Hard-To-Sample Distributions from Robust Extractors

Farzan Byramji
UC San Diego, La Jolla, CA, USA

Daniel M. Kane
UC San Diego, La Jolla, CA, USA

Jackson Morris
UC San Diego, La Jolla, CA, USA

Anthony Ostuni
UC San Diego, La Jolla, CA, USA

Abstract

We provide a unified method for constructing explicit distributions which are difficult for restricted models of computation to generate. Our constructions are based on a new notion of *robust extractors*, which are extractors that remain sound even when a small number of points violate the min-entropy constraint. Using such objects, we show that for a broad range of sampling models (e.g., low-depth circuits, small-space sources, etc.), every output of the model has distance $1 - o(1)$ from our target distribution, qualitatively recovering essentially all previously known hardness results. Our work extends that of Viola (SICOMP '14), who developed an earlier unified framework based on traditional extractors to rule out sampling with very small error.

As a further application of our technique, we leverage a recent extractor construction of Chattopadhyay, Goodman, and Gurumukhani (ITCS '24) to present the first explicit distribution with distance $1 - o(1)$ from the output of any low-degree $\mathbb{F}_2$-polynomial source. We note that a similar bound was obtained concurrently and independently by Khodabandeh and Shinkar (ECCC '26). We also describe a potential avenue toward proving a similar hardness result for $\mathsf{AC}^0[\oplus]$ circuits.

2012 ACM Subject Classification Theory of computation → Circuit complexity; Theory of computation → Pseudorandomness and derandomization

Keywords and phrases sampling, extractor, low-degree polynomials

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.37

Category RANDOM

Related Version *Full Version*: https://arxiv.org/abs/2604.26179

Funding *Farzan Byramji*: Supported by Simons Investigator Award #929894, and NSF Awards CCF-2425349 and AF: Medium 2212136.
Daniel M. Kane: Supported by NSF Medium Award CCF-2107547.
Anthony Ostuni: Partially supported by Simons Investigator Award #929894 and NSF Award CCF-2425349.

Acknowledgements We thank Jesse Goodman and Mohit Gurumukhani for a number of helpful comments on an earlier draft. In particular, we are grateful for the suggestion of and discussion about "smooth extractors" as mentioned in Subsection 3.2. We also thank anonymous reviewers for their useful feedback.

1 Introduction

The quest to prove unconditional hardness results for restricted models of computation (e.g., low-depth circuits) is one of the major programs of complexity theory. This program saw a number of major successes in the 1980s when researchers exposed the limitations of AC^0 and $\mathsf{AC}^0[p]$ circuits with simple and explicit hard-to-compute functions [19, 1, 50, 24, 25, 41, 37].

Unfortunately, the current state of affairs on this front remains rather humble: even four decades later we still cannot rule out the preposterous claim that all of PSPACE can be computed by polynomially-sized AC^0 circuits with mod 6 gates.

It would be disingenuous, however, to suggest that no modern progress has occurred. Even just the past few years have seen impressive developments in understanding the limitations of these models through the lens of pseudorandom generators [16, 17, 27, 32, 33] (see also the survey [26]), extractors [13, 20, 3], and quantum analogs [36, 5, 28, 18, 23].[1]

One particularly exciting line of inquiry is understanding the capabilities of these weak models to generate specific distributions (when taking random bits as input). The landscape here is dramatically different from the traditional task of computing specific functions. For example, if a circuit can compute a function $f\colon \{0,1\}^n \to \{0,1\}$, then it can output the uniform distribution over $(x, f(x))$ by simply computing f on each random input. The converse, however, is not true. While PARITY has long been known to be difficult for AC^0 circuits to compute [19, 1, 50, 24, 25, 41, 37], one can produce the uniform distribution over input-output pairs by simply mapping the uniformly random bits $(x_1, x_2, \ldots, x_{n+1})$ to $(x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_n \oplus x_{n+1}, x_{n+1} \oplus x_1)$ [6, 11].

Even still, for many common models including small AC^0 circuits [35, 8], small-space sources [14], and communication protocols [4, 21, 14, 52], researchers have discovered explicit distributions which have total variation distance $1 - o(1)$ from any distribution produced by the model. In comparison to our knowledge of computation, the glaring shortcoming in our understanding of sampling is that we do not know of any explicit distribution which cannot be (approximately) produced by polynomially-sized AC^0 circuits with mod p gates for any prime p.

Beyond its inherent intrigue, the complexity of distributions has intimate connections with data structure lower bounds [44, 35, 8, 48, 14, 49, 52, 29, 2], quantum supremacy [10, 29, 22], explicit codes [40], and learning theory [30]. Moreover, the field has had a fruitful relationship with the study of *extractors* [42, 45, 15, 46, 7, 39], which are objects that convert particular sources of randomness into approximately uniform ones.

This work further develops the latter connection by introducing a general paradigm based on *robust extractors* to prove sampling lower bounds. Intuitively, one can view robust extractors as extractors which remain sound even when a small number of points violate the min-entropy condition. There are several ways to formalize this notion, and we defer our exact definition and discussion of alternatives to Subsection 3.2. For now, we will say a robust extractor $\mathsf{rExt}\colon \{0,1\}^n \to \{0,1\}^m$ for a class of distributions $\mathcal{X}$ satisfies the following two conditions:

1. (Extractor Property) For every source $\mathbf{X} \in \mathcal{X}$ with sufficient min-entropy, the output of rExt on $\mathbf{X}$ is close in total variation (TV) distance to the uniform distribution over $\{0,1\}^m$, and
2. (Robustness Property) For every source $\mathbf{X} \in \mathcal{X}$, the probability that $\mathbf{x} \sim \mathbf{X}$ lands in the set of $\mathbf{X}$'s low probability points and $\mathsf{rExt}(\mathbf{x}) = 0^m$ is not much more than 2^{-m}.

With these objects in hand, we can generically prove strong sampling lower bounds. This result further advances a similar approach based on traditional extractors [46], which we discuss in greater detail in Section 2. Below, $\mathbf{U}^n$ denotes the uniform distribution over $\{0,1\}^n$, and for a function $f\colon \{0,1\}^n \to \{0,1\}^m$, $f(\mathbf{U}^n)$ denotes the output distribution of f with n uniformly random bits as input.

[1] This is only a minute fraction of contemporary work, and we encourage the interested reader to consult references within those cited.

▶ **Theorem 1** (Informal version of Theorem 8). *Let $\mathcal{X}$ be a class of distributions over $\{0,1\}^{t(n+1)}$, and let $\mathcal{Y}$ be the class of distributions over $\{0,1\}^n$ obtained from "low complexity" functions of $\mathcal{X}$. Suppose $\mathsf{rExt}\colon \{0,1\}^n \to \{0,1\}^m$ is a sufficiently good robust extractor for $\mathcal{Y}$, and define the function $f\colon \{0,1\}^n \to \{0,1\}$ by $f(x) = \mathbf{1}(\mathsf{rExt}(x) = 0^m)$. If we consider the distribution*

$$\mathbf{D} = (\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_t, f(\mathbf{U}_1), f(\mathbf{U}_2), \dots, f(\mathbf{U}_t)),$$

where $\mathbf{U}_1, \dots, \mathbf{U}_t$ are independent copies of $\mathbf{U}^n$, then every source $\mathbf{X} \in \mathcal{X}$ has TV distance

$$\|\mathbf{X} - \mathbf{D}\|_{\mathsf{TV}} \geq 1 - o_{t,n}(1),$$

where $o_{t,n}(1) \to 0$ as $t, n \to \infty$.

Moreover, if $\widetilde{\mathsf{rExt}}\colon \{0,1\}^n \to \{0,1\}$ is a sufficiently good robust extractor (in a different parameter regime) for $\mathcal{Y}$, then $\left\|\mathbf{X} - (\mathbf{U}^n, \widetilde{\mathsf{rExt}}(\mathbf{U}^n))\right\|_{\mathsf{TV}} \geq \frac{1}{4} - o_n(1)$.

▶ Remark 2. Robust extractors are the motivating object behind our approach, but actually a weaker one-sided notion in place of the typical extractor guarantee (1) suffices for our purposes. In particular, Theorem 1 holds as long as rExt on the uniform distribution assigns decent probability to 0^m. The details can be found in Section 2.

We highlight that the "moreover" part of Theorem 1 can be viewed as a stronger type of lower bound than those for approximate computation. If, for example, the functions $f, g\colon \{0,1\}^n \to \{0,1\}$ agree on a $(1-\delta)$-fraction of their inputs, then it is easy to generate the uniform distribution over $(\mathbf{U}^n, g(\mathbf{U}^n))$ to TV distance at most δ by simply computing f on each random input.

Using Theorem 1, we can recover unconditional hardness results for sampling in essentially every model where they are currently known (albeit with weaker decay rates). We are also able to provide the first explicit[2] distribution which has distance $1 - o(1)$ from the output of any low-degree $\mathbb{F}_2$-polynomial source, making progress toward a similar result for $\mathsf{AC}^0[\oplus]$ circuits.[3] Previous arguments [46, 13] only forbade distance $2^{-\Omega(n)}$.

▶ **Theorem 3** (Informal instantiations of Theorem 1). *Let $\mathcal{X}$ be one of the following classes of sources over $\{0,1\}^N$:*

- *low-degree $\mathbb{F}_2$-polynomial,*
- *local (NC^0),*
- *circuit (AC^0) ,*
- *communication,*
- *small-space,*
- *Turing machine.*

Then every source $\mathbf{X} \in \mathcal{X}$ satisfies

$$\|\mathbf{X} - \mathbf{D}\|_{\mathsf{TV}} \geq 1 - o(1),$$

where $\mathbf{D}$ is defined by taking many independent copies of $(\mathbf{U}^n, \mathbf{1}(\mathsf{rExt}(\mathbf{U}^n) = 0^m))$ for an explicit robust extractor $\mathsf{rExt}\colon \{0,1\}^n \to \{0,1\}^m$ (depending on the choice of $\mathcal{X}$).

[2] One can essentially use a standard counting argument to show such a distribution exists; see the appendix in the full version.

[3] Recall these are AC^0 circuits with mod 2 gates.

Moreover, there exists an explicit robust extractor $\widetilde{\mathsf{rExt}}\colon \{0,1\}^{N-1} \to \{0,1\}$ *(in a different parameter regime) such that every source* $\mathbf{X} \in \mathcal{X}$ *satisfies*

$$\left\|\mathbf{X} - (\mathbf{U}^{N-1}, \widetilde{\mathsf{rExt}}(\mathbf{U}^{N-1}))\right\|_{\mathsf{TV}} \geq \frac{1}{4} - o(1).$$

We emphasize that bounds of this form were already known (with better quantitative behavior) for local and circuit [35, 8], communication [4, 21, 14, 52], small-space [14], and Turing-machine sources [45, 14]. Our novel contribution is obtaining these bounds in a *unified* way, as well as providing strong bounds for low-degree $\mathbb{F}_2$-polynomial sources.

Concurrent Work

In independent and concurrent work, Khodabandeh and Shinkar also investigate the complexity of sampling with low-degree $\mathbb{F}_2$-polynomials [31]. Using very different techniques from our own, they prove that the output of any such polynomial sampler has TV distance $1 - o(1)$ from the $(1/3)$-biased product distribution. As in the present work, the $o(1)$ term has suboptimal decay, but in the case of degrees 1, 2, and 3, they prove it can be taken to be $\exp(-\Omega(N))$, $\exp(-\Omega(\log(N)/\log\log(N)))$, and $\exp(-\Omega(\sqrt{\log\log(N)}))$, respectively, which is stronger than we are able to obtain via our robust extractor framework. (Note, however, that an extractor-based approach can match their bound for degree-1 polynomials – see Section 5.)

Paper Organization

We state and prove Theorem 8, the precise version of Theorem 1, in Section 2, along with additional background and context. We then review some preliminary material in Section 3 before instantiating Theorem 8 in Section 4 to obtain Theorem 3 for polynomial sources. We address the other classes (local, circuit, communication, small-space, and Turing-machine sources) in the full version. We conclude with some open problems in Section 5. Supplementary material on non-explicit constructions can be found in the appendix of the full version.

2 Background and Main Result

In this section, we state and prove our main result, Theorem 8. Before doing so, it will be instructive to review existing arguments and their limitations to develop intuition for our approach. For concreteness, we will focus our discussion on distributions generated by low-degree $\mathbb{F}_2$-polynomials, but essentially all of the analysis holds more generally. The meaning of any unfamiliar terminology or notation in this section can be found in Section 3.

Let $P\colon \mathbb{F}_2^r \to \mathbb{F}_2^n$ be a polynomial map defined by n arbitrary degree-d polynomials $\{p_i\colon \mathbb{F}_2^r \to \mathbb{F}_2\}_{i=1}^n$ acting on the same r input bits, where we view r as some arbitrarily large integer. Our goal is to construct an explicit distribution $\mathbf{D}$ over $\{0,1\}^n$ such that regardless of how the p_i's are defined, the output distribution of P on r uniformly random bits, denoted $P(\mathbf{U}^r)$, has total variation distance at least $1 - o(1)$ from $\mathbf{D}$. (Here, we identify $\mathbb{F}_2^n$ with $\{0,1\}^n$.)

2.1 A Nonzero Lower Bound

We begin with the more modest goal of showing that P cannot *exactly* generate some explicit distribution $\mathbf{D}$. Here, we may invoke an argument of Viola [46, 47] based on extractors, whose formal definition we now recall.

▶ **Definition 4** (Extractor). *A function* $\mathsf{Ext}\colon \{0,1\}^n \to \{0,1\}$ *is an* (ε, k)-extractor *for a class* $\mathcal{X}$ *of distributions over* $\{0,1\}^n$ *if for every source* $\mathbf{X} \in \mathcal{X}$ *with min-entropy at least* k*, we have*

$$\mathbf{Pr}[\mathsf{Ext}(\mathbf{X}) = 1] = \frac{1}{2} \pm \varepsilon.$$

We will take the hard distribution to be uniform over input-output pairs of an explicit (ε, k)-extractor $\mathsf{Ext}\colon \{0,1\}^{n-1} \to \{0,1\}$ for polynomial sources of degree $2d$, where ε is some small constant and $k = n - O(1)$. That is, $\mathbf{D} = (\mathbf{U}^{n-1}, \mathsf{Ext}(\mathbf{U}^{n-1}))$. Such extractors which are computable in time $\mathsf{poly}(n)$ are known [13]. For clarity, we express $P(\mathbf{U}^r)$ similarly as $(Q(\mathbf{U}^r), q(\mathbf{U}^r))$, where $Q\colon \{0,1\}^r \to \{0,1\}^{n-1}$ and $q\colon \{0,1\}^r \to \{0,1\}$ are degree-d polynomials acting on the same set of input bits, corresponding to the first $n-1$ output bits and last output bit of P, respectively.

Suppose for contradiction that $P(\mathbf{U}^r)$ exactly generates $\mathbf{D}$. Viola's argument considers the random variable $\mathbf{M}$ over $\{0,1\}^{n-1}$ defined by sampling $\mathbf{u} \sim \mathbf{U}^r$ and outputting $Q(\mathbf{u})$ if $q(\mathbf{u}) = 1$, and otherwise $n-1$ uniformly random bits independent of $\mathbf{U}^r$, which we denote by $\widetilde{\mathbf{U}}^{n-1}$. Written suggestively similar to a polynomial, we have

$$\mathbf{M} = q(\mathbf{U}^r)Q(\mathbf{U}^r) + (1 - q(\mathbf{U}^r))\widetilde{\mathbf{U}}^{n-1}, \tag{1}$$

or equivalently by our assumption,

$$\mathbf{M} = \mathsf{Ext}(\mathbf{U}^{n-1})\mathbf{U}^{n-1} + (1 - \mathsf{Ext}(\mathbf{U}^{n-1}))\widetilde{\mathbf{U}}^{n-1}. \tag{2}$$

Now consider the effect of applying Ext to $\mathbf{M}$. By (1), $\mathbf{M}$ can be generated by a degree-$(2d)$ $\mathbb{F}_2$-polynomial source with $n - O(1)$ bits of min-entropy (recall $Q(\mathbf{U}^r)$ is assumed to be uniform), so $\mathbf{Pr}[\mathsf{Ext}(\mathbf{M}) = 1] \le \frac{1}{2} + \varepsilon$. By contrast, whenever the sample $\mathbf{u} \sim \mathbf{U}^{n-1}$ satisfies $\mathsf{Ext}(\mathbf{u}) = 1$, we must have $\mathsf{Ext}(\mathbf{M}) = 1$ by (2), so

$$\begin{aligned}
\mathbf{Pr}[\mathsf{Ext}(\mathbf{M}) = 1] &= \mathbf{Pr}\left[\mathsf{Ext}(\mathbf{M}) = 1 \,\middle|\, \mathsf{Ext}(\mathbf{U}^{n-1}) = 1\right] \cdot \mathbf{Pr}[\mathsf{Ext}(\mathbf{U}^{n-1}) = 1] \\
&\quad + \mathbf{Pr}\left[\mathsf{Ext}(\mathbf{M}) = 1 \,\middle|\, \mathsf{Ext}(\mathbf{U}^{n-1}) = 0\right] \cdot \mathbf{Pr}[\mathsf{Ext}(\mathbf{U}^{n-1}) = 0] \\
&= 1 \cdot \mathbf{Pr}[\mathsf{Ext}(\mathbf{U}^{n-1}) = 1] + \mathbf{Pr}[\mathsf{Ext}(\widetilde{\mathbf{U}}^{n-1}) = 1] \cdot \mathbf{Pr}[\mathsf{Ext}(\mathbf{U}^{n-1}) = 0] \\
&\approx \frac{1}{2} + \frac{1}{4} \gg \frac{1}{2} + \varepsilon,
\end{aligned}$$

a contradiction. That is, no low-degree polynomial can exactly sample $\mathbf{D}$. In fact, a more careful analysis forbids P from sampling $\mathbf{D}$ to distance better than $2^{-\Omega(k)}$ (which is $2^{-\Omega(n)}$ for known explicit constructions of such extractors [13]).

2.2 A Constant Lower Bound

It is initially unclear how to strengthen the previous argument to obtain even some constant distance lower bound. The primary issue is that even if $\mathbf{D}$ has large min-entropy, it might be reasonably close to some distribution $P(\mathbf{U}^n)$ with small min-entropy, so we do not have any guarantees on the output of the extractor on $\mathbf{M}$.

2.2.1 Min-Entropy Polarization

One approach that was successful in the analysis of distributions generated by low-depth circuits is *min-entropy polarization* [48]. Using random restrictions and hypercontractivity, Viola proved that any distribution $C(\mathbf{U}^r)$ produced by a small circuit $C\colon \{0,1\}^r \to \{0,1\}^n$

could be approximated by a not-too-large collection of restrictions of C whose output was either constant or had large min-entropy, in which case an extractor Ext could be meaningfully applied (see Section 4.2 in the full version for additional details). The event witnessing the total variation distance can then be defined as the union of a small set of outputs, corresponding to restrictions under which the circuit is constant, and pairs (x, b) where $\mathsf{Ext}(x) \neq b$.

Unfortunately, one cannot always polarize the output distribution of an arbitrary sampler in this fashion; consider the following example from [48].

▶ **Example 5.** Let $\mathbf{G}$ be the distribution over $\{0,1\}^n$ obtained by sampling a random string $\mathbf{u} \sim \mathbf{U}^n$, and outputting $\mathbf{u}$ if $|\mathbf{u}| \bmod 2 = 1$ and otherwise outputting 0^n. This sampler cannot have its min-entropy polarized via restrictions for the following reason. For any restriction which leaves some variable free, the min-entropy remains 1. This means that every restriction in a polarizing collection must fix every variable, but then $\Omega(2^n)$ many such restrictions are needed to approximate the distribution.

Note, however, that this sampling procedure *can* be implemented by the degree-2 $\mathbb{F}_2$-polynomial map

$$P(y) = \left(\sum_{i=1}^{n} y_i\right) \cdot (y_1, \ldots, y_n) + \left(1 - \sum_{i=1}^{n} y_i\right) \cdot 0^n$$

to express $\mathbf{G}$ as $P(\mathbf{U}^n)$. Hence, we are not able to apply this framework for our purposes.

2.2.2 Discarding Heavy Points

An alternative approach, still somewhat in the spirit of [48], was taken by Chattopadhyay, Goodman, and Zuckerman [14] when considering communication sources (see Section 4.3.1 in the full version for a formal definition). They also defined their witness event to be the union of a small "bad" set and input-output pairs which contradict a particular extractor's definition. Unlike Viola, however, they define their bad set to be strings which are assigned substantial probability by the sampler (excluding the last bit). Since there cannot be many "heavy" strings, this set is indeed small.

The downside to this approach is that the extractor one requires does not necessarily correspond to an extractor for the class of sources being considered. In particular, the argument requires an extractor that works on the source after conditioning on membership in some set. For communication sources, one can reduce to the case of being able to apply a two-source extractor, but it is unclear how to generically perform a similar reduction to known extractors in the case of other sources, such as those generated by low-degree polynomials.

2.2.3 Robust Extractors

To overcome the previous two obstacles, we ask for more from the *extractor* rather than from the *source*. Recall that the issue with trying to strengthen the argument from Subsection 2.1 is that the generated distribution $P(\mathbf{U}^r) = (Q(\mathbf{U}^r), q(\mathbf{U}^r))$ might assign too much mass to a small set of "bad" points, so the auxiliary distribution of $\mathbf{M}$ does not have enough min-entropy to apply an extractor. Our solution is to simply ask for an extractor whose soundness holds on the light (i.e., low probability) points, rather than the entire domain. In other words, we seek an extractor which is robust to a small number of points violating the min-entropy constraint.

To be a bit more precise, suppose that the (ε, k)-extractor $\mathsf{Ext}\colon \{0,1\}^{n-1} \to \{0,1\}$ we have been considering had the property that every degree-$(2d)$ $\mathbb{F}_2$-polynomial source $\mathbf{X}$ satisfies

$$\mathbf{Pr}[\mathbf{X} \in L \text{ and } \mathsf{Ext}(\mathbf{X}) = 1] \leq \frac{1}{2} + \varepsilon, \tag{3}$$

where $L = \{x \in \{0,1\}^{n-1} : \mathbf{Pr}[\mathbf{X} = x] \leq 2^{-k}\}$ is the set of light points.[4] (Recall that a "traditional" extractor is only guaranteed to satisfy (3) when $L = \{0,1\}^{n-1}$.) A priori, it is not obvious that these objects even exist, but it turns out that a straightforward modification to the extractor construction of [13] yields such an object.

In general, we believe that essentially any natural class of distributions should have extractors with a similar guarantee, and much of the present work is devoted to showing that existing extractor constructions (or simple modifications of them) are robust. In fact, one can interpret the aforementioned min-entropy polarization results in [48] as saying that any extractor for circuit sources enjoys the robustness property (see Section 4.2.1 in the full version).

Now that we have a *robust extractor*, we can run a version of our earlier argument. Once again, let $\mathbf{M}$ be the random variable over $\{0,1\}^{n-1}$ defined by sampling $\mathbf{u} \sim \mathbf{U}^r$ and outputting $Q(\mathbf{u})$ if $q(\mathbf{u}) = 1$, and otherwise $n-1$ random bits using fresh randomness. We also define $\widetilde{\mathbf{M}}$ similarly, but with $\mathbf{D} = (\mathbf{U}^{n-1}, \mathsf{Ext}(\mathbf{U}^{n-1}))$ in place of $P(\mathbf{U}^r)$. By (3), we morally have that $\mathsf{Ext}(\mathbf{M})$ is near balanced on its light points, whereas by construction, $\mathsf{Ext}(\widetilde{\mathbf{M}})$ is likely to output 1 on such points. Working out the details (see Subsection 2.4), one finds that $P(\mathbf{U}^r)$ must be $\Omega(1)$-far from $\mathbf{D}$, as desired.

2.3 A $1 - o(1)$ Lower Bound

It remains to amplify the total variation distance lower bound from constant to $1 - o(1)$. Toward this end, we will take many independent copies of the distribution above, and prove that generating this product of distributions is much harder than generating any individual distribution. We note that such *direct product theorems* have been examined previously in the sampling context [14, 22].

Proceeding more formally, we redefine the distribution $\mathbf{D}$ to be

$$\mathbf{D} = (\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_t, \mathsf{Ext}(\mathbf{U}_1), \mathsf{Ext}(\mathbf{U}_2), \ldots, \mathsf{Ext}(\mathbf{U}_t)),$$

where the $\mathbf{U}_i$'s are independent copies of $\mathbf{U}^{n-1}$ and $t \geq 1$ is an integer to be chosen later. Observe that now $\mathbf{D}$ is over $\{0,1\}^N$ for $N \coloneqq tn$, so we must redefine our polynomial map P to also be on N output bits. We emulate much of our previous analysis with the auxiliary random variable $\widetilde{\mathbf{M}}$ generalized to output $\mathbf{u}_i$ if $\mathsf{Ext}(\mathbf{u}_i) = 1$ for any $i \in \{1, \ldots, t\}$ (and $\mathbf{M}$ similarly generalized with respect to $P(\mathbf{U}^r)$). This greatly increases the probability that $\mathsf{Ext}(\widetilde{\mathbf{M}}) = 1$, which further improves our distance bound.

Unfortunately, the approach as written does not provide distance approaching 1. The problem can essentially be traced back to a loss of $\mathbf{Pr}[\mathbf{X} \in L \text{ and } \mathsf{Ext}(\mathbf{X}) = 1]$ coming from our test event. Since the set of low probability points L might be the entire domain, we cannot upper bound this quantity by anything appreciably better than $1/2$ without violating the extractor guarantee. In order to make further progress, we extend our extractor to have multiple output bits.

[4] There are other ways to capture a similar notion of robustness, but this definition seems to be particularly advantageous; see the discussion in Subsection 3.2.

This multi-output extractor $\mathsf{Ext}\colon \{0,1\}^{n-1} \to \{0,1\}^m$ is defined similarly to the single-output version (Definition 4), only now with the pseudorandomness condition being a bound on the total variation distance $\|\mathsf{Ext}(\mathbf{X}) - \mathbf{U}^m\|_{\mathsf{TV}} \leq \varepsilon$. One can again obtain a robust version (with $m = \Omega(\log \log n)$) by modifying the construction of [13]; we defer the formal definition to Subsection 3.2.

The upshot is that now we can replace the bottleneck of the previous argument with an analysis of $\mathbf{Pr}[\mathbf{X} \in L \text{ and } \mathsf{Ext}(\mathbf{X}) = 0^m]$. This, in turn, can be more tightly bounded by $2^{-m} + \varepsilon$. In our construction, ε will tend to 0 as m grows,[5] so working out the calculations, we obtain a bound roughly of the form

$$1 - \exp(-2^{-m} \cdot t) - 2^{-\Omega(n)} \cdot \mathsf{poly}(t)$$

for sufficiently large n. Setting t to be slightly larger than 2^m, we can finally obtain our desired $1 - o(1)$ bound. The full details can be found in the subsequent subsection.

2.3.1 Isolators

Before formalizing our discussion into a theorem, it is worth reflecting on exactly what properties of the robust extractor we used. (These properties may be further illuminated by consulting the proof of Theorem 8.) In analyzing our target distribution $\mathbf{D}$, all that we truly needed is for some output $y \in \{0,1\}^m$ to be assigned substantial mass by $\mathsf{Ext}(\mathbf{U}^{n-1})$. By contrast, our analysis of the generated distribution $P(\mathbf{U}^r)$ only required that the extractor did not map too many of the light points to the same output y. In other words, our argument does not require the full power of robust extractors. We distill the properties we need into the following weaker object, which we call an *isolator*, since it in some sense controls the isolated light points.

▶ **Definition 6** (Isolator). *A function* $\mathsf{Iso}\colon \{0,1\}^n \to \{0,1\}$ *is an* (α, β, k)-isolator *for a class* $\mathcal{X}$ *of distributions over* $\{0,1\}^n$ *if*

1. $\mathbf{Pr}[\mathsf{Iso}(\mathbf{U}^n) = 1] \geq \alpha$, *and*
2. *Every source* $\mathbf{X} \in \mathcal{X}$ *satisfies* $\mathbf{Pr}_{x \sim \mathbf{X}}[x \in L \text{ and } \mathsf{Iso}(x) = 1] \leq \beta$, *where* $L = \{x \in \{0,1\}^n : \mathbf{Pr}[\mathbf{X} = x] \leq 2^{-k}\}$.

It is perhaps worth highlighting that although all our constructions of isolators essentially work by "robustifying" existing extractor constructions, this is not strictly necessary to satisfy Definition 6. It is plausible one could explicitly construct an isolator for $\mathsf{AC}^0[\oplus]$ sources, for example, without first explicitly constructing seemingly elusive extractors for that class.

2.4 Main Result

We conclude Section 2 by stating our main result, Theorem 8, and its proof. Below, the following notation will be convenient for generalizing the random variable $\mathbf{M}$ considered above.

[5] We had previously defined ε to be a small constant, but our construction of robust extractors allows for this smaller setting.

▶ **Definition 7** (Randomized Addressing). *For positive integers t and n, we define the randomized function* $\mathsf{addr}\colon (\{0,1\}^n)^t \times \{0,1\}^t \to \{0,1\}^n$ *as*

$$\mathsf{addr}(A_1, A_2, \ldots, A_t, b_1, b_2, \ldots, b_t) = \begin{cases} A_1 & \text{if } b_1 = 1, \\ A_2 & \text{if } b_1 = 0, b_2 = 1, \\ A_3 & \text{if } b_1 = b_2 = 0, b_3 = 1, \\ \quad\vdots & \\ A_t & \text{if } b_1 = b_2 = \cdots = b_{t-1} = 0, b_t = 1, \\ \mathbf{U}^n & \text{if } b_1 = b_2 = \cdots = b_t = 0. \end{cases}$$

Though the parameters n and t will mostly be clear from context, we will sometimes write $\mathsf{addr}_{n,t}$ for clarity.

We can now finally present our main result.

▶ **Theorem 8.** *Let $\alpha, \beta \in [0,1]$ and t, k, n be positive integers where $k \le n-1$. Let $\mathcal{X}$ be a class of distributions over $\{0,1\}^{t(n+1)}$, and let $\mathcal{Y}$ be the class of distributions over $\{0,1\}^n$ of the form $\mathsf{addr}_{n,t}(\mathbf{X})$ for some $\mathbf{X} \in \mathcal{X}$. Suppose $\mathsf{Iso}\colon \{0,1\}^n \to \{0,1\}$ is an (α, β, k)-isolator for $\mathcal{Y}$, and define the distribution*

$$\mathbf{D} = (\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_t, \mathsf{Iso}(\mathbf{U}_1), \mathsf{Iso}(\mathbf{U}_2), \ldots, \mathsf{Iso}(\mathbf{U}_t)),$$

where $\mathbf{U}_1, \ldots, \mathbf{U}_t$ are independent copies of $\mathbf{U}^n$. Then every source $\mathbf{X} \in \mathcal{X}$ satisfies

$$\|\mathbf{X} - \mathbf{D}\|_{\mathsf{TV}} \ge 1 - (1-\alpha)^{t+1} - 2^{-(n-k)}\left(2t^3 + 1\right) - \beta.$$

In particular, $\|\mathbf{X} - (\mathbf{U}^n, \mathsf{Iso}(\mathbf{U}^n))\|_{\mathsf{TV}} \ge 2\alpha - \alpha^2 - 2^{-(n-k-2)} - \beta$.

▶ Remark 9. Depending on the sampling model and particular goal, the precise ordering of the coordinates in a target distribution may affect whether or not that distribution can be sampled (e.g., [45]). In the case of Theorem 8, the proof goes through for any permutation of the coordinates of addr and $\mathbf{D}$, although the order may affect whether $\mathsf{addr}(\mathbf{X})$ lies in a particular class.

Proof of Theorem 8. Let $\mathbf{X}$ be any source in $\mathcal{X}$. Viewing $\mathbf{X}$ as $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_t, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t)$ with the $\mathbf{X}_i$'s and $\mathbf{x}_j$'s over $\{0,1\}^n$ and $\{0,1\}$, respectively, we may apply $\mathsf{addr}_{n,t}$ to $\mathbf{X}$ to obtain the source $\mathbf{Y} = \mathsf{addr}(\mathbf{X}) \in \mathcal{Y}$. Next, define the sets of light points in $\mathbf{X}_i$ for $i = 1, 2, \ldots, t$ and $\mathbf{Y}$ by

$$L_{\mathbf{X}_i} = \{x \in \{0,1\}^n : \mathbf{Pr}[\mathbf{X}_i = x] \le 2^{-(\log(t)+k+1)}\} \text{ and } L_{\mathbf{Y}} = \{y \in \{0,1\}^n : \mathbf{Pr}[\mathbf{Y} = y] \le 2^{-k}\}.$$

Observe that $\bigcap_i L_{\mathbf{X}_i} \subseteq L_{\mathbf{Y}}$, since any string $z \in \bigcap_i L_{\mathbf{X}_i}$ satisfies

$$\begin{aligned} \mathbf{Pr}[\mathbf{Y} = z] = \Pr_{\mathbf{X}}[\mathsf{addr}(\mathbf{X}) = z] &\le \mathbf{Pr}[\mathbf{U}^n = z] + \sum_{i=1}^{t} \mathbf{Pr}[\mathbf{X}_i = z] \qquad \text{(by union bound)} \\ &\le 2^{-n} + t \cdot 2^{-(\log(t)+k+1)} \le 2^{-k}. \end{aligned}$$

Thus, the isolator property of Iso guarantees that

$$\begin{aligned} \mathcal{E}(\mathbf{X}) &\coloneqq \Pr_{\mathbf{X}}\left[\bigcap_i \left(\mathbf{X}_i \in \bigcap_j L_{\mathbf{X}_j}\right) \text{ and } \mathsf{Iso}(\mathsf{addr}(\mathbf{X})) = 1\right] \\ &\le \Pr_{\mathbf{X}}[\mathsf{addr}(\mathbf{X}) \in L_{\mathbf{Y}} \text{ and } \mathsf{Iso}(\mathsf{addr}(\mathbf{X})) = 1] + \Pr_{\mathbf{X}}\left[\mathsf{addr}(\mathbf{X}) \notin L_{\mathbf{Y}} \,\middle|\, \bigcap_i \left(\mathbf{X}_i \in \bigcap_j L_{\mathbf{X}_j}\right)\right] \\ &\le \beta + \mathbf{Pr}[\mathbf{U}^n \notin L_{\mathbf{Y}}] \le \beta + 2^{-(n-k)}, \end{aligned} \tag{4}$$

where the final inequality follows from the fact that $\mathbf{Y}$ can assign 2^{-k} probability mass to at most 2^k points.

We will choose $\mathcal{E}$ to be the test witnessing the claimed TVD between $\mathbf{X}$ and $\mathbf{D}$, although it will be slightly more convenient in this part of the analysis to consider the complement event. In a similar fashion to $\mathbf{X}$, we view $\mathbf{D}$ as $(\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_t, \mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_t)$, where recall each $\mathbf{D}_i$ is an independent copy of the uniform distribution $\mathbf{U}^n$. By the union bound, we have

$$\begin{aligned}
&\Pr_{\mathbf{D}}\left[\bigcup_i \left(\mathbf{D}_i \notin \bigcap_j L_{\mathbf{X}_j}\right) \text{ or } \mathsf{Iso}(\mathsf{addr}(\mathbf{D})) = 0\right] \\
&\leq \sum_i \Pr_{\mathbf{D}}\left[\mathbf{D}_i \notin \bigcap_j L_{\mathbf{X}_j}\right] + \Pr_{\mathbf{D}}[\mathsf{Iso}(\mathsf{addr}(\mathbf{D})) = 0] \\
&\leq \sum_{i,j} \Pr_{\mathbf{D}}\left[\mathbf{D}_i \notin L_{\mathbf{X}_j}\right] + \Pr_{\mathbf{D}}[\mathsf{Iso}(\mathsf{addr}(\mathbf{D})) = 0]. \qquad (5)
\end{aligned}$$

Note that the first term is at most $t^2 \cdot 2^{-(n-(\log(t)+k+1))}$, since no $\mathbf{X}_j$ can assign $2^{-(\log(t)+k+1)}$ probability mass to more than $2^{\log(t)+k+1}$ points. To address the second term, let us consider the effect of applying addr to $\mathbf{D}$. Whenever any $\mathbf{D}_i$ satisfies $\mathsf{Iso}(\mathbf{D}_i) = 1$, we have $\mathsf{Iso}(\mathsf{addr}(\mathbf{D})) = \mathsf{Iso}(\mathbf{D}_i) = 1$. Otherwise, $\mathsf{Iso}(\mathbf{D}_i) = 0$ for all i, and we have $\mathsf{Iso}(\mathsf{addr}(\mathbf{D})) = \mathsf{Iso}(\mathbf{U}^n)$. Thus,

$$\begin{aligned}
\mathbf{Pr}[\mathsf{Iso}(\mathsf{addr}(\mathbf{D})) = 0] &= \Pr_{\mathbf{D}}\left[\mathsf{Iso}(\mathbf{U}^n) = 0 \text{ and } \bigcap_i (\mathsf{Iso}(\mathbf{D}_i) = 0)\right] \\
&= \mathbf{Pr}[\mathsf{Iso}(\mathbf{U}^n) = 0]^{t+1} \leq (1-\alpha)^{t+1}.
\end{aligned}$$

Plugging these two bounds back into (5) yields

$$\Pr_{\mathbf{D}}\left[\bigcap_i \left(\mathbf{D}_i \in \bigcap_j L_{\mathbf{X}_j}\right) \text{ and } \mathsf{Iso}(\mathsf{addr}(\mathbf{D})) = 1\right] \geq 1 - (1-\alpha)^{t+1} - t^2 \cdot 2^{-(n-(\log(t)+k+1))}.$$

Recalling the upper bound from (4), we conclude

$$\|\mathbf{X} - \mathbf{D}\|_{\mathsf{TV}} \geq 1 - (1-\alpha)^{t+1} - 2^{-(n-k)}\left(2t^3+1\right) - \beta. \qquad \blacktriangleleft$$

3 Preliminaries

We now briefly review some notation and formal definitions used throughout the work.

3.1 The Basics

For a positive integer n, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. For a binary string x, we use $|x|$ to denote the Hamming weight of x. All logarithms given in the paper are base 2. For two real numbers a and b, we write $a \pm b$ to denote a value in the range $[a-b, a+b]$. The indicator function is denoted by $\mathbf{1}(\cdot)$. The notation $\binom{n}{\leq k}$ is shorthand for $\sum_{i=0}^{k}\binom{n}{i}$. We use $\mathbb{F}_2$ to denote the finite field of two elements; we often identify it with $\{0,1\}$. The concatenation of two strings x and y is denoted $x \circ y$.

3.1.1 Asymptotics

We use the standard $\Omega(\cdot), O(\cdot), \Theta(\cdot)$ asymptotic notation to hide universal positive constants, although we will sometimes use $\gg$ and $\ll$ in more informal contexts. Occasionally, we will use subscripts to indicate an unspecified dependence on a particular parameter (e.g., $\Omega_d(n)$). Additionally, we write $o_t(1)$ to denote a positive quantity tending to 0 as t tends to infinity; we often omit the subscript when t is clear from context. The shorthand $\mathsf{poly}(n)$ corresponds to a polynomial in n of some fixed, but unspecified, degree.

3.1.2 Probability

We endeavor to use bold capital letters to denote probability distributions, and use bold lowercase letters to denote randomly drawn samples. That is, $\mathbf{x} \sim \mathbf{X}$ denotes a sample $\mathbf{x}$ drawn from the distribution $\mathbf{X}$. Oftentimes, we will refer to a distribution as a *source* if it is being fed into an extractor-like object. We reserve $\mathbf{U}$ for the uniform distribution over $\{0,1\}$. We use calligraphic letters, such as $\mathcal{X}$, for classes of distributions. For an event $\mathcal{E}$, we define $\mathbf{X}(\mathcal{E})$ to be the probability mass assigned to $\mathcal{E}$ by $\mathbf{X}$. For a function f, we use $f(\mathbf{X})$ to denote the output distribution of $f(\mathbf{x})$ on randomly drawn $\mathbf{x} \sim \mathbf{X}$. The *min-entropy* of a distribution $\mathbf{X}$, denoted $H_\infty(\mathbf{X})$, is given by $-\log \max_{x \in \mathsf{supp}(\mathbf{X})} \mathbf{Pr}[\mathbf{X} = x]$, where the support $\mathsf{supp}\,(\mathbf{X}) = \{x : \mathbf{Pr}[\mathbf{X} = x] > 0\}$.

Given a distribution $\mathbf{X}$ and positive integer t, we use $\mathbf{X}^t$ to denote the t-fold product distribution $\mathbf{X} \times \cdots \times \mathbf{X}$. If s is a finite set, we write $\mathbf{X}^s$ to emphasize that the coordinates of $\mathbf{X}^{|s|}$ are indexed by s. We refer to $\mathbf{X}$ as a mixture if it can be written as a convex combination of other distributions. That is, there exists $c_1, \ldots, c_k \in [0,1]$ and distributions $\mathbf{X}_1, \ldots, \mathbf{X}_k$ such that $\mathbf{X}(\mathcal{E}) = \sum_{i=1}^k c_i \cdot \mathbf{X}_i(\mathcal{E})$ for every event $\mathcal{E}$. Occasionally, we write this more concisely as $\mathbf{X} = \sum_{i=1}^k c_i \mathbf{X}_i$.

We measure the similarity of two (discrete) distributions $\mathbf{P}$ and $\mathbf{Q}$ by the *total variation (TV) distance*

$$\|\mathbf{P} - \mathbf{Q}\|_{\mathsf{TV}} = \max_{\text{event } \mathcal{E}} \mathbf{P}(\mathcal{E}) - \mathbf{Q}(\mathcal{E}) = \frac{1}{2} \sum_x |\mathbf{P}(x) - \mathbf{Q}(x)| \,.$$

We say $\mathbf{P}$ is ε-*close* to $\mathbf{Q}$ if $\|\mathbf{P} - \mathbf{Q}\|_{\mathsf{TV}} \leq \varepsilon$, and ε-*far* otherwise.

3.1.3 Explicit Constructions

The focus of this work is on *explicitly* constructing distributions with certain properties, by which we mean $\mathbf{D}$ can be sampled in $\mathsf{poly}(n)$ time. Note that as in the computational world, it is straightforward to non-constructively prove strong hardness results via a counting argument. More precisely, for any class $\mathcal{X}$ of distributions over $\{0,1\}^n$ of size $2^{2^{cn}}$ for a constant $c < 1$, there exists a uniform distribution $\mathbf{D}$ with distance $1 - 2^{-\Omega(n)}$ from every distribution in $\mathcal{X}$ (see the full version for details).

There is a small subtlety, however, in that the sampling models we consider have unbounded input length, so one cannot naively apply the above claim. Fortunately, the classes of interest can typically be approximated by a subset of the class where the input length is bounded (e.g., [14]), which allows the argument to go through. We are unaware of any work formally describing these counting arguments giving hard distributions for low-degree polynomial sources and $\mathsf{AC}^0[\oplus]$ sources, so we record the details in the appendix of the full version.

3.2 Extractors, Robust Extractors, and Isolators

There are three pseudorandom objects at the core of our work: extractors, robust extractors, and isolators. We state the formal definitions of the first two below.

▶ **Definition 10** ((Robust) Extractor). *A function* $\mathsf{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ *is an* (ε, k)-extractor *for a class* $\mathcal{X}$ *of distributions over* $\{0,1\}^n$ *if for every source* $\mathbf{X} \in \mathcal{X}$ *with min-entropy at least* k*, we have*

$$\|\mathsf{Ext}(\mathbf{X}) - \mathbf{U}^m\|_{\mathsf{TV}} \leq \varepsilon.$$

If additionally, there exists some $z \in \{0,1\}^m$ *such that every source* $\mathbf{X} \in \mathcal{X}$ *satisfies*

$$\mathbf{Pr}[\mathbf{X} \in L \text{ and } \mathsf{Ext}(\mathbf{X}) = z] \leq \frac{1}{2^m} + \delta,$$

where $L = \{x \in \{0,1\}^n : \mathbf{Pr}[\mathbf{X} = x] \leq 2^{-k}\}$*, then we call* Ext *an* (ε, δ, k)-robust extractor.

It is worth mentioning that there are other ways to formalize this notion of robustness. For example, one could ask for an extractor which works for a source $\mathbf{X}$, as long as $\mathbf{X}$ has TV distance no more than some parameter γ to a high min-entropy source $\mathbf{X}'$ (i.e., an extractor for sources of high smooth min-entropy [38]). Assuming such extractors existed for the classes of distributions we consider, it would be possible to obtain a constant distance lower bound as in the second part of Theorem 8.

Unfortunately, it does not seem feasible to get distance approaching 1. Fix some choice of γ and a sampler $f(\mathbf{U}^r)$ for the hard distribution $\mathbf{D}$ considered in Theorem 8. If $f(\mathbf{U}^r)$ is γ-far from $\mathbf{D}$, we trivially have a lower bound of γ, so assume this is not the case. Note that we cannot expect a "smooth extractor" to have a better upper bound probability guarantee than $2^{-m} + \gamma$, since it has to apply to distributions which are a point mass with probability γ and uniform otherwise. Tracing through the remainder of the argument, one finds that $f(\mathbf{U}^r)$ has distance roughly $1 - \varepsilon - \gamma$ from $\mathbf{D}$ for some small ε. In other words, we can only guarantee a bound around $\min(\gamma, 1 - \varepsilon - \gamma) \ll 1$.

One of the benefits of Definition 10 is that we can obtain tighter distance guarantees, because we do not need to consider the behavior of the extractor on heavy points. Hence, we can avoid the above issue. Returning back to our selected formalizations, we restate the definition of an isolator for the reader's convenience.

▶ **Definition 6** (Isolator). *A function* $\mathsf{Iso}\colon \{0,1\}^n \to \{0,1\}$ *is an* (α, β, k)-isolator *for a class* $\mathcal{X}$ *of distributions over* $\{0,1\}^n$ *if*

1. $\mathbf{Pr}[\mathsf{Iso}(\mathbf{U}^n) = 1] \geq \alpha$*, and*
2. *Every source* $\mathbf{X} \in \mathcal{X}$ *satisfies* $\mathbf{Pr}_{x \sim \mathbf{X}}[x \in L \text{ and } \mathsf{Iso}(x) = 1] \leq \beta$*, where* $L = \{x \in \{0,1\}^n : \mathbf{Pr}[\mathbf{X} = x] \leq 2^{-k}\}$.

The role of robust extractors in the present work is primarily as a convenient device for constructing isolators.

▶ **Fact 11.** *Let* $\mathcal{X}$ *be a class of distributions over* $\{0,1\}^n$ *which includes the uniform distribution* $\mathbf{U}^n$*. If* $\mathsf{rExt}\colon \{0,1\}^n \to \{0,1\}^m$ *is an* (ε, δ, k)*-robust extractor for* $\mathcal{X}$*, then the function* $\mathsf{Iso}(x) \coloneqq \mathbf{1}(\mathsf{rExt}(x) = z)$ *is a* $(2^{-m} - \varepsilon, 2^{-m} + \delta, k)$*-isolator for* $\mathcal{X}$ *(where* z *is the same string as in Definition 10).*

Of course, one still has to construct a robust extractor to apply Fact 11. Typically, this is not much more difficult than constructing a traditional extractor, and in the following section, we illustrate such a modification with a number of natural examples. We note that in certain cases, the robustness can even be obtained for free.

▷ Claim 12. Let $\mathcal{X}$ be a class of distributions over $\{0,1\}^n$ such that each $\mathbf{X} \in \mathcal{X}$ is the uniform distribution on some set $S_{\mathbf{X}} \subseteq \{0,1\}^n$ (i.e., $\mathcal{X}$ is a family of *flat* sources). If $\mathsf{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ is an (ε, k)-extractor for $\mathcal{X}$, then Ext is also an $(\varepsilon, \varepsilon, k)$-robust extractor for $\mathcal{X}$.

Proof. Consider any source $\mathbf{X} \in \mathcal{X}$, which is uniform over a set $S_{\mathbf{X}}$. If $|S_{\mathbf{X}}| \geq 2^k$, then $\mathbf{X}$ has min-entropy at least k and the set

$$L \coloneqq \{x \in \{0,1\}^n : \mathbf{Pr}[\mathbf{X} = x] \leq 2^{-k}\} = \{0,1\}^n.$$

Thus, the extractor guarantee implies

$$\mathop{\mathbf{Pr}}_{x\sim\mathbf{X}}[x \in L \text{ and } \mathsf{Ext}(x) = 0^m] = \mathop{\mathbf{Pr}}_{x\sim\mathbf{X}}[\mathsf{Ext}(x) = 0^m] \leq 2^{-m} + \varepsilon.$$

Otherwise $|S_{\mathbf{X}}| < 2^k$, and $\mathbf{Pr}[\mathbf{X} = x] = 0$ for every $x \in L$. Hence,

$$\mathop{\mathbf{Pr}}_{x\sim\mathbf{X}}[x \in L \text{ and } \mathsf{Ext}(x) = 0^m] \leq \mathop{\mathbf{Pr}}_{x\sim\mathbf{X}}[x \in L] = 0. \qquad \triangleleft$$

4 Hard-to-Sample Distributions for Polynomial Sources

In this section, we instantiate Theorem 8 for polynomial sources.

▶ **Definition 13** (Polynomial Source). *A degree-d* polynomial source $\mathbf{X}$ *is defined by a polynomial map $P\colon \mathbb{F}_2^m \to \mathbb{F}_2^n$, where $P = (p_1, p_2, \ldots, p_n)$ and each p_i is an $\mathbb{F}_2$-polynomial of degree at most d, such that $\mathbf{X} = P(\mathbf{U}^m)$.*

Prior to our work, the best explicit distribution was only known to have distance $2^{-\Omega(n)}$ from any low-degree polynomial map; this is achieved by combining an argument of [46] (described in Subsection 2.1) with an extractor construction of [13]. We improve this distance bound to $1 - o(1)$.

▶ **Theorem 14.** *There exists a constant $\delta > 0$ such that the following holds. Let N and Δ be positive integers satisfying $\Delta \leq \delta \log\log N$. There exist positive integers n, t, d with $(n+1)t \leq N$ and an isolator $\mathsf{Iso}\colon \{0,1\}^n \to \{0,1\}$ with suitable parameters for the class of polynomial sources on $\{0,1\}^n$ of degree d such that the following holds.*

Define the distribution

$$\mathbf{D} = (\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_t, \mathsf{Iso}(\mathbf{U}_1), \mathsf{Iso}(\mathbf{U}_2), \ldots, \mathsf{Iso}(\mathbf{U}_t)),$$

where $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_t$ are independent copies of $\mathbf{U}^n$. Let $\mathcal{X}$ be the class of all polynomial sources $\mathbf{X}$ on $\{0,1\}^{(n+1)t}$ of degree Δ. Then for any $\mathbf{X} \in \mathcal{X}$,

$$\|\mathbf{D} - \mathbf{X}\|_{\mathsf{TV}} \geq 1 - O\left(\frac{\Delta}{\log\log N} \log\left(\frac{\log\log N}{\Delta}\right)\right).$$

Moreover, the distribution $\mathbf{D}$ can be sampled in $\mathsf{poly}(N)$ time.

Additionally, there exists an isolator $\widetilde{\mathsf{Iso}}\colon \{0,1\}^n \to \{0,1\}$ (with possibly different parameters) such that the distribution $\widetilde{\mathbf{D}} = (\mathbf{U}^n, \widetilde{\mathsf{Iso}}(\mathbf{U}^n))$ satisfies $\left\|\widetilde{\mathbf{D}} - \mathbf{Y}\right\|_{\mathsf{TV}} \geq 1/4 - n^{-\Omega(1)}$ for all degree-Δ polynomial sources $\mathbf{Y}$ on $\{0,1\}^{n+1}$. $\widetilde{\mathbf{D}}$, too, can be sampled in $\mathsf{poly}(N)$ time.

We prove Theorem 14 by modifying the construction of an explicit extractor from [13] to create an isolator Iso for low-degree polynomial sources, and invoke Theorem 8 using Iso.

The extractor from [13] is constructed by brute-forcing over many possible extractors on a small number of input bits. The existence of such an extractor is guaranteed by the probabilistic method (as is standard), but what makes their argument not straightforward is that the number of polynomial sources is not bounded, since a polynomial source can have any number of input bits. Their key ingredient is an input reduction technique [13, Theorem 4.1], which shows that it is enough to consider only sources where the number of input bits is at most a constant factor times the min-entropy.

We observe below that a variant of their input reduction argument also works for isolators. We first prove a variant of their simple entropy smoothing claim [13, Claim 4.3].

▷ Claim 15. For any random variable $\mathbf{X}$ over $\{0,1\}^n$ and positive integer k, there exists a function $S\colon \{0,1\}^n \to \{0,1\}^{k+1}$ with the following property. For every $x \in \{0,1\}^n$ such that $\mathbf{Pr}[\mathbf{X} = x] \le 2^{-k}$, we have $\mathbf{Pr}[S(\mathbf{X}) = S(x)] \le 2^{-k}$.

Proof. We perform the following merging operation. Start with each string in $\{0,1\}^n$ in its own bucket. If there are two buckets whose combined probability mass under $\mathbf{X}$ is at most 2^{-k}, then merge the two buckets. This operation is repeated for as long as possible. (This process necessarily terminates since the number of buckets decreases in each step.)

At the end, there can be at most one bucket whose corresponding probability is at most $2^{-(k+1)}$, since otherwise we could merge two such buckets. This implies that there are at most 2^{k+1} buckets in total, and we can define S to simply map each string $x \in \{0,1\}^n$ to the bucket containing it (where we associate each bucket with a distinct string in $\{0,1\}^{k+1}$).

To verify the desired property of S, first observe that since any bucket containing two or more elements is the result of some merge operation, it must have total probability at most 2^{-k}. This handles all x such that $|S(x)| \ge 2$. On the other hand, if $\mathbf{Pr}[\mathbf{X} = x] \le 2^{-k}$ and $|S(x)| = 1$, then clearly $\mathbf{Pr}[S(\mathbf{X}) = S(x)] = \mathbf{Pr}[\mathbf{X} = x] \le 2^{-k}$. ◁

We also need the following general lemma, which is implicit [13]. We sketch the proof for completeness.

▶ **Lemma 16.** *Let $f\colon \mathbb{F}_2^r \to \mathbb{F}_2^n$ be a function and $0 < \varepsilon < 1/4$. Let $\ell = \lceil n + 3\log(1/\varepsilon) \rceil$. If $r > \ell$, there exist $A \in \mathbb{F}_2^{r \times \ell}$ and $b \in \mathbb{F}_2^r$ such that if we define $h\colon \mathbb{F}_2^\ell \to \mathbb{F}_2^n$ by $h(x) = f(Ax+b)$, we have $\left\| f(\mathbf{U}^r) - h(\mathbf{U}^\ell) \right\|_{\mathsf{TV}} \le 2\varepsilon$.*

Proof. Lemma 4.2 in [13] gives a full rank[6] linear map $M\colon \mathbb{F}_2^r \to \mathbb{F}_2^{r-\ell}$ such that

$$\left\| f(\mathbf{U}^r) \circ M(\mathbf{U}^r) - f(\mathbf{U}^r) \circ \mathbf{U}^{r-\ell} \right\|_{\mathsf{TV}} \le 2\varepsilon. \tag{6}$$

We start by observing that

$$\left\| f(\mathbf{U}^r) \circ M(\mathbf{U}^r) - f(\mathbf{U}^r) \circ \mathbf{U}^{r-\ell} \right\|_{\mathsf{TV}} = \mathop{\mathbb{E}}_{v \sim M(\mathbf{U}^r)} \Big[\left\| (f(\mathbf{U}^r) \,|\, M(\mathbf{U}^r) = v) - f(\mathbf{U}^r) \right\|_{\mathsf{TV}} \Big],$$

where we have used that M has full rank, so $M(\mathbf{U}^r)$ is $\mathbf{U}^{r-\ell}$. Combining with (6), there must exist some $v \in \mathbb{F}_2^{r-\ell}$ such that

$$\left\| (f(\mathbf{U}^r) \,|\, M(\mathbf{U}^r) = v) - f(\mathbf{U}^r) \right\|_{\mathsf{TV}} \le 2\varepsilon.$$

Now $(f(\mathbf{U}^r) \,|\, M(\mathbf{U}^r) = v)$ can be expressed as $h(\mathbf{U}^\ell)$ for a function h on ℓ inputs of the desired form $h(x) = f(Ax + b)$, since conditioning on $M(\mathbf{U}^r) = v$ (where M has full rank) is equivalent to replacing $r - \ell$ of the r input bits for f by affine functions of the other ℓ inputs. Thus, we have $\left\| h(\mathbf{U}^\ell) - f(\mathbf{U}^r) \right\|_{\mathsf{TV}} \le 2\varepsilon$ as desired. ◀

[6] The conference version of [13] does not explicit state that M is full rank, but this can be found in the updated arXiv version: `https://arxiv.org/abs/2309.11019`.

We now proceed to the input reduction lemma.

▶ **Lemma 17.** *Suppose* $\mathsf{Iso}\colon \mathbb{F}_2^n \to \mathbb{F}_2$ *is an* $(\alpha, \beta, k-1)$*-isolator for the class of degree-*d *polynomial sources with at most* $4(k+1)$ *inputs. Then* Iso *is also an* $(\alpha, \beta + 2^{-k}, k)$*-isolator for the class of all degree-*d *polynomial sources.*

Proof. We clearly have $\mathbf{Pr}[\mathsf{Iso}(\mathbf{U}^n) = 1] \geq \alpha$ since Iso is an $(\alpha, \beta, k-1)$-isolator, so it remains to verify that Iso satisfies the second condition in the definition of an $(\alpha, \beta + 2^{-k}, k)$-isolator for the class of all degree-d polynomial sources.

Let $f\colon \mathbb{F}_2^r \to \mathbb{F}_2^n$ be a degree-d polynomial map with light points $L_f \coloneqq \{x : \mathbf{Pr}[f(\mathbf{U}^r) = x] \leq 1/2^k\}$. Let $\ell = 4(k+1)$, and let $S\colon \{0,1\}^n \to \{0,1\}^{k+1}$ be the map given by Claim 15 applied to $f(\mathbf{U}^r)$ and k. Apply Lemma 16 to the function $S(f(\cdot))$ with $\varepsilon = 2^{-(k+1)}$ to get $A \in \mathbb{F}_2^{r \times \ell}, b \in \mathbb{F}_2^r$ such that

$$\left\| S(f(A \cdot \mathbf{U}^\ell + b)) - S(f(\mathbf{U}^r)) \right\|_{\mathsf{TV}} \leq 2^{-k}. \tag{7}$$

Let $g\colon \mathbb{F}_2^\ell \to \mathbb{F}_2^n$ be defined by $g(x) = f(Ax + b)$. Observe that g is a degree-d polynomial since we have only substituted linear polynomials for the inputs of f.

Define $L_g = \{x : \mathbf{Pr}[g(\mathbf{U}^\ell) = x] \leq 2^{-(k-1)}\}$. Observe that $L_f \subseteq L_g$, since (7) implies every $x \in L_f$ satisfies

$$\mathbf{Pr}[g(\mathbf{U}^\ell) = x] \leq \mathbf{Pr}[S(g(\mathbf{U}^\ell)) = S(x)] \leq \mathbf{Pr}[S(f(\mathbf{U}^r)) = S(x)] + 2^{-k} \leq 2^{-(k-1)}.$$

Therefore, we conclude

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\mathbf{x} \sim f(\mathbf{U}^r)}[\mathbf{x} \in L_f \text{ and } \mathsf{Iso}(\mathbf{x}) = 1] &\leq \mathop{\mathbf{Pr}}_{\mathbf{x} \sim g(\mathbf{U}^\ell)}[\mathbf{x} \in L_f \text{ and } \mathsf{Iso}(\mathbf{x}) = 1] + 2^{-k} && \text{(by (7))} \\
&\leq \mathop{\mathbf{Pr}}_{\mathbf{x} \sim g(\mathbf{U}^\ell)}[\mathbf{x} \in L_g \text{ and } \mathsf{Iso}(\mathbf{x}) = 1] + 2^{-k} && \text{(since } L_f \subseteq L_g\text{)} \\
&\leq \beta + 2^{-k}.
\end{aligned}$$

The last inequality above uses the isolator guarantee for $g(\mathbf{U}^\ell)$. This finishes the proof. ◀

To apply Lemma 17, we need to construct an isolator for the class of polynomial sources with bounded input length. The following lemma will allow us to find such an object more efficiently than a naive brute force. Below, recall that a family $\{f_i\colon \{0,1\}^n \to \{0,1\}^m\}_i$ of t-wise uniform hash functions is defined by the property that for any string $(y_1, y_2, \ldots, y_t) \in (\{0,1\}^m)^t$ and all distinct strings $x_1, x_2, \ldots, x_t \in \{0,1\}^n$, a function $\mathbf{f}$ chosen uniformly at random from the family satisfies $(\mathbf{f}(x_1), \mathbf{f}(x_2), \ldots, \mathbf{f}(x_t)) = (y_1, y_2, \ldots, y_t)$ with probability 2^{-mt}.

▶ **Lemma 18.** *Let* $0 < \alpha < \beta < 1$*. Let* k, n, ℓ*, and* d *be positive integers. Set* $K = 2^k$ *and* $N = 2^n$*. Let* $t \geq 4$ *be an even integer. Suppose there exists an integer* m *such that* $p \coloneqq 2^{-m}$ *satisfies* $\alpha < p < \beta$ *and the following inequality holds:*

$$\left(\frac{Npt + t^2}{N^2(p-\alpha)^2}\right)^{t/2} + 2^{\binom{\ell}{\leq d}n}\left(\frac{Kpt + t^2}{K^2(\beta - p)^2}\right)^{t/2} < 1/8.$$

Additionally, let $\mathcal{H}$ *be a family of* t*-wise uniform hash functions from* $\{0,1\}^n$ *to* $\{0,1\}^m$*. Then there exists a function* $h \in \mathcal{H}$ *such that the function* g *defined by* $g(x) = \mathbf{1}(h(x) = 0^m)$ *is an* (α, β, k)*-isolator for the class of degree-*d *polynomial sources with* ℓ *inputs and* n *outputs.*

Proof. Let $\mathbf{h}$ be a function drawn uniformly at random from $\mathcal{H}$, and define $\mathbf{g}\colon \{0,1\}^n \to \{0,1\}$ by $\mathbf{g}(x) = \mathbf{1}(\mathbf{h}(x) = 0^m)$. By the assumption on $\mathcal{H}$, $\{\mathbf{g}(x)\}_{x\in\{0,1\}^n}$ is a collection of t-wise independent random variables, and for each x, $\mathbf{Pr}[\mathbf{g}(x) = 1] = 2^{-m}$. By using a tail inequality for sums of t-wise independent random variables (see [9, Lemma 2.3]), we have

$$\mathop{\mathbf{Pr}}_{\mathbf{g}}[\mathbf{Pr}[\mathbf{g}(\mathbf{U}^n) = 1] \le \alpha] \le 8\left(\frac{Npt + t^2}{N^2(p-\alpha)^2}\right)^{t/2}.$$

This is the probability that $\mathbf{g}$ fails to satisfy the first condition for being an (α, β, k)-isolator.

We will now estimate the probability that for some source, the second condition for being an (α, β, k)-isolator is not satisfied. Fix any source $\mathbf{X}$ on $\{0,1\}^n$, and define $L = \{x : \mathbf{Pr}[\mathbf{X} = x] \le 2^{-k}\}$. We wish to show that with high probability over $\mathbf{g}$, $\mathbf{Pr}[\mathbf{g}(\mathbf{X}) = 1 \text{ and } \mathbf{X} \in L] \le \beta$. Let $\mathbf{Z} = \mathbf{Pr}[\mathbf{g}(\mathbf{X}) = 1 \text{ and } \mathbf{X} \in L]$ be the random variable of interest. For each $x \in L$, define the indicator random variable $\mathbf{Z}_x = \mathbf{1}(\mathbf{g}(x) = 1)$, each of which is a Bernoulli random variable with probability p of being 1. Then $\mathbf{Z} = \sum_{x\in L} \mathbf{Pr}[\mathbf{X} = x]\mathbf{Z}_x$ and $\mathbb{E}[\mathbf{Z}] = \mathbf{Pr}[\mathbf{X} \in L] \cdot p \le p$.

By a tail inequality for t-wise independent random variables as above and using that $\mathbf{Pr}[\mathbf{X} = x] \le 2^{-k}$ for all $x \in L$, we obtain

$$\begin{aligned}\mathbf{Pr}[\mathbf{Z} \ge \beta] &= \mathbf{Pr}[\mathbf{Z} \ge \mathbb{E}[\mathbf{Z}] + \beta - \mathbf{Pr}[\mathbf{X} \in L] \cdot p] \\ &\le \mathbf{Pr}[\mathbf{Z} \ge \mathbb{E}[\mathbf{Z}] + \beta - p] \\ &\le 8\left(\frac{Kpt + t^2}{K^2(\beta-p)^2}\right)^{t/2}.\end{aligned}$$

Now a union bound over the $2^{\binom{\ell}{\le d}n}$ many degree-d polynomial sources with ℓ inputs and n outputs gives that the second condition for being an (α, β, k)-isolator for such sources does not hold with probability at most $8 \cdot 2^{\binom{\ell}{\le d}n} \cdot \left(\frac{Kpt+t^2}{K^2(\beta-p)^2}\right)^{t/2}$. Combining this with the failure probability for the first condition shows that g fails to be an (α, β, k)-isolator for such sources with probability at most

$$8\left(\left(\frac{Npt + t^2}{N^2(\alpha-p)^2}\right)^{t/2} + 2^{\binom{\ell}{\le d}n}\left(\frac{Kpt + t^2}{K^2(\beta-p)^2}\right)^{t/2}\right) < 1$$

by assumption. Hence there exists some $\mathbf{g}$ which is an (α, β, k)-isolator, as desired. ◀

By iterating through a t-wise uniform family of hash functions, we must find an isolator in a reasonable amount of time.

▶ **Corollary 19.** *Let k, n, d, and m be positive integers. Suppose $k \ge 10(d + \log n)$, $k \le n$, and $m \le 0.01k$. For $p = 2^{-m}$, there exist $\alpha = p - 2^{-\Omega(n)}$ and $\beta = p + 2^{-\Omega(k)}$ such that there exists an (α, β, k)-isolator* Iso *for the class of degree-d polynomials with n outputs, which can be computed in time $2^{O(\binom{\Theta(k)}{\le d}n^2)}$.*

Proof. Set $\ell = 5k$, $K = 2^k$, and $t = 2\left(\binom{\ell}{\le d}n + 4\right)$; note that $K^{0.99} \ge 4t$. By using Lemma 18 under the conditions on k, d in the statement, there exists a function Iso which is an $(\alpha', \beta', k-1)$-isolator for the class of degree-d polynomial sources with $4(k+1) \le 5k$ inputs and n outputs, where $\alpha' = p - 2^{-\Omega(n)}$ and $\beta' = p + 2^{-\Omega(k)}$. Furthermore by Lemma 17, Iso is an (α, β, k)-isolator for the class of degree-d polynomial sources with n outputs, where $\alpha = \alpha' = p - 2^{-\Omega(n)}$ and $\beta = \beta' + 2^{-k} = p + 2^{-\Omega(k)}$.

We compute such a function by going over any fixed t-wise uniform family of hash functions from $\{0,1\}^n$ to $\{0,1\}^m$ until we find a function that lets us construct an isolator as described in Lemma 18. It is well known that there is such a family $\mathcal{H}$ of size 2^{tn} where each

function in the family can be evaluated in $\mathsf{poly}(n, m, t)$ time (see, for instance, [43, Corollary 3.34]). For any such fixed function $h \in \mathcal{H}$, we check whether g defined by $g(x) = \mathbf{1}(h(x) = 0^m)$ is an (α, β, k)-isolator. This can be done in time $2^{O(\binom{\ell}{\leq d} n)}$. ◀

By invoking Theorem 8 with the above isolator, we obtain an explicit hard distribution for polynomial sources, proving Theorem 14. We restate the theorem below for the reader's convenience.

▶ **Theorem 14.** *There exists a constant $\delta > 0$ such that the following holds. Let N and Δ be positive integers satisfying $\Delta \leq \delta \log\log N$. There exist positive integers n, t, d with $(n+1)t \leq N$ and an isolator $\mathsf{Iso}\colon \{0,1\}^n \to \{0,1\}$ with suitable parameters for the class of polynomial sources on $\{0,1\}^n$ of degree d such that the following holds.*

Define the distribution

$$\mathbf{D} = (\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_t, \mathsf{Iso}(\mathbf{U}_1), \mathsf{Iso}(\mathbf{U}_2), \ldots, \mathsf{Iso}(\mathbf{U}_t)),$$

where $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_t$ are independent copies of $\mathbf{U}^n$. Let $\mathcal{X}$ be the class of all polynomial sources $\mathbf{X}$ on $\{0,1\}^{(n+1)t}$ of degree Δ. Then for any $\mathbf{X} \in \mathcal{X}$,

$$\|\mathbf{D} - \mathbf{X}\|_{\mathsf{TV}} \geq 1 - O\left(\frac{\Delta}{\log\log N} \log\left(\frac{\log\log N}{\Delta}\right)\right).$$

Moreover, the distribution $\mathbf{D}$ can be sampled in $\mathsf{poly}(N)$ time.

Additionally, there exists an isolator $\widetilde{\mathsf{Iso}}\colon \{0,1\}^n \to \{0,1\}$ (with possibly different parameters) such that the distribution $\widetilde{\mathbf{D}} = (\mathbf{U}^n, \widetilde{\mathsf{Iso}}(\mathbf{U}^n))$ satisfies $\left\|\widetilde{\mathbf{D}} - \mathbf{Y}\right\|_{\mathsf{TV}} \geq 1/4 - n^{-\Omega(1)}$ for all degree-Δ polynomial sources $\mathbf{Y}$ on $\{0,1\}^{n+1}$. $\widetilde{\mathbf{D}}$, too, can be sampled in $\mathsf{poly}(N)$ time.

Proof. We choose parameters to optimize the final distance with respect to the distribution length $(n+1)t$ while ensuring the construction takes only $\mathsf{poly}(N)$ time. Set $n = \lfloor (\log N)^{1/(2+\lambda)} \rfloor$ for a constant $\lambda > 0$ to be determined later. Let $d = \log n$ and $t = \lfloor d/\Delta \rfloor - 1$ so that $\Delta(t+1) \leq d$. By picking δ to be small enough, we have $t \geq 1$. Let $k = \lceil 20 \log n \rceil$ so that the condition $k \geq 10(\log n + d)$ in Corollary 19 is satisfied. Let m be the largest integer such that for $p = 2^{-m}$, we have $t \geq \frac{1}{p} \log \frac{1}{p}$. Since $t \leq d = \log n$ and $m \leq \log t$, we have $m \leq 0.01k$. Now let Iso be the (α, β, k)-isolator given by Corollary 19 with the chosen parameters and $\alpha = p - 2^{-\Omega(n)}, \beta = p + 2^{-\Omega(k)}$.

Let $\mathbf{X} \in \mathcal{X}$. Note that $\mathsf{addr}_{n,t}(\mathbf{X})$ can be computed by a degree $\Delta(t+1) \leq d$ polynomial source since $\mathsf{addr}_{n,t}$ can be computed by a degree-$(t+1)$ polynomial. By Theorem 8, we have

$$\begin{aligned}
1 - \|\mathbf{D} - \mathbf{X}\|_{\mathsf{TV}} &\leq (1-\alpha)^t + \beta + (2t^3+1)2^{-(n-k)} \\
&\leq (1 - p + 2^{-\Omega(n)})^t + p + 2^{-\Omega(k)} + (2t^3+1) \cdot 2^{-\Omega(n)} \\
&\leq (1-p)^t + t \cdot 2^{-\Omega(n)} + p + 2^{-\Omega(k)} + 2^{-\Omega(n)} \\
&\leq \exp(-pt) + p + n^{-\Omega(1)} \\
&\leq O(p) \leq O\left(\frac{\log t}{t}\right) \\
&\leq O\left(\frac{\Delta}{\log\log N} \log\left(\frac{\log\log N}{\Delta}\right)\right).
\end{aligned}$$

Finally, we show that $\mathbf{D}$ can be sampled in time $\mathsf{poly}(N)$. The isolator Iso can be computed in time $2^{O(\binom{\Theta(k)}{\leq d} n^2)}$. We have

$$\binom{\Theta(k)}{\leq d} \leq \left(\frac{e \cdot \Theta(k)}{d}\right)^d \leq C_0^{\log n} = n^{\log C_0}$$

for some constant $C_0 > 1$. Set $\lambda = \log C_0$, so Iso can be computed in time $2^{O(n^{2+\lambda})} = \mathsf{poly}(N)$. Given that Iso is computable in $\mathsf{poly}(N)$ time, it is straightforward to see that $\mathbf{D}$ can be sampled in $\mathsf{poly}(N)$ time.

We now specialize to the case of $t = 1$ and $\widetilde{\mathbf{D}} = (\mathbf{U}^n, \widetilde{\mathsf{Iso}}(\mathbf{U}^n))$, where $\widetilde{\mathsf{Iso}}$ is an $(1/2 - 2^{-\Omega(n)}, 1/2 + n^{-\Omega(1)}, 20\log n)$-isolator. Invoking Theorem 8, we obtain

$$\left\|\widetilde{\mathbf{D}} - \mathbf{Y}\right\|_{\mathsf{TV}} \geq 1 - (1/2 + 2^{-\Omega(n)})^2 - (1/2 + n^{-\Omega(1)}) - 2^{-\Omega(n)} \geq 1/4 - n^{-\Omega(1)}. \quad \blacktriangleleft$$

Note that our hard distribution above has at most N bits instead of exactly N bits. To obtain a distribution with exactly N bits, one can pad with some bits that are fixed to, say, 0. It is clear that the padded distribution does not become easier for polynomial sources of degree Δ since TV distance cannot increase by applying a projection. It is also easy to see that the padded distribution is computable in $\mathsf{poly}(N)$ time if the original distribution is computable in $\mathsf{poly}(N)$ time.

5 Open Problems

We conclude by highlighting several directions for future research. One notable goal is to bring our understanding of sampling by small circuits with parity gates (i.e., $\mathsf{AC}^0[\oplus]$) up to the frontier of what is known in the computational setting.

1. Provide an explicit distribution over $\{0,1\}^n$ which has distance $1 - o(1)$ from the output of any $\mathsf{AC}^0[\oplus]$ circuit with n outputs, $\mathsf{poly}(n)$ many gates, and arbitrarily many random input bits.

Even obtaining a much weaker bound would be novel; to the best of our knowledge, no hardness results for explicit[7] distributions exist beyond the trivial arguments of finding a hard distribution for extremely small circuits via brute force, or appealing to the limitations posed by binary precision. For example, the $(1/3)$-biased distribution over $\{0,1\}$ cannot be exactly generated by a circuit with r random bits as input, since each output occurs with probability that is an integer multiple of 2^{-r}. Of course, one could also ask for a similarly hard distribution in the case of circuits with mod p gates for any other prime p. Without these more powerful gates, it is known that such circuits cannot accurately generate the uniform distribution over the codewords of a good code [35, 8]. Unfortunately, the arguments do not seem to generalize.

In light of Theorem 14, a reasonable approach to Question 1 is to construct an isolator for $\mathbb{F}_2$-polynomial sources of polylogarithmic degree, and then appeal to classical results on polynomial approximations of circuits [37, 41]. This is a more general degree constraint than the best-known explicit extractors can handle [13], but we are optimistic further improvements in these extractors will be flexible enough for our techniques to apply. It is also worth emphasizing that extractors and isolators are incomparable objects, and it is plausible that the latter may be easier to construct in certain settings.

An alternative strengthening of Theorem 14 is to improve the quantitative behavior; while we can prove a $1 - o(1)$ bound, the specific decay rate of the $o(1)$ term is suboptimal.

2. Provide an explicit distribution over $\{0,1\}^n$ which has distance $1 - \exp(-n^{\Omega_d(1)})$ from the output of any degree-d $\mathbb{F}_2$-polynomial source.

[7] Existential results can be found in the appendix of the full version.

We note that one can prove such a bound in the case of $d = 1$, where the source $P(\mathbf{U}^r)$ is uniform over an affine subspace. Letting $\mathsf{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ be an (ε, k)-extractor for affine sources, we can set the hard distribution $\mathbf{D}$ to be uniform over the preimage $\mathsf{Ext}^{-1}(0^m)$. If $P(\mathbf{U}^r)$ has min-entropy at least k, then it has distance at least $1 - 2^{-m} - \varepsilon$ from $\mathbf{D}$. Otherwise, $P(\mathbf{U}^r)$ is uniform over a set of size less than 2^k, so it has distance at least $1 - 2^{-(n-k)}/(2^{-m} - \varepsilon)$ from $\mathbf{D}$. By known explicit constructions of such objects [12, 51, 34], we can take $k = \Omega(n)$, $\varepsilon = 2^{-\Omega(n)}$, and $m = \Omega(n)$ with a sufficiently small implicit constant to conclude $\|P(\mathbf{U}^r) - \mathbf{D}\|_{\mathsf{TV}} \geq 1 - \exp(-\Omega(n))$.

Sadly, this argument already breaks down for $d = 2$, since we no longer get meaningful information about $\mathsf{supp}\,(P(\mathbf{U}^r))$ in the low min-entropy case. Moreover, our current framework does not appear strong enough to address Question 2, and it would be interesting to determine whether the framework itself can be strengthened.

3. Prove a quantitatively stronger version of Theorem 8.

More specifically, it would be desirable to have the β term in Theorem 8 decrease as t increases. It also seems worthwhile to try to improve the $t = 1$ setting to be able to obtain the optimal form $\frac{1}{2} - o(1)$, rather than our current $\frac{1}{4} - o(1)$. In both cases, one may wish to consider a variant of extractors even beyond robust extractors or isolators.

References

1 Miklós Ajtai. Σ_1^1-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983. doi:10.1016/0168-0072(83)90038-6.

2 Yaroslav Alekseev, Mika Göös, Konstantin Myasnikov, Artur Riazanov, and Dmitry Sokolov. Sampling permutations with cell probes is hard. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, 2026.

3 Omar Alrabiah, Jesse Goodman, Jonathan Mosheiff, and João Ribeiro. Low-degree polynomials are good extractors. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 353 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 38, 25. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/lipics.approx/random.2025.38.

4 Andris Ambainis, Leonard J. Schulman, Amnon Ta-Shma, Umesh Vazirani, and Avi Wigderson. The quantum communication complexity of sampling. *SIAM J. Comput.*, 32(6):1570–1585, 2003. doi:10.1137/S009753979935476.

5 Anurag Anshu, Yangjing Dong, Fengning Ou, and Penghui Yao. On the computational power of QAC0 with barely superlinear ancillae. In *STOC'25—Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1476–1487. ACM, New York, 2025. doi:10.1145/3717823.3718189.

6 László Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Information Processing Letters*, 26(1):51–53, 1987. doi:10.1016/0020-0190(87)90036-6.

7 Marshall Ball, Ronen Shaltiel, and Jad Silbak. Extractors for samplable distributions with low min-entropy. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 596–603, 2025. doi:10.1145/3717823.3718143.

8 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. IEEE, 2012. doi:10.1109/FOCS.2012.82.

9 Mihir Bellare and John Rompel. Randomness-efficient oblivious sampling. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 276–287. IEEE, 1994. doi:10.1109/SFCS.1994.365687.

10 Adam Bene Watts and Natalie Parham. Unconditional quantum advantage for sampling with shallow circuits. In *17th Innovations in Theoretical Computer Science Conference*, volume 362 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 17, 12. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. doi:10.4230/lipics.itcs.2026.17.

11 Ravi B Boppana and Jeffrey C Lagarias. One-way functions and circuit complexity. *Information and Computation*, 74(3):226–240, 1987. doi:10.1016/0890-5401(87)90022-8.

12 Jean Bourgain. On the construction of affine extractors. *Geom. Funct. Anal.*, 17(1):33–57, 2007. doi:10.1007/s00039-007-0593-z.

13 Eshan Chattopadhyay, Jesse Goodman, and Mohit Gurumukhani. Extractors for polynomial sources over $\mathbb{F}_2$. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.ITCS.2024.28.

14 Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The space complexity of sampling. In *13th Innovations in Theoretical Computer Science Conference (ITCS)*, 2022.

15 Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory (TOCT)*, 4(1):1–21, 2012. doi:10.1145/2141938.2141941.

16 Harm Derksen, Peter Ivanov, Chin Ho Lee, and Emanuele Viola. Pseudorandomness, symmetry, smoothing: I. In *39th Computational Complexity Conference*, volume 300 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 18, 27. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/lipics.ccc.2024.18.

17 Dean Doron and William M. Hoza. Implications of better PRGs for permutation branching programs. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 353 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 28, 20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/lipics.approx/random.2025.28.

18 Stephen Fenner, Daniel Grier, Daniel Padé, and Thomas Thierauf. Tight bounds on depth-2 QAC-circuits computing parity. *arXiv preprint*, 2025. doi:10.48550/arXiv.2504.06433.

19 Merrick Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Math. Systems Theory*, 17(1):13–27, 1984. doi:10.1007/BF01744431.

20 Alexander Golovnev, Zeyu Guo, Pooya Hatami, Satyajeet Nagargoje, and Chao Yan. Hilbert functions and low-degree randomness extractors. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 317 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 41, 24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/lipics.approx/random.2024.41.

21 Mika Göös and Thomas Watson. A lower bound for sampling disjoint sets. *ACM Trans. Comput. Theory*, 12(3):Art. 20, 13, 2020. doi:10.1145/3404858.

22 Daniel Grier, Daniel M. Kane, Jackson Morris, Anthony Ostuni, and Kewen Wu. Quantum advantage from sampling shallow circuits: beyond hardness of marginals. In *17th Innovations in Theoretical Computer Science Conference*, volume 362 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 73, 14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. doi:10.4230/lipics.itcs.2026.73.

23 Daniel Grier, Jackson Morris, and Kewen Wu. QAC^0 contains TC^0 (with many copies of the input). *arXiv preprint*, 2026. arXiv:2601.03243.

24 Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, pages 6–20, 1986. doi:10.1145/12130.12132.

25 Johan Håstad. *Computational limitations for small depth circuits*. PhD thesis, Massachusetts Institute of Technology, 1986.

26 Pooya Hatami and William Hoza. Theory of unconditional pseudorandom generators. *Foundations and Trends in Theoretical Computer Science*, 16(1-2):1–210, February 2024. doi:10.1561/0400000109.

27 William M. Hoza and Zelin Lv. On sums of INW pseudorandom generators. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 353 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 67, 24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/lipics.approx/random.2025.67.

28 Malvika Raj Joshi, Avishay Tal, Francisca Vasconcelos, and John Wright. Improved lower bounds for QAC^0. *arXiv preprint*, 2025. arXiv:2512.14643.

29 Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locality bounds for sampling Hamming slices. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1279–1286, 2024. doi:10.1145/3618260.3649670.

30 Daniel M. Kane, Anthony Ostuni, and Kewen Wu. Symmetric distributions from shallow circuits. *arXiv preprint*, 2025. doi:10.48550/arXiv.2511.14127.

31 Mohammad Mahdi Khodabandeh and Igor Shinkar. On sampling lower bounds for polynomials, 2026. Available at https://eccc.weizmann.ac.il/report/2026/066/.

32 Vinayak M. Kumar. New pseudorandom generators and correlation bounds using extractors. In *16th Innovations in Theoretical Computer Science Conference*, volume 325 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 68, 23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/lipics.itcs.2025.68.

33 Chin Ho Lee and Emanuele Viola. Pseudorandom bits for non-commutative programs. In *40th Computational Complexity Conference*, volume 339 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 9, 22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/lipics.ccc.2025.9.

34 Xin Li. A new approach to affine extractors and dispersers. In *26th Annual IEEE Conference on Computational Complexity*, pages 137–147. IEEE Computer Soc., Los Alamitos, CA, 2011. doi:10.1109/CCC.2011.27.

35 Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 243–251. IEEE, 2011. doi:10.1109/CCC.2011.11.

36 Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. On the Pauli spectrum of QAC0. In *STOC'24—Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1498–1506. ACM, New York, 2024. doi:10.1145/3618260.3649662.

37 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.

38 R. Renner and S. Wolf. Smooth Renyi entropy and applications. In *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, pages 233–, 2004. doi:10.1109/ISIT.2004.1365269.

39 Ronen Shaltiel. Multiplicative extractors for samplable distributions. In *40th Computational Complexity Conference*, volume 339 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 22, 22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/lipics.ccc.2025.22.

40 Ronen Shaltiel and Jad Silbak. Explicit codes for poly-size circuits and functions that are hard to sample on low entropy distributions. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 2028–2038, 2024. doi:10.1145/3618260.3649735.

41 Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 77–82, 1987. doi:10.1145/28395.28404.

42 Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 32–42. IEEE, 2000. doi:10.1109/SFCS.2000.892063.

43 Salil P Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1-3):1–336, 2012. doi:10.1561/0400000010.

44 Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012. doi:10.1137/100814998.

45 Emanuele Viola. Extractors for Turing-machine sources. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 663–671. Springer, 2012. doi:10.1007/978-3-642-32512-0_56.

46 Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014. doi:10.1137/11085983X.

47 Emanuele Viola. Quadratic maps are hard to sample. *ACM Transactions on Computation Theory (TOCT)*, 8(4):1–4, 2016. doi:10.1145/2934308.

48 Emanuele Viola. Sampling lower bounds: boolean average-case and permutations. *SIAM Journal on Computing*, 49(1):119–137, 2020. doi:10.1137/18M1198405.

49 Emanuele Viola. New Sampling Lower Bounds via the Separator. In *38th Computational Complexity Conference (CCC 2023)*, volume 264 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 26:1–26:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.CCC.2023.26.

50 Andrew Chi-Chih Yao. Separating the polynomial-time hierarchy by oracles. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 1–10. IEEE, 1985.

51 Amir Yehudayoff. Affine extractors over prime fields. *Combinatorica*, 31(2):245–256, 2011. doi:10.1007/s00493-011-2604-9.

52 Huacheng Yu and Wei Zhan. Sampling, Flowers and Communication. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 100:1–100:11. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.ITCS.2024.100.