}, \boldsymbol{Y}$ are $\Omega(N)$-similar with respect to 2-local samplers.

It also follows that $\boldsymbol{X}, \boldsymbol{Y}$ are $\Omega(N)$-indistinguishable by the following argument: more generally, suppose that $\boldsymbol{X'}, \boldsymbol{Y'}$ are k-independent distributions and C is a linear NC^0 circuit such that $\boldsymbol{X} = C(\boldsymbol{X'})$ and $\boldsymbol{Y} = C(\boldsymbol{Y'})$, and suppose that each output of C depends on at most B bits; then $\boldsymbol{X}, \boldsymbol{Y}$ are (k/B)-indistinguishable, because any subset of k/B outputs of C depends on at most k of its inputs.

Thus far, we showed a concrete pair of indistinguishable distributions that are similar with respect to our 2-local sampler, and we argued indistinguishability using similarity. It turns out that the converse is true as well, namely that indistinguishability of distributions based on a parity tree and a linear constraint implies their similarity. This is summarized in the following proposition.

Proposition 7.3 (indistinguishability implies similarity under local sampling)**.** *Let π be a parity over $\{0,1\}^N$. Let $\boldsymbol{x}$ be the uniform distribution over all vectors $x \in \{0,1\}^N$ satisfying $\pi(x) = 0$, and let $\boldsymbol{y}$ be the uniform distribution over all vectors $y \in \{0,1\}^N$ satisfying $\pi(y) = 1$. Define $\boldsymbol{X} \triangleq \mathsf{PT}(\boldsymbol{x})$ and $\boldsymbol{Y} \triangleq \mathsf{PT}(\boldsymbol{y})$. If $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable, then $\boldsymbol{X}, \boldsymbol{Y}$ are k-similar with respect to 2-local samplers.*

Proof. Suppose that $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable. Consider the parity tree on N variables, and think of π as a subset S of the tree leaves. Perform the following simplification repeatedly: if S contains two siblings, replace them with their parent. Eventually we reach a set S of vertices in the parity tree with the following two properties: (i) if a vertex v belongs to S then the sibling of v is not in S; (ii) the constraint $\pi(x) = b$ is the same as the constraint $\bigoplus_{v \in S} \pi_v(x) = b$, where π_v is the parity associated with the vertex v. Since $\boldsymbol{X}, \boldsymbol{Y}$ are k-indistinguishable, it must be that $|S| > k$; in fact, the optimal indistinguishability of $\boldsymbol{X}, \boldsymbol{Y}$ is exactly $|S| - 1$.

The 2-local construction of the parity tree expresses every vertex as a sum of two z_i's. It can be checked that for any two siblings, there is a variable that appears in both siblings and only in them. Therefore, the constraint $\pi(x) = b$ translates to a constraint $\rho(z) = b$, where the number of z_i's that appear in ρ is greater than k. Such a constraint corresponds to an r-independent distribution with $r \geq k$. Hence, we can define $\boldsymbol{X'}$ and $\boldsymbol{Y'}$ to be uniformly distributed over vectors z satisfying $\rho(z) = 0$ and $\rho(z) = 1$, respectively, thus showing that $\boldsymbol{X}, \boldsymbol{Y}$ are k-similar with respect to 2-local samplers. □

7.2 A distribution that is not locally self-similar

In this section, we exhibit a distribution $\boldsymbol{X}$ that cannot be k-locally sampled by a sampler from any $O(k)$-wise independent source. The distribution of interest is uniform over the codewords of a low-density parity-check code with sufficiently large unique expansion.

Such codes can be constructed using bipartite graphs with desired expansion properties. We call a bipartite graph $G(L \cup R, E)$, where $|L| = n$ and $|R| = m$, an $(n, m, D, \epsilon, \gamma)$-*bipartite expander graph* if it is D-left-regular, and any $S \subseteq L$ with $|S| \leq \gamma n$ has at least $(1-\epsilon)D|S|$ neighbors in R. For a bipartite expander graph G, let $C(G)$ denote the linear code over $\mathbb{F}_2$ with the bipartite adjacency matrix $P \in \mathbb{F}_2^{m \times n}$ of G as its parity-check matrix; that is, $C(G) = \{z \in \mathbb{F}_2^n \mid Pz = 0\}$. Let us denote $\theta \triangleq m/n$. The code $C(G)$ is an $[n, (1-\theta)n]_2$-code and is called an *expander code*. If $D = \Theta(\frac{1}{\epsilon}\log\frac{1}{\theta})$ and we pick a graph at random, it satisfies the following property: every set $S \subseteq L$ of up to γn nodes on the left has at least $(1-\epsilon)D|S|$ neighbors on the right, where $\gamma = \Theta(\epsilon^2\theta/\log\frac{1}{\theta})$ (cf. [Gur10]). This implies that S has at least $(1-2\epsilon)D|S|$ *unique* neighbors on the right, by which it follows that the minimum distance of $C(G)$ is at least γn (assuming $\epsilon < 1/2$).

We can now state the main result of this section.

Proposition 7.4. *For every large enough D there exists $\beta > 0$ such that for every k and sufficiently large n there exists a D-left-regular G for which the uniform distribution $\boldsymbol{X}$ over $C(G)$ is not samplable from any k-independent distribution by βk-local samplers.*

By a result of Lovett and Viola [LV11], the same conclusion holds when the sampler is in AC^0, but is given a uniformly random distribution instead of it being k-independent.

Proposition 7.4 explains why Braverman's theorem cannot be directly used to argue the security of the LRCC construction of Fig. 4. See Section 9.2.2 for details.

We show that for any sampler F in NC^0 and any k-independent distribution $\boldsymbol{z}$, $F(\boldsymbol{z})$ cannot be distributed according to $\boldsymbol{X}$.

In what follows, we assume that the parameters θ, ϵ are constant, and n grows to infinity. Let $C(G)$ be an expander code. As a warm up, we show that given a uniform distribution $\mathbf{z} \in \{0,1\}^t$, no $O(1)$-local sampler can sample uniformly from $C(G)$.

Claim 7.5. *Let G be an $(n, \theta n, D, \epsilon, \gamma)$-bipartite expander, $B > 0$ a constant, and t any integer. Let $F\colon \{0,1\}^t \to \{0,1\}^n$ be an arbitrary B-local function. If $n > B/\gamma(1-\theta)$ then F cannot sample uniformly from $C(G)$ given the uniform distribution $\mathbf{z} \in \{0,1\}^t$.*

Proof. We can assume that each bit in $\mathbf{z}$ is used by the function F. Since each of the n output bits depends on at most B input bits, the average input bit affects nB/t output bits. Therefore, there is an input bit which affects at most nB/t many output bits. Moreover, since F depends on all bits in $\mathbf{z}$, there exists a setting of the remaining bits such that flipping this input bit will change $0 \neq a \leq nB/t$ many output bits, and so by the minimum distance property, $nB/t \geq \gamma n$. Therefore, $t \leq B/\gamma$. This implies that the entropy of the output $H(F(\boldsymbol{z}))$, which is bounded by $H(\boldsymbol{z})$, is at most B/γ, which is impossible for large enough n, since the entropy of a random sample from $C(G)$ is at least $(1-\theta)n$. □

We extend this argument to show that the same holds if we replace the uniform distribution by a k-independent distribution, which we use to complete the proof of Proposition 7.4.

Claim 7.6. *Let G be a $(n, \theta n, D, \epsilon, \gamma)$-bipartite expander, where $\epsilon < 1/4$. Let $F\colon \{0,1\}^t \to \{0,1\}^n$ be a B-local function. Then, for $k > \max\{B/\epsilon\gamma, BD/\theta, B/((1-\theta)\epsilon) + BD/(1-\theta)\}$ and large enough n, F cannot sample uniformly from $C(G)$ given a k-independent distribution $\boldsymbol{z}$ over $\{0,1\}^t$.*

Proof. Let P be the $\theta n \times n$ bipartite adjacency matrix of G. By assumption, each output bit of F depends on at most B input bits. Since $F(\boldsymbol{z})$ must lie in $C(G)$, we must have $PF(\boldsymbol{z}) = 0$ always.

We extend G to a graph on $W \cup L \cup R$, where $W \triangleq [t]$, by adding an edge between every $w \in W$ and $\ell \in L$ if the output bit ℓ of F depends on the input bit w. Thus each vertex in L has D neighbors in R and at most B neighbors in W. Hence there are at most nBD paths between vertices in R and W.

We will classify vertices of L and R as *good* or *bad* (the typical vertex will be good), starting with R. A vertex $r \in R$ is *good* if the number of paths starting from r to some vertex in W is at most k. This means that the r'th entry of $PF(\boldsymbol{z})$ can be expressed as a function of at most k variables in z, say $(PF(\boldsymbol{z}))_r = f_r(\boldsymbol{z}_{i_1}, \dots, \boldsymbol{z}_{i_s})$, where $s \leq k$. By k-independence, the marginal distribution of $(\boldsymbol{z}_{i_1}, \dots, \boldsymbol{z}_{i_s})$ is uniform over $\{0,1\}^s$, and so $f_r \equiv 0$. In other words, the r'th coordinate of $PF(z)$ must always evaluate to zero, for *any* $z \in \{0,1\}^t$ (not necessarily one in the support of $\boldsymbol{z}$). Since there are most BDn many paths between vertices in R and W, by Markov's inequality there can be at most $(BD/k)n$ bad vertices in R.

A vertex $\ell \in L$ is *bad* if it has at least $2\epsilon D$ bad neighbors in R; otherwise it is good. How many vertices in L can be bad? Let $S \subseteq L$ be a set of at most γn bad vertices. On the one hand, by expansion, S must have at least $(1-\epsilon)D|S|$ neighbors in R. On the other hand, each vertex in S has at most $(1-2\epsilon)D$ neighbors in R which are good, and so considering also the $(BD/k)n$ bad vertices in R, we see that S has at most $(BD/k)n + (1-2\epsilon)D|S|$ neighbors in R. It follows that $\epsilon D|S| \leq (BD/k)n$ and so $|S| \leq (B/\epsilon k)n$. We conclude that at most $(B/\epsilon k)n$ vertices of L are bad.

From now on, we restrict attention only to good vertices. Let $L' \subseteq L$ and $R' \subseteq R$ be the good vertices, and let G', P', F' be the corresponding restrictions of G, P, F. Let $W' \subseteq W$ be all the input bits that F' depends on, let $t' = |W'|$, and let $\boldsymbol{z'}$ be the corresponding restriction of $\boldsymbol{z}$. Thus $|L'| \geq (1 - B/\epsilon k)n$, $|R'| \geq (\theta - BD/k)n$, $|W| - |W'| \leq (BD/k)n$, and $P'F'(z') = 0$ for any $z' \in \{0,1\}^{t'}$.

Let us consider the code $C(G')$. We claim that its minimum distance is at least γn. Indeed, if $S \subseteq L'$ is a subset of up to γn vertices then S has at least $(1-2\epsilon)D|S|$ unique neighbors in R and so at least $(1-4\epsilon)D|S|$ unique neighbors in R' (since all vertices in S are good). Since $\epsilon < 1/4$, this is positive, and so the vector corresponding to S doesn't belong to $C(G')$.

Since $P'F'(z') = 0$ for all $z' \in \{0,1\}^{t'}$, $F'(z')$ must be a codeword of $C(G')$, and so it either consists entirely of zeroes, or contains at least γn many ones. On the other hand, each output bit of F' depends on most B many inputs, and so there exists an input bit that affects at most $B|L'|/|W'|$ many outputs. Again, since F' depends on all bits in z', there exists a setting of the remaining bits such that flipping this bit will change $0 \neq a \leq B|L'|/|W'| \leq Bn/|W'|$ output bits, giving $Bn/|W'| \geq \gamma n$. It follows that F' depends on at most B/γ many coordinates of z'. Consequently, the entropy of $F'(\mathbf{z}')$ is at most B/γ.

Since F is obtained from F' by removing at most $(B/\epsilon k)n$ many output bits and $(BD/k)n$ many input bits, the entropy of output $H(F(\mathbf{z}))$, is at most $H(F'(\mathbf{z}')) + (B/\epsilon k + BD/k)n$. On the other hand, $F(\mathbf{z})$ is uniformly distributed over $C(G)$, which has rate at least $(1-\theta)n$. We conclude that $n \leq (B/\gamma)/(1 - \theta - B/\epsilon k - BD/k)$. □

Proposition 7.4 now follows. Indeed, for our choice of ϵ and θ, we fix $D = \Theta(\frac{1}{\epsilon}\log\frac{1}{\theta})$ and $\gamma = \Theta(\epsilon^2\theta/\log\frac{1}{\theta})$ so that an $(n, \theta n, D, \epsilon, \gamma)$-bipartite expander G is guaranteed to exist. For any ℓ, let F be a ℓ-local sampler. Applying Claim 7.6 for this choice of parameters and for sufficiently large n tells that for $\ell > k\beta$, where $\beta = \max\{1/\epsilon\gamma, D/\theta, 1/((1-\theta)\epsilon) + D/(1-\theta)\}$, the sampler F cannot sample uniformly from $C(G)$ given a k-independent distribution.

8 Simplifying sources via randomized encoding

In this section we explore a generalization of the technique that was used in Section 5.3 to convert a positive result for sources sampled by $\mathsf{poly}(n)$-size decision trees to a positive result for sources sampled by degree $O(\log n)$ polynomial maps. The high-level idea is to replace each output bit of the sampler by an s-bit *randomized encoding* (RE) of this bit that admits a lower-complexity sampler. This transformation respects k-indistinguishability, and can respect the distinguishing advantage by incorporating the decoder of the RE into the distinguisher.

The rest of this section is organized as follows. After formalizing the notion of randomized encoding of functions and its application to encoding samplers (Section 8.1), in Section 8.2 we use known RE constructions to show that a (hypothetical) positive result with $o(\log\log n)$-local $n^{\Omega(1)}$-indistinguishable sources and an ϵ-distinguisher in AC^0 can be converted into a similar positive result with 4-local sources. Finally, in Section 8.3

we put forward a natural conjecture about the complexity of RE for AC^0 that may be viewed as a barrier to negative results for local sources.

8.1 Encoding samplers

We start by recalling the standard notion of a randomized encoding of functions, focusing on the case of perfect privacy and statistical correctness.

Definition 8.1 (Randomized encoding of functions [IK00, AIK06])**.** Let $f\colon \{0,1\}^m \to \{0,1\}$. A function $\hat{f}\colon \{0,1\}^m \times \{0,1\}^\rho \to \{0,1\}^s$ is a (perfectly private) *randomized encoding (RE) of* f with decoder Dec and error δ if there exist distributions D_0, D_1 on $\{0,1\}^s$ such that for every $x \in \{0,1\}^m$:

- *Privacy*: If we choose $\mathbf{r} \in \{0,1\}^\rho$ at random then $\hat{f}(x, \mathbf{r}) \sim D_{f(x)}$;
- *Correctness*: $\Pr_{\mathbf{r}}[\mathsf{Dec}(\hat{f}(x, \mathbf{r})) = f(x)] \geq 1 - \delta$.

We use RE in a natural way to encode a pair of k-indistinguishable sources $\boldsymbol{X}, \boldsymbol{Y}$. Without loss of generality, we consider here the coset variant of our main question, where $\boldsymbol{X} = F(0, \boldsymbol{x})$ and $\boldsymbol{Y} = F(1, \boldsymbol{x})$.[7]

Lemma 8.2 (Encoding samplers)**.** *Let* $F\colon \{0,1\}^m \to \{0,1\}^n$ *be a* sampler *defining a pair of sources* $\boldsymbol{X} = F(0, \boldsymbol{x})$ *and* $\boldsymbol{Y} = F(1, \boldsymbol{x})$*. Let* $f_i\colon \{0,1\}^m \to \{0,1\}$ *be the ith output of* F *and* $\hat{f}_i\colon \{0,1\}^m \times \{0,1\}^{\rho_i} \to \{0,1\}^{s_i}$ *an RE of* f_i *with decoder* Dec_i *and error* δ_i*. Consider the* encoded sampler $\hat{F}$ *defined by* $\hat{F}(b, x, r_1, \ldots, r_n) = (\hat{f}_1(b, x, r_1), \ldots, \hat{f}_n(b, x, r_n))$ *and the encoded sources* $\hat{\boldsymbol{X}} = \hat{F}(0, \boldsymbol{x}, \boldsymbol{r_1}, \ldots, \boldsymbol{r_n})$ *and* $\hat{\boldsymbol{Y}} = \hat{F}(1, \boldsymbol{x}, \boldsymbol{r_1}, \ldots, \boldsymbol{r_n})$*. Then:*

- *If* $\boldsymbol{X}, \boldsymbol{Y}$ *are k-indistinguishable then so are* $\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}$*;*
- *If* $C\colon \{0,1\}^n \to \{0,1\}$ ϵ*-distinguishes between* $\boldsymbol{X}, \boldsymbol{Y}$ *then* $\hat{C}\colon \{0,1\}^{s_1} \times \cdots \times \{0,1\}^{s_n} \to \{0,1\}$*, defined by* $\hat{C}(\hat{y}_1, \ldots, \hat{y}_n) = C(\mathsf{Dec}_1(\hat{y}_1), \ldots, \mathsf{Dec}_n(\hat{y}_n))$*,* $\hat{\epsilon}$*-distinguishes between* $\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}$ *for* $\hat{\epsilon} = (1-\delta)\epsilon - \delta$*, where* $\delta = \sum_{i=1}^n \delta_i$*.*

Proof. The k-indistinguishability of $\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}$ follows from the privacy requirement of the RE and the fact that each RE instance $\hat{f}_i$ uses fresh randomness r_i. Indeed, the latter ensures that a sample from the output distribution of $\hat{F}(b, \boldsymbol{x}, \boldsymbol{r_1}, \ldots, \boldsymbol{r_n})$ can be obtained via the following two-step process: (1) sample an output from $F(b, \boldsymbol{x})$; (2) independently replace each bit σ_i in this output by a fresh sample from the RE output distribution D_{σ_i} corresponding to $\hat{f}_i$. In particular, the joint distribution of every k bits in the output of $\hat{F}$ is fully determined by the joint distribution of a corresponding set of k bits in the output of F. If the latter is insensitive to the choice of b, then so is the former.

Finally, to see that the distinguishing advantage of $\hat{C}$ is at least $\hat{\epsilon}$, note that conditioned on the good event that none of the decoders Dec_i errs, which occurs with at least $1 - \delta$ probability, the distinguishing advantage of $\hat{C}$ is the same as that of C. □

8.2 Useful instances

In order to effectively apply Lemma 8.2 to instances of our problem with AC^0 distinguishers, we need the RE decoders Dec_i to be implemented in AC^0 and have low error probability. For instance, $\delta = o(1/n)$ suffices for constant ϵ.

There are two kinds of useful RE constructions in the literature. The first, which is implicit in the circuit lower bound proofs of Razborov [Raz87] and Smolensky [Smo87], can be used to encode an AC^0 function $f\colon \{0,1\}^m \to \{0,1\}$ by a degree-$\mathsf{polylog}(m)$ function $\hat{f}\colon \{0,1\}^m \times \{0,1\}^{\mathsf{poly}(m)} \to \{0,1\}$ with error $\delta = 2^{-\mathsf{polylog}(m)}$. (Here Dec is simply the identity function.) However, a crucial problem with this general

[7] The coset variant is equivalent to the main variant up to a difference of 1 in the locality and degree of the sampler, which will not matter for the sampler classes considered in this section. Alternatively, one could directly handle the main variant by encoding each source separately using a pair of encoders that share the same decoder.

construction is that it also has *privacy error* $\phi = 2^{-\mathsf{polylog}(m)}$, namely it only satisfies a relaxed version of Definition 8.1 where $\hat{f}(x, r)$ and $D_{f(x)}$ should be ϕ-close in statistical distance. This limitation seems inherent to the Razborov–Smolensky-based RE technique (see [BI05]), and is at odds with the goal of respecting (perfect) k-indistinguishability.[8] Fortunately, for decision trees or (more generally) *unambiguous* DNFs, it is possible to eliminate the privacy error completely while keeping the correctness error δ sufficiently small. See Lemma 5.10 for the exact quantitative statement. We leave open the question of obtaining a similar RE for general AC^0 or even just DNF:

Open Question 4. Does every AC^0 function $f\colon \{0,1\}^m \to \{0,1\}$, or even just DNF, admit a degree-$\mathsf{polylog}(m)$ (perfectly private) RE $\hat{f}\colon \{0,1\}^m \times \{0,1\}^{\mathsf{poly}(m)} \to \{0,1\}$ with error $\delta = 0.1$?

A different class of useful RE constructions encode functions f in complexity classes such as NC^1, $\oplus\mathsf{L}$, or NL by polynomial-size $\hat{f} \in \mathsf{NC}^0$. Unlike the Razborov–Smolensky-based RE, here $\hat{f}$ has multiple bits of output, and the correctness can be perfect in most cases (namely, $\delta = 0$). However, unlike the typical applications of such RE (see, e.g., [Ish13, App17]), where polynomial-time decoding suffices, here we must insist on decoding in AC^0. With $\hat{f}$ in NC^0, this is only possible when $f \in \mathsf{AC}^0$. The existence of such RE for every $f \in \mathsf{AC}^0$, or even just for the m-input OR function, is open; see Section 8.3 below. Here we observe that existing RE constructions suffice when applied to functions f with low (but super-constant) locality, yielding an AC^0-decodable RE with locality 4.

Claim 8.3 (Encoding formulas and branching programs [AIK06])**.** *Suppose $f\colon \{0,1\}^m \to \{0,1\}$ can be computed by a Boolean formula or branching program of size S. Then f admits a perfectly correct, degree 3, 4-local RE $\hat{f}\colon \{0,1\}^m \times \{0,1\}^{\mathsf{poly}(S)} \to \{0,1\}^{\mathsf{poly}(S)}$.*

Combining Lemma 8.2 with Claim 8.3, we get the following.

Theorem 8.4 (From low locality to locality 4)**.** *Let $\boldsymbol{X}, \boldsymbol{Y}$ be ℓ-local k-indistinguishable sources over $\{0,1\}^n$, where $\ell = o(\log\log n)$. Suppose $\boldsymbol{X}, \boldsymbol{Y}$ can be ϵ-distinguished by $\mathsf{poly}(n)$-size, depth-d AC^0 circuits. Then there exist k-indistinguishable degree 3, 4-local sources $\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}$ over $\{0,1\}^N$, where $N = n^{1+o(1)}$, that can be ϵ-distinguished by $\mathsf{poly}(n)$-size, depth $(d+1)$ AC^0 circuits.*

Proof. First, using b as a selector between the sources, both $\boldsymbol{X}, \boldsymbol{Y}$ can be sampled by a single "coset" sampler $F(b, \boldsymbol{x})$ with locality $\ell + 1$. Since each output f_i of F can be computed by a formula (or branching program) of size $2^{\ell+1}$, by Claim 8.3 it admits an RE $\hat{f}_i$ with output length $s = 2^{O(\ell)}$, with a "brute force" (depth 2) DNF or CNF decoder of size $2^{2^{O(\ell)}} = o(n)$. The sources $\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}$ are obtained by applying Lemma 8.2 to F with the RE $\hat{f}_i$, suitably using either a CNF or a DNF decoder to increase the depth of the distinguisher C by 1 instead of 2. □

Remark 8.5 (Extensions). If we allow the depth of the distinguisher in the conclusion of Theorem 8.4 to be $d + O(1)$ instead of $d + 1$, then the theorem can be extended in several ways. First, the $o(\log\log n)$ locality bound in the assumption can be relaxed to $O(\log\log n)$. In fact, it suffices to assume that each bit of the source is samplable by a $\mathsf{polylog}(n)$-size formula. This follows from the fact that, in the case of formulas, the decoder in Claim 8.3 has formula size $\mathsf{poly}(S)$. When $S = \mathsf{polylog}(n)$, a size-$\mathsf{poly}(S)$ formula can be emulated by a $\mathsf{poly}(n)$-size, $O(1)$-depth AC^0 circuit (cf. [Bus13], Theorem 1). Finally, in the case of $\mathsf{polylog}(n)$-local *linear* sources, a simple RE for linear functions (cf. [AIK06]) implies a similar conclusion where $\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}$ are *2-local* linear sources.

[8] The RE-based compiler from Lemma 8.2 can also be applied with a ϕ-private RE if one relaxes k-indistinguishability to tolerate a $k\phi$ statistical distance between k-projections of $\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}$. However, this relaxed form of k-indistinguishability, with inverse-quasipolynomial statistical error, is qualitatively different from the perfect one. In particular, it can totally break down when the projection set is chosen in an adaptive fashion, and it gives rise to simpler and stronger positive results that are not possible with perfect k-indistinguishability (or even with $2^{-\Omega(k)}$ error). See [BW17] for further discussion.

8.3 A conjecture on local RE with AC^0 decoders

To obtain a positive result for NC^0 sources with AC^0 distinguishers, settling Open Question 3 in the affirmative, it would suffice to obtain an RE in NC^0 with decoder in AC^0 for every $f \in \mathsf{AC}^0$, or even just for polynomial-size decision trees. We conjecture this to be impossible.

Conjecture 7. There are $f \in \mathsf{AC}^0$ that do *not* admit an RE in NC^0 with an AC^0 decoder.

In fact, we believe that Conjecture 7 holds even for $f = \mathsf{OR}$. Combining Lemma 8.2 with the decision-tree samplable sources from Theorem 5.1, we get that a negative answer to Open Question 3, which seems likely, would imply Conjecture 7 for decision trees. Concretely:

Claim 8.6 (RE barrier)**.** *Suppose every polynomial-size decision tree admits RE in* NC^0 *with* AC^0 *decoder. Then there exists a pair* $\boldsymbol{X}, \boldsymbol{Y}$ *of* NC^0*-samplable sources over* $\{0,1\}^n$ *that are* $n^{\Omega(1)}$*-indistinguishable and are* $\Omega(1)$*-distinguishable by* AC^0*.*

Thus, proving Conjecture 7 serves as a natural barrier to settling Open Question 3 in the negative.

9 Applications to Leakage-Resilient Cryptography

In this section, we describe cryptographic applications of our conjectures. Specifically, we show that these conjectures lead to better constructions of Leakage-Resilient Circuit Compilers (LRCC) [ISW03] that are secure against global leakage functions (such as AC^0 [FRR+14]). This section is organized as follows. In Section 9.1, we give the definition of an LRCC. In Section 9.2, we show that assuming the degree 1 version of our main conjecture (see Conjecture 3), the construction of stateless LRCC given in [ISW03, BIS19] for computing linear functions satisfies a stronger security property. In Section 9.3, we give an efficient construction of LRCC for computing general functions assuming the degree 2 version of our main conjecture.

9.1 Leakage-resilient circuit compilers

We give the definition of a leakage-resilient circuit compiler taken verbatim from [BIS19].

Definition 9.1 ($(\mathcal{L}, \epsilon)$-leakage resilient implementation)**.** Let $C\colon \{0,1\}^\ell \to \{0,1\}^m$ be a deterministic stateless circuit, $\mathcal{L}$ be a leakage class, and ϵ be an error parameter. We say that $(I, \widehat{C}, O)$ is a $(\mathcal{L}, \epsilon)$-leakage resilient implementation of C if:

- $I\colon \{0,1\}^\ell \to \{0,1\}^{\widehat{\ell}}$ is a randomized input encoder which maps an input x to an encoded input $\widehat{x}$.
- $\widehat{C}$ is a randomized circuit that maps an encoded input $\widehat{x}$ to an encoded output $\widehat{y} \in \{0,1\}^{\widehat{m}}$.
- $O\colon \{0,1\}^{\widehat{m}} \to \{0,1\}^m$ is the deterministic output decoder that maps an encoded output $\widehat{y}$ to y.
- **Correctness:** For every input $x \in \{0,1\}^\ell$, $\Pr[O(\widehat{C}(I(x))) = f(x)] = 1$ where the probability is over the random coins of I and $\widehat{C}$.
- **Security:** For any two inputs $x_0, x_1 \in \{0,1\}^\ell$, let $(\mathsf{W}_0, \widehat{y}_0) \Leftarrow \widehat{C}(I(x_0))$ and $(\mathsf{W}_1, \widehat{y}_1) \Leftarrow \widehat{C}(I(x_1))$ where W_0 (resp. W_1) represents the assignment to every wire of $\widehat{C}$ on input $I(x_0)$ (resp. $I(x_1)$). For any leakage function $\ell \in \mathcal{L}$, the statistical distance between $\ell(\mathsf{W}_0)$ and $\ell(\mathsf{W}_1)$ is at most ϵ.

Definition 9.2 (LRCC for Circuits)**.** Let λ be the security parameter and let $\mathcal{C}$ be a class of circuits taking ℓ input bits and having m output bits. A leakage resilient stateless circuit compiler for the class $\mathcal{C}$ is a tuple of polynomial-time algorithms $(\mathsf{Enc}, \mathsf{T}, \mathsf{Dec})$ where

- Enc is a randomized input encoder which maps the security parameter 1^λ and an input $x \in \{0,1\}^\ell$ to an encoded input $\widehat{x}$.

- T is a deterministic algorithm that maps the security parameter 1^λ and a deterministic circuit in $C \in \mathcal{C}$ to another randomized circuit $\widehat{C}$. $\widehat{C}$ maps an encoded input $\widehat{x}$ to an encoded output $\widehat{y}$.
- Dec is the deterministic output decoder that maps an encoded output $\widehat{y}$ to $y \in \{0,1\}^m$.

For a leakage class $\mathcal{L}(\lambda)$ and the error parameter $\epsilon(\lambda)$, we say that $(\mathsf{Enc}, \mathsf{T}, \mathsf{Dec})$ is a $(\mathcal{L}(\lambda), \epsilon(\lambda))$-leakage resilient circuit compiler for $\mathcal{C}$ if for every $C \in \mathcal{C}$, $(\mathsf{Enc}(1^\lambda, \star), \mathsf{T}(1^\lambda, C), \mathsf{Dec})$ is a $(\mathcal{L}(\lambda), \epsilon(\lambda))$-leakage resilient implementation of C.

Prior Work. The works of Rothblum [Rot12] and Bogdanov, Ishai, and Srinivasan [BIS19] gave constructions of leakage resilient circuit compiler that are secure against AC^0 circuits. Rothblum's [Rot12] construction relied on the IPPP conjecture whereas the work of [BIS19] gave an unconditional result with the same asymptotic efficiency as that of Rothblum. We now recall the main result from their work.

Theorem 9.3 ([BIS19]). *Let λ denote the security parameter, $d \in \mathbb{N}$ and let C denote the class of $\mathsf{poly}(\lambda)$ size circuits mapping ℓ input bits to m output bits. There exists a construction $(\mathsf{Enc}, \mathsf{T}, \mathsf{Dec})$ of a LRCC for C that is secure against leakage by size 2^λ, depth d circuits with error parameter $2^{-\lambda}$. Furthermore, for every $C \in \mathsf{C}$, the size of $\mathsf{T}(1^\lambda, C)$ is $O(\lambda^{2d}|C|)$.*

If we restrict the class C to only consist of XOR gates, then the size of $\mathsf{T}(1^\lambda, C)$ is $O(\lambda^d|C|)$.

9.2 Application 1: LRCC for linear functions

In this section, we prove that the construction of LRCC given in [ISW03, BIS19] for computing linear functions satisfies a stronger security property assuming the degree 1 version of our conjecture. Before we state the stronger property, we first recall the construction of [ISW03, BIS19] in Figure 4 for computing linear functions.

9.2.1 Stronger Security Property

The construction given in Figure 4 was proved in [BIS19] to unconditionally satisfy Definition 9.2 when the leakage class is restricted to AC^0 functions. However, this proof crucially relied on the fact that the wires of the output decoder are not subject to leakage and only the wires of $\mathsf{T}(1^\lambda, C)$ are subject to leakage by an AC^0 function. In other words, they assumed that the decoder is implemented using a trusted hardware whose wires are not subject to leakage. A natural question is whether this assumption is necessary or, can we show that even if the wires of the decoder are subject to leakage, the above construction is resilient against AC^0 leakage functions. Before we try to answer this question, we first augment Definition 9.1 to satisfy this stronger property (the correctness property remains the same).

Stronger Security: For any two inputs $x_0, x_1 \in \{0,1\}^\ell$ such that $C(x_0) = C(x_1)$, let $(\mathsf{W}_0, \widehat{y}_0) \Leftarrow O \circ \widehat{C}(I(x_0))$ and $(\mathsf{W}_1, \widehat{y}_1) \Leftarrow O \circ \widehat{C}(I(x_1))$ where W_0 (resp. W_1) represents the assignment to every wire of $O \circ \widehat{C}$ on input $I(x_0)$ (resp. $I(x_1)$). For any leakage function $\ell \in \mathcal{L}$, the statistical distance between $\ell(\mathsf{W}_0)$ and $\ell(\mathsf{W}_1)$ is at most ϵ.

In the above definition, it is necessary to restrict the two inputs x_0 and x_1 to have the same output as there is a trivial distinguishing attack in the case where the outputs are not the same.

We show is that assuming the degree 1 version of our main conjecture, the construction in Figure 4 satisfies the stronger security property when the leakage function is restricted to AC^0 circuits.

9.2.2 Proof of Stronger Security

Before we move on to our security analysis, we discuss some barriers to give a reduction to Braverman's theorem.

Let λ denote the security parameter. Let C be a linear function (i.e., consisting only of XOR gates) mapping ℓ input bits to m output bits.

- $\mathsf{Enc}(1^\lambda, x \in \{0,1\}^\ell)$:
  1. Parse x as $(x_1, \ldots, x_\ell)$ where each $x_i \in \{0,1\}$.
  2. For each $i \in [\ell]$, choose a random λ-out-of-λ additive secret sharing of x_i given by $(x_{i,1}, \ldots, x_{i,\lambda})$.
  3. Output $\{(x_{i,1}, \ldots, x_{i,\lambda})\}_{i \in [\ell]}$.
- $\mathsf{T}(1^\lambda, C \colon \{0,1\}^\ell \to \{0,1\}^m)$:
  1. Each wire w in the original circuit C is transformed into a bundle of λ wires $\mathbf{w}$ in $\mathsf{T}(1^\lambda, C)$. The invariant that will be maintained is that the parity of the wire bundle will be equal to the value carried by the wire.
  2. Each XOR gate in C taking in wires a and b is replaced by a gate gadget that takes in $\mathbf{a}$ and $\mathbf{b}$ and outputs $\mathbf{a} + \mathbf{b}$ (where $+$ denotes bitwise XOR).
- $\mathsf{Dec}(\widehat{y})$:
  1. For each output wire in C, compute the parity of the corresponding bundle in $T(1^\lambda, C)$ and output the parity.

Figure 4: [ISW03, BIS19] Construction for Linear Functions

Barriers in Reducing the Security to Braverman's theorem. Fix C to be the generator matrix of an expander code (see Section 7.2 for the definition) and C' be another linear function with a single output bit. Now, consider a linear function $C \| C'$ that takes an ℓ bit string (x_1, x_2) (where x_1, x_2 are $\ell/2$ bit strings) and outputs $C(x_1) \| C'(x_2)$.[9] We argue that there does not exist an NC^0 circuit that takes any input $(x_1, x_2) \in \{0,1\}^\ell$ and some k-wise independent distribution z and outputs the distribution of all the wires of $\mathsf{Dec} \circ \mathsf{T}(1^\lambda, C \| C')$ on $\mathsf{Enc}(1^\lambda, (x_1, x_2))$. If such a circuit exists, then there exists a reduction from the stronger security property to the Braverman's theorem. Assume for the sake of contradiction there exists such a circuit. Then, if x is randomly chosen then the output distribution of C corresponds to a random codeword of this expander code. However, by Proposition 7.4, this is impossible. We note that in Section 7.1, we gave an instance where the reduction to Braverman's theorem is possible when the function is computing the parity of all the bits. In other words, we showed that such a reduction is possible when the output of the linear function is a single bit. However, when we consider linear functions that may output multiple bits, there exist some barriers as mentioned above.

Security. We show that assuming the conjecture for linear sources stated below, the construction given in Figure 4 satisfies the stronger security property.

Conjecture 8 (MAINLIN$(\mathcal{C}, n', k, \epsilon)$ restated:)**.** Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be sources on n' bits that are samplable by linear functions. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are k-indistinguishable, then any circuit in $\mathcal{C}$ can only distinguish between $\boldsymbol{X}$ and $\boldsymbol{Y}$ with $\epsilon(n', k)$ advantage.

Theorem 9.4. *Let λ be the security parameter and let C be a circuit that computes a linear function from $\{0,1\}^\ell$ to $\{0,1\}^m$. Set n' to be equal to the number of wires in the Boolean implementation of* $\mathsf{Dec} \circ \mathsf{T}(1^\lambda, C)$

[9] We include C' so that there exists two different (x_1, x_2) and (x'_1, x'_2) s.t. $C \| C'(x_1, x_2) = C \| C'(x'_1, x'_2)$.

(described in Figure 4) and $k = \lambda - 1$. Assume $\mathrm{MAINLIN}(\mathcal{C}, n', k, \epsilon)$*. Then,* $(\mathsf{Enc}(1^\lambda, \star), \mathsf{T}(1^\lambda, C), \mathsf{Dec})$ *described in Figure 4 satisfies the stronger security property.*

Proof. Let x_0 and x_1 be two inputs such that $C(x_0) = C(x_1)$. Let $\mathsf{W}_0 \Leftarrow \mathsf{Dec} \circ \mathsf{T}(1^\lambda, C)(\mathsf{Enc}(1^\lambda, x_0))$ and $\mathsf{W}_1 \Leftarrow \mathsf{Dec} \circ \mathsf{T}(1^\lambda, C)(\mathsf{Enc}(1^\lambda, x_1))$ where W_0 (resp. W_1) represents the assignment to every wire in the Boolean implementation of $\mathsf{Dec} \circ \mathsf{T}(1^\lambda, C)$ on input $\mathsf{Enc}(1^\lambda, x_0)$ (resp. $\mathsf{Enc}(1^\lambda, x_1)$). Assume for the sake of contradiction that there exists a function $g \in \mathcal{C}$ such that the statistical distance between $g(\mathsf{W}_0)$ and $g(\mathsf{W}_1)$ is greater than $\epsilon(n', k)$. We will show that this contradicts Conjecture 8.

Consider two sources $\mathbf{X_0}$ and $\mathbf{X_1}$ where $\mathbf{X_b}$ is same as W_b. We first observe that $\mathbf{X_0}$ and $\mathbf{X_1}$ are both linear sources since the circuit C consists only of XOR gates. We now argue that $\mathbf{X_0}$ and $\mathbf{X_1}$ are k-wise indistinguishable. We, in fact, show a stronger property where we argue that the joint distribution of any k wires of $\mathsf{T}(1^\lambda, C)$ along with all the wires of Dec in W_0 and W_1 are identically distributed.

Note that each XOR gadget in $\mathsf{T}(1^\lambda, C)$ can be viewed as λ computational components where the j-th component involves computing the XOR of j-th bits of the bundles $\mathbf{a}$ and $\mathbf{b}$. We first consider a partition $P_1, \ldots, P_\lambda$ of the wires of $\mathsf{T}(1^\lambda, C)$ where for each $j \in [\lambda]$, P_j contains all the wires in the j-th computational component of each XOR gadget. Naturally, any set of k wires in $\mathsf{T}(1^\lambda, C)$ corresponds to at most k of these partitions. To complete the proof, it is sufficient to show that the joint distribution of wires in at most k partitions along with the wires of Dec in W_0 and W_1 are identically distributed. We argue this via an hybrid argument. Fix any k partition indices $i_1, \ldots, i_k$.

- Hyb_0 : This corresponds to the joint distribution of all the wires in $P_{i_1}, \ldots, P_{i_k}$ along with the wires of Dec in W_0.
- Hyb_1 : In this hybrid, we do not make any changes to the distribution of the wires in $P_{i_1}, \ldots, P_{i_k}$ but change the input to Dec, namely $\widehat{y}$, to be a fresh additive sharing of $C(x_0)$ conditioned on the values in $\{P_j\}_{j \in \{i_1, \ldots, i_k\}}$. This hybrid is identically distributed to the previous hybrid since the partition $\{P_j\}_{j \notin \{i_1, \ldots, i_k\}}$ is not revealed.
- Hyb_2 : In this hybrid, we change the distribution of $P_{i_1}, \ldots, P_{i_k}$ to be sampled from W_1. This hybrid is identically distributed to the previous hybrid from the security of additive secret sharing.
- Hyb_3 : In this hybrid, we reverse the change made in Hyb_1. Since $C(x_0) = C(x_1)$, via an identical argument as before, this hybrid is identically distributed to the previous hybrid. Note that Hyb_3 is identical to $P_{i_1}, \ldots, P_{i_k}$ and the wires of Dec sampled from the distribution W_1.

Thus, $\mathbf{X_0}$ and $\mathbf{X_1}$ satisfy the premise of Conjecture 8 and hence, the existence of a function g mentioned above contradicts this conjecture. □

Assume that our degree 1 conjecture has the same error parameter as Braverman's theorem (the formal statement appears below).

Conjecture 9. Let λ denote the security parameter and $d \in \mathbb{N}$. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be sources on $\mathsf{poly}(\lambda)$ bits that are samplable by linear functions. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $\lambda^{O(d)}$-indistinguishable, then any circuit of size 2^λ and depth d can only distinguish between $\boldsymbol{X}$ and $\boldsymbol{Y}$ with $2^{-\lambda}$ advantage.

By setting $k = \lambda^{O(d)}$ in the above construction (i.e., using a $(k+1)$-out-of-$(k+1)$ additive secret sharing scheme), we get the following corollary.

Corollary 9.5. *Let λ denote the security parameter, $d \in \mathbb{N}$ and let C be a circuit computing a linear function from $\{0,1\}^\ell$ to $\{0,1\}^m$. Assume Conjecture 9. Then, the construction described in Figure 4 satisfies the stronger security property against leakage by 2^λ-size and depth-d circuits with error parameter $2^{-\lambda}$. Furthermore, the size of $\mathsf{T}(1^\lambda, C)$ is $O(|C|\lambda^{O(d)})$.*

9.2.3 Efficiency Improvement

We observe that we can improve the efficiency of this construction further. We first observe that the proof of Theorem 9.4 holds if we use a linear secret sharing scheme that is secure against k corruptions. Thus, instead of relying on additive secret sharing, we can use packed Shamir secret sharing [FY92] over a finite field of characteristic 2 as the underlying secret sharing scheme. Specifically, we consider packed secret sharing scheme among n parties where the security holds against $k = \lambda^{O(d)}$ parties. We set $n = O(k)$ and pack $O(n)$ secrets in each sharing. As a result, we get the following corollary (using a proof that is similar to Theorem 9.4).

Corollary 9.6. *Let λ denote a security parameter, $d \in \mathbb{N}$, and C be a circuit computing a linear function mapping ℓ bits to m bits. Assume Conjecture 9. Then, for any ρ, there exists a leakage-resilient implementation of ρ instantiations of C (on possibly different inputs) that satisfies stronger security against 2^{λ}-size and depth-d circuits with error parameter $2^{-\lambda}$. Furthermore, for large enough ρ, the size of $\mathsf{T}(1^{\lambda}, (1^{\rho}, C))$ is $\widetilde{O}(\rho|C|)$.*

9.3 Application 2: LRCC for general functions

In this section, we give an efficient construction of LRCC for general functions that is resilient to global leakage functions such as AC^0. We show the security of this construction based on the degree-2 variant of our main conjecture. The construction uses the following building blocks.

9.3.1 Building Blocks

Multiplicative Secret Sharing. We make use of a multiplicative secret sharing scheme which is a generalization of threshold secret sharing discussed below.

Definition 9.7 (Threshold Secret Sharing)**.** A t-out-of-n threshold secret sharing scheme for secrets in $\mathbb{F}$ is a tuple of algorithms $(\mathsf{Share}, \mathsf{Rec})$ where:

- Share is a randomized algorithm that takes a secret $s \in \mathbb{F}$ and outputs $(\mathsf{Sh}_1, \ldots, \mathsf{Sh}_n)$ where each $\mathsf{Sh}_i \in \mathbb{F}$.
- Rec is a deterministic algorithm that takes $T \subseteq [n]$ of size at least t and $\{\mathsf{Sh}_i\}_{i \in T}$ and outputs the secret s.
- **Correctness:** For any secret $s \in \mathbb{F}$ and for any $T \subseteq [n]$ of size at least t, we have:

$$\Pr[\mathsf{Rec}(T, \{\mathsf{Sh}_i\}_{i \in T}) = s] = 1$$

  where $(\mathsf{Sh}_1, \ldots, \mathsf{Sh}_n) \leftarrow \mathsf{Share}(s)$.

- **Secrecy:** For any two secrets $s_0, s_1 \in \mathbb{F}$ and for any $K \subset [n]$ of size $< t$, we have:

$$\left\{\{\mathsf{Sh}_i\}_{i \in K} : (\mathsf{Sh}_1, \ldots, \mathsf{Sh}_n) \leftarrow \mathsf{Share}(s_0)\right\} \equiv \left\{\{\mathsf{Sh}_i\}_{i \in K} : (\mathsf{Sh}_1, \ldots, \mathsf{Sh}_n) \leftarrow \mathsf{Share}(s_1)\right\}$$

Definition 9.8 (Multiplicative Secret Sharing)**.** A t-out-of-n (for $n \geq 2t - 1$) threshold secret sharing $(\mathsf{Share}, \mathsf{Rec})$ for secrets in a finite field $\mathbb{F}$ (where $|\mathbb{F}| > n$) is said to be a multiplicative secret sharing if:

1. Share is a linear function of the input and the randomness.
2. There exists an linear function Rec' such that for any $(a_1, \ldots, a_n)$ which is in the support of $\mathsf{Share}(a)$ for some $a \in \mathbb{F}$ and $(b_1, \ldots, b_n)$ which is in the support of $\mathsf{Share}(b)$ for some $b \in \mathbb{F}$, we have $\mathsf{Rec}'(a_1 \cdot b_1, \ldots, a_n \cdot b_n) = a \cdot b$.

We note that Shamir secret sharing [Sha79] is an example of a multiplicative secret sharing.

Arithmetic Emulation of a Boolean Circuit. Let C be a circuit that maps ℓ bit inputs to m bit outputs. We consider the characteristic 2 finite field $\mathbb{F} = \mathbb{F}_{2^q}$ and consider the circuit C' over $\mathbb{F}^\ell \to \mathbb{F}^m$ which is same as C except that each XOR gate is replaced with $+$ and each AND gate is replaced with $\times$. Note that 0 and 1 in the Boolean world naturally map to elements 0 and 1 in $\mathbb{F}_{2^q}$. Hence, for any input $x \in \{0,1\}^n$, $C(x) = C'(x)$. We call this C' to be the arithmetic emulation of C. One useful property of $\mathbb{F}_{2^q}$ is that the addition operation is implemented using a Boolean circuit where each wire is a linear function of the input bits. Another property is that the multiplication operation is implemented by a Boolean circuit where each wire of this circuit is a quadratic function of the input bits.

Linear Circuit Encoder. A linear circuit encoder is obtained by the circuit implementation of an MPC protocol for computing linear functions on inputs held by different parties which is secure against all but one corruptions.

Lemma 9.9 (Linear Circuit Encoding)**.** *Let $f \colon \mathbb{F}^n \times \mathbb{F}^r \to \mathbb{F}^n$ be a randomized linear function that takes in n field elements as inputs, r field elements as randomness and outputs n field elements. There exists another randomized linear circuit $\widehat{f} \colon \mathbb{F}^n \times \mathbb{F}^{r'} \to \mathbb{F}^n$ such that:*

1. *For any input $x \in \mathbb{F}^n$, the output distribution of f on input x is identically distributed to the output distribution of $\widehat{f}$ on input x.*
2. *There exists n partitions $P_1, \ldots, P_n$ of the wires of the Boolean circuit implementation of $\widehat{f}$ such that:*
   - *(a) The bundle of wires representing the i-th input and the i-th output wire belong to partition P_i.*
   - *(b) For any input $x \in \mathbb{F}^n$ and any subset $K \subset [n]$ of size at most $n-1$, the joint distribution of the wires in the partitions $\{P_i\}_{i \in K}$ when $\widehat{f}$ is run on input x and uniform randomness is identically distributed to a randomized function $D(\{x_i, y_i\}_{i \in K})$ where y_i is the i-th output of $f(x; U_{\mathbb{F}^r})$.*

9.3.2 Leakage-Resilient Circuit Compiler for General Functions

Construction. We give the description of the construction in Figure 5.

Security. We state the quadratic version of the conjecture which is used in our main theorem.

Conjecture 10 (Quadratic Variant (MAINQUAD$(\mathcal{C}, n', k, \epsilon)$))**.** Let $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ be quadratic sources over n' bits. If $(\boldsymbol{X}_0, \boldsymbol{X}_1)$ are k-indistinguishable, then the pair $\boldsymbol{X}_0, \boldsymbol{X}_1$ $\epsilon(n', k)$-fool any circuit $C \in \mathcal{C}$.

Theorem 9.10. *Let λ be the security parameter and let C be an arbitrary circuit mapping $\{0,1\}^\ell \to \{0,1\}^m$. Set n' to be equal to the number of wires in $\mathsf{Dec} \circ \mathsf{T}(1^\lambda, C)$ (described in Figure 5) and $k = t - 1$. Assume* MAINQUAD$(\mathcal{C}, n', k, \epsilon)$. *Then,* $(\mathsf{Enc}(1^\lambda, \star), \mathsf{T}(1^\lambda, C), \mathsf{Dec})$ *described in Figure 5 is a $(\mathcal{C}, \epsilon(n', k))$-leakage resilient implementation of C.*

Proof. Let x_0 and x_1 be two inputs such that $C(x_0) = C(x_1)$. Let $\mathsf{W}_0 \Leftarrow \mathsf{T}(1^\lambda, C)(\mathsf{Enc}(1^\lambda, x_0))$ and $\mathsf{W}_1 \Leftarrow \mathsf{T}(1^\lambda, C)(\mathsf{Enc}(1^\lambda, x_1))$ where W_0 (resp. W_1) represents the assignment to every wire of the Boolean implementation of $\mathsf{T}(1^\lambda, C)$ on input $\mathsf{Enc}(1^\lambda, x_0)$ (resp. $\mathsf{Enc}(1^\lambda, x_1)$). Assume for the sake of contradiction that there exists a function $h \in \mathcal{C}$ such that the statistical distance between $h(\mathsf{W}_0)$ and $h(\mathsf{W}_1)$ is greater than $\epsilon(n', k)$. We will show that this contradicts Conjecture 10.

Consider two sources $\mathbf{X_0}$ and $\mathbf{X_1}$ where $\mathbf{X_b}$ is same as W_b. In the following two claims, we show that $\mathbf{X_0}$ and $\mathbf{X_1}$ are quadratic sources and show that they are k-wise indistinguishable.

Claim 9.11. *For each $b \in \{0,1\}$, $\mathbf{X_b}$ is a quadratic source.*

Proof. Let $\widehat{C} = \mathsf{T}(1^\lambda, C)$. We first argue that (i) the output wire bundle $(w_1, \ldots, w_n)$ of each gate gadget in $\widehat{C}$ is in the support of $\mathsf{Share}(w)$ where w is the actual value carried by that wire in C when run on input x_b, and (ii) the elements of the wire bundle are an affine function L_w of the randomness used in $\widehat{C}$ and Enc. We

Let λ be the security parameter and let $\mathbb{F} = \mathbb{F}_{2^q}$. Set $t = \lambda$ and $n \geq 2t - 1$. Let $C\colon \{0,1\}^\ell \to \{0,1\}^m$. Let $(\mathsf{Share}, \mathsf{Rec})$ be a t-out-of-n multiplicative threshold secret sharing scheme over $\mathbb{F}$.

- $\mathsf{Enc}(1^\lambda, x \in \{0,1\}^\ell)$:
  1. Parse x as $(x_1, \ldots, x_\ell)$ where each x_i is 0 or 1 in $\mathbb{F}$.
  2. For each $i \in [\ell]$, compute $(\widehat{x}_{i,1}, \ldots, \widehat{x}_{i,n}) \leftarrow \mathsf{Share}(x_i)$.
  3. Output $\{(\widehat{x}_{i,1}, \ldots, \widehat{x}_{i,n})\}_{i \in [\ell]}$.
- $\mathsf{T}(1^\lambda, C\colon \{0,1\}^\ell \to \{0,1\}^m)$:
  1. Let $C'\colon \mathbb{F}^\ell \to \mathbb{F}^m$ be the arithmetic emulation of C.
  2. Every wire $w \in \mathbb{F}$ in C' is replaced by a wire bundle $(w_1, \ldots, w_n) \in \mathbb{F}^n$ in $\widehat{C} = \mathsf{T}(1^\lambda, C)$ that represents a t-out-of-n secret sharing of the value carried in w.
  3. Each addition gate in C' taking in wires a and b is replaced by a gate gadget in $\widehat{C}$ that takes wire bundles $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ and outputs another bundle $(a_1 + b_1, \ldots, a_n + b_n)$.
  4. Each multiplication gate in C' taking wires a and b is replaced by a gate gadget in $\widehat{C}$ that takes wire bundles $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ that does the following:
     (a) It computes $(a_1 \cdot b_1, \ldots, a_n \cdot b_n)$.
     (b) Let g be the randomized linear circuit that computes $\mathsf{Share} \circ \mathsf{Rec}'$ and let $\widehat{g}$ be the randomized linear circuit implementation of g from Lemma 9.9.
     (c) It computes $(c_1, \ldots, c_n) \leftarrow \widehat{g}(a_1 \cdot b_1, \ldots, a_n \cdot b_n)$.
     (d) It outputs $(c_1, \ldots, c_n)$.
- $\mathsf{Dec}(\widehat{y} \in \mathbb{F}^{mn})$:
  1. Parse $\widehat{y}$ as $\{\widehat{y}_{i,1}, \ldots, \widehat{y}_{i,n}\}_{i \in [m]}$.
  2. For each $i \in [m]$, output $y_i := \mathsf{Rec}([n], \widehat{y}_{i,1}, \ldots, \widehat{y}_{i,n})$.

Figure 5: Leakage-Resilient Circuit Compiler for General Functions

show this through an induction on the depth of the circuit. The base case is the input wire bundles and the induction hypothesis trivially holds. Assume that the hypothesis holds for all wire bundles up to a depth d from the input level. Let $(w_1, \ldots, w_n)$ be a wire bundle in depth $d+1$ and let w be the value carried by the wire in C. If $(w_1, \ldots, w_n)$ is the output of an addition gate gadget, then the induction hypothesis holds from the fact that Share is linear function (see Property 1 in Definition 9.8). If $(w_1, \ldots, w_n)$ is the output bundle of a multiplication gadget, it follows from Property 2 of Definition 9.8 that the output distribution of $g = \mathsf{Share} \circ \mathsf{Rec}'$ is identical to $\mathsf{Share}(w)$ (with uniform randomness). Since Property 1 of Lemma 9.9 states that the output distribution of $\widehat{g}$ is identical to g, the induction hypothesis holds.

We now argue that for any gate gadget, the internal wires of the Boolean circuit implementation of the gadget is a quadratic function of the input to this gadget and the random bits. For the case of addition gadget, this follows directly from property of $\mathbb{F}_{2^q}$ that the addition operation can be implemented by a Boolean circuit which is computing a linear function on the input bits. For the case of multiplication gadget, we first observe that since multiplication over $\mathbb{F}_{2^q}$ can be implemented by a Boolean circuit where every wire is a quadratic function of the input bits, there exists a Boolean circuit computing $(a_1 \cdot b_1, \ldots, a_n \cdot b_n)$ where each wire of this circuit is a quadratic function of the bits representing $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$. Since $\widehat{g}$ is a linear randomized function, it again follows from the above mentioned property of computing addition operations over $\mathbb{F}_{2^q}$ that the wires of $\widehat{g}$ is a linear function of the input and the random bits and hence, each wire of the multiplication gate gadget is a quadratic function of the input and random bits.

Since (i) the input wire bundle to each gate gadget is a linear function of the randomness and the value carried by the wire, and (ii) the internal wires of each gate gadget is a quadratic function of the input and random bits, it follows that $\mathbf{X}_b$ is a quadratic source. □

Claim 9.12. *$\mathbf{X_0}$ and $\mathbf{X_1}$ are k-indistinguishable.*

Proof. Let $\widehat{C} = \mathsf{T}(1^\lambda, C)$. We view the execution of $\widehat{C}$ as the joint computation done by each party in an n-party protocol. Specifically, in this protocol, the j-th party receives $\{x_{i,j}\}_{i\in[\ell]}$ as the initial input. The parties now jointly compute each gate of the circuit C via a secure distributed protocol where at the end of the protocol the parties hold a secret sharing of the output.

Specifically, to compute an addition gate where the parties hold a t-out-of-n secret sharing of the input wires, each party locally adds the shares to obtain the share of the output. This precisely corresponds to the implementation of the addition gate gadget in $\widehat{C}$. To compute a multiplication gate, the j-th party locally multiply the shares of the inputs to obtain $a_j \cdot b_j$. Now, the parties run a distributed protocol to securely implement the randomized linear functionality $\mathsf{Share} \circ \mathsf{Rec}'$. This precisely corresponds to running the circuit $\widehat{g}$ from Lemma 9.9.

Via a standard argument (see [AL17, Theorem 4.2]), this protocol can be shown to be perfectly secure against k corruptions. To show k-wise indistinguishability, we consider a partition $P_1, \ldots, P_n$ of the wires of $\widehat{C}$ where for each $j \in [n]$, P_j corresponds to all the computation done by the j-th party in the above described multiparty protocol. Naturally, any set of k wires in $\widehat{C}$ corresponds to at most k of these partitions. To complete the proof, it is sufficient to show that the joint distribution of wires in at most k partitions in W_0 and W_1 are identically distributed. This directly follows from the security of above MPC protocol against k corruptions.

□

From Claim 9.11 and Claim 9.12, we infer that $\mathbf{X_0}$ and $\mathbf{X_1}$ satisfy the premise of Conjecture 10. Thus, the function $h \in \mathcal{C}$ such that the statistical distance between $h(\mathbf{X_0})$ and $h(\mathbf{X_1})$ is greater than $\epsilon(n', k)$ contradicts this conjecture. □

Remark 9.13. We note that the above construction can be proved to satisfy the stronger security property given in the previous subsection. Specifically, the k-wise indistinguishability of $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ holds even when the wires of the output decoder are leaked. This actually follows from the security property of the MPC protocol.

Assume the quadratic version of our main conjecture has the same parameters as that of Braverman's theorem.

Conjecture 11. Let λ denote the security parameter and $d \in \mathbb{N}$. Let $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ be quadratic sources over $\mathsf{poly}(\lambda)$ bits. If $(\boldsymbol{X}_0, \boldsymbol{X}_1)$ are $\lambda^{O(d)}$-indistinguishable, then the pair $\boldsymbol{X}_0, \boldsymbol{X}_1$ $2^{-\lambda}$-fool any 2^λ size and depth d circuit.

By setting $t = \lambda^{O(d)}$ (instead of λ) and instantiating the linear circuit encoder with a protocol from [BGW88]), we get

Corollary 9.14. *Let λ be the security parameter, $d \in \mathbb{N}$ and let C be an arbitrary $\mathsf{poly}(\lambda)$ size circuit mapping $\{0,1\}^\ell \to \{0,1\}^m$. Assume Conjecture 11. Then, $(\mathsf{Enc}(1^\lambda, \star), \mathsf{T}(1^\lambda, C), \mathsf{Dec})$ described in Figure 5 is a leakage resilient implementation of C against leakage by size 2^λ and depth d circuits with error parameter $2^{-\lambda}$. Furthermore, the size of $\mathsf{T}(1^\lambda, C)$ is $\tilde{O}(\lambda^{O(d)}|C|)$.*

Though the efficiency of this construction is worse than that of [BIS19], we build on this construction and give a more efficient instantiation of LRCC in the next subsection.

9.3.3 Extension to SIMD Circuits

We now give a leakage-resilient circuit compilers for special class of circuits called SIMD circuits. In a SIMD circuit, each gate is replaced with an instruction (either + or $\times$) that act on multiple data points. Further, we will assume a special structure on the data points that are given as inputs to each instruction (see Figure 6). We then rely on routing networks as in [DIK10] to transform any general circuit to a SIMD circuit (with a poly logarithmic overhead).

Construction. We give the construction of a leakage resilient circuit compiler for SIMD circuits in Figure 6. The difference between this construction and one presented in Figure 5 is that here, we use a packed secret sharing scheme. Specifically, we use packed t-out-of-n Shamir secret sharing and let $t' = O(t)$ denote the number of secrets that are packed. Note that packed secret sharing is multiplicative (see Definition 9.8).

Proof of Security. The proof of security follows via an identical argument to the proof of Theorem 9.10.

Theorem 9.15. *Let λ be the security parameter and let C be an arbitrary SIMD circuit mapping $\mathbb{F}^\ell \to \mathbb{F}^m$. Set n' to be equal to the number of wires in $\mathsf{Dec} \circ \mathsf{T}(1^\lambda, C)$ (described in Figure 6) and $k = t - 1$. Assume $\mathrm{MAINQUAD}(\mathcal{C}, n', k, \epsilon)$. Then, $(\mathsf{Enc}(1^\lambda, \star), \mathsf{T}(1^\lambda, C), \mathsf{Dec})$ described in Figure 6 is a $(\mathcal{C}, \epsilon(n', k))$-leakage resilient implementation of C.*

Instantiation. We will instantiate the construction in Figure 6 with the linear circuit encoder from [DN07].

Theorem 9.16 ([DN07])**.** *Let $g = \mathsf{Share} \circ \mathsf{Rec}'$. There exists a linear circuit $\widehat{g}$ that satisfies the conditions of Lemma 9.9 and for large enough ρ, implementing ρ instantiations of $\widehat{g}$ (on possibly different inputs) can be done by a circuit of size $\widetilde{O}(\rho \cdot n)$.*

By setting $t = \lambda^{O(d)}$ (instead of λ) and using the above linear circuit encoder in Figure 6, we get the following corollary.

Corollary 9.17. *Let λ be the security parameter, $d \in \mathbb{N}$ and let C be an arbitrary $\mathsf{poly}(\lambda)$ size SIMD circuit mapping $\mathbb{F}^\ell \to \mathbb{F}^m$. Suppose that Conjecture 11 holds. Then there exists a leakage-resilient implementation $(\mathsf{Enc}(1^\lambda, \star), \mathsf{T}(1^\lambda, C), \mathsf{Dec})$ of C that is secure against circuits of size 2^λ and depth d and error parameter $2^{-\lambda}$. Furthermore, for large enough $|C|$, the size of $\mathsf{T}(1^\lambda, C)$ (described in Figure 6) is $\widetilde{O}(|C| \cdot \lambda^{O(d)})$.*

Damgård et al. [DIK10] gave a transformation from arbitrary circuits to SIMD circuits that incurs a polylogarithmic (amortized) overhead to the circuit size. The transformation uses a routing network to compile an arithmetic circuit C (which is assumed to be at least ℓ gates wide) into another circuit C' with the following properties:

1. $C'(x) = C(x)$ for all inputs x.
2. Every layer of C' contains only one type of gate.
3. If the values in each layer are arranged in blocks of size ℓ (a power of 2), the action between any two layers to achieve correct line-up is to permute the blocks and then in some blocks permute the elements within the block, where the same permutation applies to all blocks in the layer. In the entire circuit, only $\log \ell$ different permutations are needed to handle permutations within blocks.
4. $|C'| = O(|C| \log |C| + d^2 \lambda \log^3 |C|)$ where d is the depth of the circuit C and depth of C' is $O(d \log^2 |C|)$.

To achieve permutation π within each block, we generate a packed secret sharing of r and $\pi(r)$. To permute a block x which is packed secret shared, we add the corresponding shares of r and x and then reconstruct to obtain $x + r$. We then apply the permutation π on this value to obtain $\pi(x) + \pi(r)$. We then subtract this value from secret shares of $\pi(r)$ to obtain a secret sharing of $\pi(x)$. We generate the sharing of r and $\pi(r)$ with $\widetilde{O}(1)$ amortized cost using the protocol from [DN07] (this is possible since there are only $\log l$ different permutations needed). Thus, the total computational cost needed to achieve the correct permutation within each block throughout the circuit is $\widetilde{O}(|C'|)$.

Corollary 9.18. *Let λ be the security parameter, $d \in \mathbb{N}$ and let C be an arbitrary $\mathsf{poly}(\lambda)$ circuit mapping $\{0,1\}^\ell \to \{0,1\}^m$. Suppose that Conjecture 11 holds. Then there exists a leakage-resilient implementation $(\mathsf{Enc}(1^\lambda, \star), \mathsf{T}(1^\lambda, C), \mathsf{Dec})$ of C that is secure against leakage by size 2^λ and depth d circuits with error parameter $2^{-\lambda}$. Furthermore, for large enough $|C|$, the size of $\mathsf{T}(1^\lambda, C)$ is $\widetilde{O}(|C| \log |C| + d^2 \lambda \log^3 |C|)$.*

Let λ be the security parameter and let $\mathbb{F} = \mathbb{F}_{2^q}$. Set $t = \lambda$, $t' = O(t)$ and $n = \Omega(t)$. Let $C : \mathbb{F}^\ell \to \mathbb{F}^m$ be a SIMD circuit composed of $+$ and $\times$ instructions where each instruction acts on t' data points. Specifically, $+((z_1, \ldots, z_{t'}), (z'_1, \ldots, z'_{t'})) = (z_1 + z'_1, \ldots, z_{t'} + z'_{t'})$ (and defined similarly for $\times$ gate). Further, we assume that the data points $(z_1, \ldots, z_{t'})$ and $(z'_1, \ldots, z'_{t'})$ are the outputs of SIMD instructions from the previous layer.

- $\mathsf{Enc}(1^\lambda, (x_1, \ldots, x_\ell))$: Split the inputs into blocks of t' elements (with padding if necessary). For each block consisting of the elements $(x_{i_1}, \ldots, x_{i_{t'}})$,

  1. Choose a random polynomial p of degree $t + t' - 2$ such that for t' distinct elements $1, 2, \ldots, t'$ on the field, we have $p(j) = x_{i_j}$ for all $j \in [t']$.
  2. The encoding consists of $(p(a_1), \ldots, p(a_n))$ where $a_1, \ldots, a_n$ are distinct elements (different from $1, 2, \ldots, t'$) in the field.

  The encoding of the inputs is the concatenation of the encoding of each block.

- The transformer $\mathsf{Tr}(1^\lambda, C)$:

  1. Each set of t' data points in the original circuit C is transformed into a bundle of n field elements in $\widehat{C}$ where the field elements represent the shares of a packed secret sharing scheme where the secrets are the data points.
  2. A gate g in C acting on the data points $(a_1, \ldots, a_{t'})$ and $(b_1, \ldots, b_{t'})$ is replaced by a gadget $\widehat{g}(\mathbf{a}, \mathbf{b})$ where $\mathbf{a}$ and $\mathbf{b}$ are the packed secret shares corresponding to $(a_1, \ldots, a_{t'})$ and $(b_1, \ldots, b_{t'})$ respectively. The descriptions of $\widehat{+}$ and $\widehat{\times}$ are given below.
  3. $\widehat{+}(\mathbf{a}, \mathbf{b})$: It outputs $\mathbf{a} + \mathbf{b}$.
  4. $\widehat{\times}(\mathbf{a}, \mathbf{b})$:
     (a) It computes $\mathbf{c} = \mathbf{a} \star \mathbf{b}$ where $\star$ denotes point-wise product.
     (b) Let g be the randomized affine circuit that computes $\mathsf{Share} \circ \mathsf{Rec}'$ of the packed Shamir secret sharing and let $\widehat{g}$ be the randomized affine circuit implementation of g from Lemma 9.9.
     (c) It computes $\mathbf{d} \leftarrow \widehat{g}(\mathbf{c})$ and outputs it.

- $\mathsf{Dec}(1^\lambda, \mathbf{y}_1, \ldots, \mathbf{y}_{m'})$: To decode each block of the output, apply Lagrange's interpolation on $\mathbf{y}_i$ to obtain the i-th output block consisting of $(y_{(i-1)t'+1}, \ldots, y_{it'})$.

Figure 6: Description of LRCC based on packed secret sharing.

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. 4, 16

[AIK06] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography in nc^0. *SIAM J. Comput.*, 36(4):845–888, 2006. 3, 8, 44, 45

[AIM$^+$03] Kazuyuki Amano, Kazuo Iwama, Akira Maruoka, Kenshi Matsuo, and Akihiro Matsuura. Inclusion-exclusion for k-cnf formulas. *Inf. Process. Lett.*, 87(2):111–117, 2003. 10, 36

[AL17] Gilad Asharov and Yehuda Lindell. A full proof of the BGW protocol for perfectly secure multiparty computation. *J. Cryptol.*, 30(1):58–151, 2017. 53

[App17] Benny Applebaum. Garbled circuits as randomized encodings of functions: a primer. In Yehuda Lindell, editor, *Tutorials on the Foundations of Cryptography*, pages 1–44. Springer International Publishing, 2017. 45

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004. 1

[ASTS$^+$98] A. Ambainis, L.J. Schulman, A. Ta-Shma, U. Vazirani, and A. Wigderson. The quantum communication complexity of sampling. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No.98CB36280)*, pages 342–351, 1998. 4

[AW21] Shyan Akmal and Ryan Williams. Majority-3sat (and related problems) in polynomial time, 2021. 10

[Bab87] László Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Inf. Process. Lett.*, 26(1):51–53, 1987. 22

[BBC$^+$01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001. 1

[BDF$^+$22] Andrej Bogdanov, Krishnamoorthy Dinesh, Yuval Filmus, Yuval Ishai, Avi Kaplan, and Akshayaram Srinivasan. Bounded indistinguishability for simple sources. In *ITCS*, 2022. 1

[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. 16

[BG13] Eli Ben-Sasson and Ariel Gabizon. Extractors for polynomial sources over fields of constant order and small characteristic. *Theory Comput.*, 9:665–683, 2013. 2

[BGW88] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 1–10, 1988. 1, 39, 53

[BI05] Omer Barkol and Yuval Ishai. Secure computation of constant-depth circuits with applications to database search problems. In *Advances in Cryptology - CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005, Proceedings*, pages 395–411, 2005. 45

[BIL12] Chris Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for ac0-circuits. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 101–110, 2012. 4

[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. 1, 2, 4, 5, 46, 47, 48, 53

[BIVW16] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded indistinguishability and the complexity of recovering secrets. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part III*, volume 9816 of *Lecture Notes in Computer Science*, pages 593–618, 2016. 1, 2, 3, 6, 7, 16, 17

[BKT19] Mark Bun, Robin Kothari, and Justin Thaler. Quantum algorithms and approximating polynomials for composed functions with shared inputs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 662–678, 2019. 4, 16

[BL86] Ravi B. Boppana and J. C. Lagarias. One- way functions and circuit complexity. In *Structure in Complexity Theory, Proceedings of the Conference hold at the University of California, Berkeley, California, USA, June 2-5, 1986*, volume 223 of *Lecture Notes in Computer Science*, pages 51–65, 1986. 22

[Bra11] Mark Braverman. Poly-logarithmic independence fools bounded-depth boolean circuits. *Commun. ACM*, 54(4):108–115, 2011. 1, 4, 39

[BT13] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and markov-bernstein inequalities. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part I*, volume 7965 of *Lecture Notes in Computer Science*, pages 303–314, 2013. 1, 7, 8, 20

[BT16] Mark Bun and Justin Thaler. Dual polynomials for collision and element distinctness. *Theory Comput.*, 12:Paper No. 16, 34, 2016. 1

[BT18] Mark Bun and Justin Thaler. Approximate degree and the complexity of depth three circuits. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 116 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 35, 18. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018. 1

[BT19] Mark Bun and Justin Thaler. The large-error approximate degree of AC^0. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 145 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 55, 16. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019. 1

[BT20a] Mark Bun and Justin Thaler. Guest column: Approximate degree in classical and quantum computing. *SIGACT News*, 51(4):48–72, 2020. 1, 7, 18

[BT20b] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC^0. *SIAM J. Comput.*, 49(4), 2020. 1

[Bus13] Sam Buss. Lecture notes in math 262A: Circuit complexity, November 2013. 45

[BW17] Andrej Bogdanov and Christopher Williamson. Approximate bounded indistinguishability. In *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*, pages 53:1–53:11, 2017. 8, 45

[CCD88] David Chaum, Claude Crépeau, and Ivan Damgård. Multiparty unconditionally secure protocols (extended abstract). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 11–19, 1988. 1

[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. 4, 16

[CS16] Gil Cohen and Igor Shinkar. The complexity of DNF of parities. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, Cambridge, MA, USA, January 14-16, 2016*, pages 47–58, 2016. 4, 5, 16

[DGRV11] Zeev Dvir, Dan Gutfreund, Guy N. Rothblum, and Salil P. Vadhan. On approximating the entropy of polynomial mappings. In Bernard Chazelle, editor, *Innovations in Computer Science - ICS 2011, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings*, pages 460–475. Tsinghua University Press, 2011. 2

[DGW09] Zeev Dvir, Ariel Gabizon, and Avi Wigderson. Extractors and rank extractors for polynomial sources. *Comput. Complex.*, 18(1):1–58, 2009. 2

[DIK10] Ivan Damgård, Yuval Ishai, and Mikkel Krøigaard. Perfectly secure multiparty computation and the computational overhead of cryptography. In *Advances in Cryptology - EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Monaco / French Riviera, May 30 - June 3, 2010. Proceedings*, volume 6110 of *Lecture Notes in Computer Science*, pages 445–465, 2010. 2, 5, 54

[DN07] Ivan Damgård and Jesper Buus Nielsen. Scalable and unconditionally secure multiparty computation. In *Advances in Cryptology - CRYPTO 2007, 27th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2007, Proceedings*, volume 4622 of *Lecture Notes in Computer Science*, pages 572–590, 2007. 5, 54

[dW08] Ronald de Wolf. A note on quantum algorithms and the minimal degree of ϵ-error polynomials for symmetric functions. *Quantum Inf. Comput.*, 8(10):943–950, 2008. 23

[DW12] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Trans. Comput. Theory*, 4(1), March 2012. 4

[FRR$^+$14] Sebastian Faust, Tal Rabin, Leonid Reyzin, Eran Tromer, and Vinod Vaikuntanathan. Protecting circuits from computationally bounded and noisy leakage. *SIAM J. Comput.*, 43(5):1564–1614, 2014. 1, 5, 46

[FY92] Matthew K. Franklin and Moti Yung. Communication complexity of secure computation (extended abstract). In S. Rao Kosaraju, Mike Fellows, Avi Wigderson, and John A. Ellis, editors, *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 699–710. ACM, 1992. 5, 50

[GGN10] Oded Goldreich, Shafi Goldwasser, and Asaf Nussboim. On the implementation of huge random objects. *SIAM Journal on Computing*, 39(7):2761–2822, 2010. 4

[GMW87] Oded Goldreich, Silvio Micali, and Avi Wigderson. How to play any mental game or A completeness theorem for protocols with honest majority. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA*, pages 218–229, 1987. 1

[Gro94] Vince Grolmusz. A weight-size trade-off for circuits with MOD m gates. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 68–74, 1994. 16

[Gur10] Venkatesan Guruswami. Notes 8: Expander codes and their decoding, 2010. `https://www.cs.cmu.edu/~venkatg/teaching/codingtheory/notes/notes8.pdf`, Last accessed on 2021-05-09. 42

[HHL19] Hamed Hatami, Pooya Hatami, and Shachar Lovett. Higher-order fourier analysis and applications. *Found. Trends Theor. Comput. Sci.*, 13(4):247–448, 2019. 28

[HS10] Elad Haramaty and Amir Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 331–340, 2010. 24

[IK00] Yuval Ishai and Eyal Kushilevitz. Randomizing polynomials: A new representation with applications to round-efficient secure computation. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*, pages 294–304, 2000. 44

[IKO$^+$11] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, Manoj Prabhakaran, and Amit Sahai. Efficient non-interactive secure computation. In Kenneth G. Paterson, editor, *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*, volume 6632 of *Lecture Notes in Computer Science*, pages 406–425. Springer, 2011. 2

[Ish13] Yuval Ishai. Randomization techniques for secure computation. In Manoj Prabhakaran and Amit Sahai, editors, *Secure Multi-Party Computation*, volume 10 of *Cryptology and Information Security Series*, pages 222–248. IOS Press, 2013. 45

[ISW03] Yuval Ishai, Amit Sahai, and David A. Wagner. Private circuits: Securing hardware against probing attacks. In Dan Boneh, editor, *Advances in Cryptology - CRYPTO 2003, 23rd Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 2003, Proceedings*, volume 2729 of *Lecture Notes in Computer Science*, pages 463–481. Springer, 2003. 1, 5, 46, 47, 48

[Jac94] Jeffrey C. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. In *35th Annual Symposium on Foundations of Computer Science, Santa Fe, New Mexico, USA, 20-22 November 1994*, pages 42–53, 1994. 10, 14

[Juk06] Stasys Jukna. On graph complexity. *Comb. Probab. Comput.*, 15(6):855–876, 2006. 16

[KNY16] Ilan Komargodski, Moni Naor, and Eylon Yogev. How to share a secret, infinitely. In *TCC (B2)*, pages 485–514. Springer, 2016. 3, 6, 22

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *FOCS*, pages 168–177, 2016. 2

[LP07] Yehuda Lindell and Benny Pinkas. An efficient protocol for secure two-party computation in the presence of malicious adversaries. In Moni Naor, editor, *Advances in Cryptology - EUROCRYPT 2007, 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain, May 20-24, 2007, Proceedings*, volume 4515 of *Lecture Notes in Computer Science*, pages 52–78. Springer, 2007. 2

[LV11] Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity, CCC 2011, San Jose, California, USA, June 8-10, 2011*, pages 243–251, 2011. 4, 42

[MG02] Daniele Micciancio and Shafi Goldwasser. *Complexity of Lattice Problems - A Cryptograhic Perspective*, volume 671 of *The Kluwer international series in engineering and computer science*. Springer, 2002. 24

[MR04] Silvio Micali and Leonid Reyzin. Physically observable cryptography (extended abstract). In Moni Naor, editor, *Theory of Cryptography, First Theory of Cryptography Conference, TCC 2004, Cambridge, MA, USA, February 19-21, 2004, Proceedings*, volume 2951 of *Lecture Notes in Computer Science*, pages 278–296. Springer, 2004. 5

[MS78] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland Publishing Company, 2nd edition, 1978. 32

[NS92] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 462–467, 1992. 1, 16

[NS94] Moni Naor and Adi Shamir. Visual cryptography. In *Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings*, volume 950 of *Lecture Notes in Computer Science*, pages 1–12, 1994. 1, 2, 17

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '92, page 468–474, New York, NY, USA, 1992. Association for Computing Machinery. 1

[PRS88] Pavel Pudlák, Vojtěch Rödl, and Petr Savický. Graph complexity. *Acta Inform.*, 25(5):515–535, 1988. 16

[Rao09] Anup Rao. Extractors for low-weight affine sources. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 95–101. IEEE Computer Society, 2009. 2

[Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987. 3, 8, 20, 44

[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. 1, 5, 16, 47

[Sha79] Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979. 1, 50

[She13] Alexander A. Sherstov. Approximating the AND-OR tree. *Theory Comput.*, 9:653–663, 2013. 1

[She20] Alexander A. Sherstov. Algorithmic polynomials. *SIAM J. Comput.*, 49(6):1173–1231, 2020. 1

[Shi00] Yaoyun Shi. Lower bounds of quantum black-box complexity and degree of approximating polynomials by influence of Boolean variables. *Inform. Process. Lett.*, 75(1-2):79–83, 2000. 1

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA*, pages 77–82, 1987. 3, 8, 20, 44

[Spa08] Robert Spalek. A dual polynomial for OR, 2008. 7

[SV12] Rocco A. Servedio and Emanuele Viola. On a special case of rigidity. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:144, 2012. 4, 16

[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. 1, 4, 39

[Tre04] Luca Trevisan. A note on approximate counting for k-dnf. In Klaus Jansen, Sanjeev Khanna, José D. P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 417–425, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. 10

[Vio10] Emanuele Viola. The complexity of distributions. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 202–211, 2010. 4

[Vio12] Emanuele Viola. Extractors for turing-machine sources. In Anupam Gupta, Klaus Jansen, José Rolim, and Rocco Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 663–671, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. 4

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014. 4

[Vio16] Emanuele Viola. Quadratic maps are hard to sample. *ACM Trans. Comput. Theory*, 8(4), June 2016. 4

[Vio20] Emanuele Viola. Sampling lower bounds: Boolean average-case and permutations. *SIAM Journal on Computing*, 49(1):119–137, 2020. 4, 6

[Vio21] Emanuele Viola. Lower bounds for samplers and data structures via the cell-probe separator. *Electron. Colloquium Comput. Complex.*, 28:73, 2021. 4

[VT00] S. Vadhan and L. Trevisan. Extracting randomness from samplable distributions. In *2000 IEEE 41st Annual Symposium on Foundations of Computer Science*, page 32, Los Alamitos, CA, USA, nov 2000. IEEE Computer Society. 4

[Wil18] R. Ryan Williams. Counting solutions to polynomial systems via reductions. In Raimund Seidel, editor, *1st Symposium on Simplicity in Algorithms (SOSA 2018)*, volume 61 of *OpenAccess Series in Informatics (OASIcs)*, pages 6:1–6:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 10

[Yao86] Andrew Chi-Chih Yao. How to generate and exchange secrets. In *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, pages 162–167, 1986. 1

A Mixtures of iid

The main theorem of Section 5 constructs a pair of $\Theta(\sqrt{n})$-indistinguishable mixtures of iid over $\{0,1\}^n$ which OR can $\Omega(1)$-distinguish. The construction starts with an arbitrary pair of $\Theta(\sqrt{n})$-indistinguishable distributions over $\{0,1\}^{\Theta(n)}$ which OR can $\Omega(1)$-distinguish, and then applies a simple resampling procedure, Lemma 5.5. In this section, we outline a direct construction from first principles. In the interest of space, we skip some tedious calculations.

As in the proof of Theorem 5.11, we construct our indistinguishable pair from a measure which is orthogonal to all low-degree polynomials.

Lemma A.1. *Let $1 = x_0 > x_1 > \cdots > x_m \geq -1$, and let $f\colon \{x_0, \ldots, x_m\} \to \mathbb{R}$ satisfy*

$$\sum_{i=1}^{m} f(x_i)p(x_i) = 0$$

for all polynomials $p(x)$ of degree at most k.

Define two mixtures of iid over $\{0,1\}^n$ as follows. Let $X_+ = \{x_i : f(x_i) > 0\}$ and $X_- = \{x_i : f(x_i) < 0\}$.

The source $\boldsymbol{X}_+$ is sampled by first sampling a bias $x_i \in X_+$ with probability $2f(x_i)/\sum_j |f(x_j)|$, and then sampling n many independent $\frac{1-x_i}{2}$-biased bits.

The source $\boldsymbol{X}_-$ is sampled by first sampling a bias $x_i \in X_-$ with probability $-2f(x_i)/\sum_j |f(x_j)|$, and then sampling n many independent $\frac{1-x_i}{2}$-biased bits.

The two sources $\boldsymbol{X}_+, \boldsymbol{X}_-$ are k-indistinguishable, and OR can distinguish them with advantage at least

$$\frac{2|f(x_0)|}{\sum_j |f(x_j)|} - \left(\frac{1+x_1}{2}\right)^n.$$

Proof. By assumption $\sum_j f(x_j) = 0$, and this shows that $\sum_{x \in X_+} f(x) = \sum_{x \in X_-} -f(x) = \sum_j |f(x_j)|/2$. Therefore the measures according to which the biases in $\boldsymbol{X}_+, \boldsymbol{X}_-$ are sampled are indeed probability distributions.

Let $z \in \{0,1\}^k$ consist of a zeroes and $k-a$ ones. Then

$$\Pr[\boldsymbol{X}_+|_{\{1,\ldots,k\}} = z] - \Pr[\boldsymbol{X}_-|_{\{1,\ldots,k\}} = z] = \sum_{i=1}^{m} \frac{2f(x_i)}{\sum_j |f(x_j)|} \left(\frac{1+x_i}{2}\right)^a \left(\frac{1-x_i}{2}\right)^{k-a} = 0,$$

since the right-hand side is a polynomial of degree k. Therefore $\boldsymbol{X}_+, \boldsymbol{X}_-$ are k-indistinguishable.

Now suppose without loss of generality that $f(x_0) > 0$. On the one hand,

$$\Pr[\boldsymbol{X}_+ = 0] \geq \frac{2f(x_0)}{\sum_j |f(x_j)|}.$$

On the other hand,

$$\Pr[\boldsymbol{X}_- = 0] \leq \left(\frac{1+x_1}{2}\right)^n. \qquad \square$$

We will aim at $f(x_0) = \Theta(\sum_j |f(x_j)|)$ and $x_1 = 1 - \Theta(1/n)$, which will guarantee a constant distinguishing advantage for OR (given appropriate hidden constants).

Given arbitrary $k+2$ points $1 = x_0 > x_1 > \cdots > x_{k+1} \geq -1$, it turns out that the unique (up to a constant factor) function f satisfying the requirements of Lemma A.1 alternates in sign, and is given by

$$f(x_i) = \prod_{j \neq i} (x_i - x_j)^{-1}. \tag{6}$$

Of particular interest are the points $x_i = \cos(\theta_i)$, where

$$\theta_0, \theta_2, \theta_3, \ldots, \theta_{k+1} = \frac{0}{k}\pi, \frac{1}{k}\pi, \ldots, \frac{k}{k}\pi,$$

and $\theta_1 \in (0, \pi/k)$ is arbitrary. Tedious calculations show that up to a constant factor (which is different from the constant factor in (6)), the resulting f is given by

$$\begin{aligned}
f(0) &= (1 - \cos\theta_1)^{-1} \\
f(1) &= 4k(\cos((k+1)\theta_1) - \cos((k-1)\theta_1))^{-1} \\
f(i) &= (-1)^i 2(\cos(\theta_1) - \cos(\theta_i))^{-1} \\
f(k+1) &= (-1)^{k+1}(\cos(\theta_1) - \cos(\theta_{k+1}))^{-1}
\end{aligned}$$

where $2 \leq i \leq k$. Furthermore,

$$\frac{2|f(x_0)|}{\sum_j |f(x_j)|} = \left(1 + k \tan\frac{k\theta_1}{2} \tan\frac{\theta_1}{2}\right)^{-1}.$$

If $\theta_1 = \frac{\rho}{k}\pi$, where $\rho \in (0,1)$, then as $k \to \infty$ we have

$$\begin{aligned}
\frac{1-x_1}{2} &\sim \frac{\rho^2\pi^2}{4k^2} \\
\frac{2|f(x_0)|}{\sum_j |f(x_j)|} &\to \left(1 + \frac{\rho\pi}{2}\tan\frac{\rho\pi}{2}\right)^{-1}
\end{aligned}$$

Choosing $k = \Theta_\rho(\sqrt{n})$, Lemma A.1 construct a pair of k-indistinguishable mixtures of iid which OR distinguishes with constant advantage. We can get the weight-complexity down to $n^{O(n)}$ by discretizing the x_i to multiples of $\frac{1}{N}$ for $N = \Theta(n)$. The parameters appearing Lemma A.1 change only slightly, as the explicit formula (6) implies.

We close this section by mentioning the limiting form of the function $g(x) = f(x)/\sum_j |f(x_j)|$ obtained by fixing ρ and letting $k \to \infty$:

$$\begin{aligned}
g(0) &\to \left(2 + \rho\pi\tan\frac{\rho\pi}{2}\right)^{-1} \\
g(1) &\to -\left(\frac{2}{\rho\pi}\sin(\rho\pi) + 2\sin^2\frac{\rho\pi}{2}\right)^{-1} \\
g(i) &\to (-1)^i\left(\left(\left(\frac{i-1}{\rho}\right)^2 - 1\right)\left(1 + \frac{\rho\pi}{2}\tan\frac{\rho\pi}{2}\right)\right)^{-1}
\end{aligned}$$