Sampling and Certifying Symmetric Functions

Yuval Filmus
Technion – Israel Institute of Technology, Haifa, Israel

Itai Leigh
Tel-Aviv University, Israel

Artur Riazanov
École Polytechnique Fédérale de Lausanne, Switzerland

Dmitry Sokolov
École Polytechnique Fédérale de Lausanne, Switzerland

Abstract

A circuit $\mathcal{C}$ samples a distribution $\boldsymbol{X}$ with an error ε if the statistical distance between the output of $\mathcal{C}$ on the uniform input and $\boldsymbol{X}$ is ε. We study the hardness of sampling a uniform distribution over the set of n-bit strings of Hamming weight k denoted by $\boldsymbol{U}_k^n$ for *decision forests*, i.e. every output bit is computed as a decision tree of the inputs. For every k there is an $O(\log n)$-depth decision forest sampling $\boldsymbol{U}_k^n$ with an inverse-polynomial error [26, 11]. We show that for every $\varepsilon > 0$ there exists τ such that for decision depth $\tau \log(n/k)/\log\log(n/k)$, the error for sampling $\boldsymbol{U}_k^n$ is at least $1 - \varepsilon$. Our result is based on the recent robust sunflower lemma [1, 23].

Our second result is about matching a set of n-bit strings with the image of a *d-local* circuit, i.e. such that each output bit depends on at most d input bits. We study the set of all n-bit strings whose Hamming weight is at least $n/2$. We improve the previously known locality lower bound from $\Omega(\log^* n)$ [5] to $\Omega(\sqrt{\log n})$, leaving only a quartic gap from the best upper bound of $O(\log^2 n)$.

2012 ACM Subject Classification Theory of computation → Circuit complexity; Theory of computation → Generating random combinatorial structures

Keywords and phrases sampling, lower bounds, robust sunflowers, decision trees, switching networks

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.36

Category RANDOM

Related Version *Full Version*: https://arxiv.org/abs/2305.04363

Funding This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 802020-ERC-HARMONIC. This work was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number MB22.00026.

Acknowledgements We thank Mika Göös, Aleksandr Smal, and Anastasia Sofronova for fruitful discussions and suggestions. We thank the RANDOM 2023 anonymous referees for their helpful comments.

1 Introduction

Studying the hardness of sampling has been proposed in the paper [26], which spurred an active line of research [27, 21, 3, 28, 29, 30, 14, 31, 6, 8]. The basic setting is the following: we are given an infinite supply of independent uniform random bits as input, and our goal is to design a circuit with multiple output bits whose output is close in statistical distance to a given target distribution. Sampling circuit constructions have been applied in

cryptography [18, 7] and algorithms [15]. Sampling hardness results[1] have been instrumental in inspiring and improving two-source extractor constructions [28, 9, 10] (see [30] for an extended discussion), and have yielded lower bounds on succinct data structures [26, 30, 31].

Sampling hardness results are more challenging than computational ones. For example, while it is known since Smolensky's classical work [25] that parity requires an exponential number of gates to compute by an $\mathbf{AC}^0[3]$ circuit, no hard distributions are known for the circuit class $\mathbf{AC}^0[p]$ for any p, and while $\mathbf{AC}^0$ requires exponentially many gates to compute parity [17], a random vector with parity 0 can be sampled by an $\mathbf{NC}^0$ circuit. Moreover, this distribution can be sampled by a 2-local circuit, in the sense that each output bit depends only on two input bits [2, 19]. A very simple mapping achieves this: $(x_1, \ldots, x_n) \mapsto (x_1, x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_{n-1} \oplus x_n, x_n)$. A more general and striking fact is that $\mathbf{AC}^0$ can sample random permutations and all distributions of form $(\boldsymbol{X}, f(\boldsymbol{X}))$ where $\boldsymbol{X}$ is uniform over $\{0,1\}^n$ and f is *symmetric* i.e. its value depends only on the Hamming weight of the input [26].

We conjecture that the power of $\mathbf{NC}^0$ in regard to sampling symmetric distributions is in essence limited to the parity example above. Observe that a function is computable by an $\mathbf{NC}^0$ circuit if and only if it is $O(1)$-*local* i.e. each of its output bits depends on at most a constant number of input bits. For simplicity, we focus only on uniform distributions with symmetric support: let $\boldsymbol{U}_S^n$ be the uniform distribution over strings in $\{0,1\}^n$ with Hamming weight in the set $S \subseteq \{0, \ldots, n\}$.

▶ **Conjecture 1.** *For every $d \in \mathbb{N}, \varepsilon \in (0,1)$ for all large enough n, if $\boldsymbol{X}$ is samplable by a d-local function and is ε-close to $\boldsymbol{U}_S^n$ for some $S \subseteq \{0, \ldots, n\}$, then $\boldsymbol{X}$ is $O(\varepsilon)$-close to $\boldsymbol{U}_T^n$, where T is one of the following: $\{0\}, \{n\}, \{0, n\}, \{0, 2, 4, \ldots\}, \{1, 3, 5, \ldots\}, [n]$.*

Our results on sampling slices (namely, Theorem 2) imply this conjecture for all sets S which only contain small values, in the sense that $\max_{x \in S} x = o(n)$.

Quantum separations

A stronger version of Conjecture 1 would identify the family of sets S such that every $\mathbf{NC}^0$-samplable distribution is $1 - o(1)$-far from $\boldsymbol{U}_S^n$. There exists a set S for which this implies a separation between $\mathbf{NC}^0$ and $\mathbf{QNC}^0$ for sampling. This is due to the recent partial separation in [32]: they show that there exists a symmetric function f such that $(\boldsymbol{X}, f(\boldsymbol{X}))$ for uniform $\boldsymbol{X} \sim \{0,1\}^n$ can be sampled by a $\mathbf{QNC}^0$ circuit. Observe, however, that if an $\mathbf{NC}^0$-samplable distribution $\boldsymbol{Y}$ is at distance η from $(\boldsymbol{X}, f(\boldsymbol{X}))$, then the first n bits of $\boldsymbol{Y}$ are $(1/2 + \eta + o(1))$-close to the uniform distribution over $f^{-1}(0)$, due to the fact that the function f used in [32] is almost balanced (in the sense that $|f^{-1}(0)| = (1 + o(1)) \cdot |f^{-1}(1)|$). Now, if the uniform distribution over $f^{-1}(0)$ is not $\mathbf{NC}^0$-samplable within the distance $1 - \Omega(1)$, we get the separation. Conjecture 1 implies a weaker lower bound for the distance: a constant instead of a function approaching 1, but most of the known lower bounds for distribution sampling have very strong distance guarantees.

Local certificates

One interesting class of sets which is also not covered by our and prior results is $M_a = \{x \in \{0, \ldots, n\} \mid x \bmod a = 0\}$, where a is a constant. However, the simple construction for sampling parity-0 vectors can be adapted to *match the support* of any $\boldsymbol{U}_{M_a}^n$, i.e. generate

[1] That is, showing that any circuit from a certain class produces a distribution that is far from the target.

a (not necessarily uniform) distribution whose support is exactly the set of strings with Hamming weight in M_a. This construction was given in [5, Proposition 3.1], where it was presented as a *proof system*. The idea is to interpret the input bits as a *certificate* that the output is in the target language (in this case, all n-bits strings whose Hamming weight is divisible by a). This connection motivates our study of locality in the context of proof systems. We drastically simplify and improve the locality lower bound of [20] for the language of n-bit strings whose Hamming weight is at least $n/2$ (in other words, 1-inputs of the majority function).

1.1 Notation

We use boldface letters for random variables, e.g. $\boldsymbol{a}, \boldsymbol{A}, \boldsymbol{b}, \boldsymbol{B}$. We write $\boldsymbol{a} \sim A$ to say that $\boldsymbol{a}$ is distributed according to a distribution A, or if A is a set, according to the uniform distribution over it. For $x \in \{0,1\}^n$, we denote its Hamming weight by $w(x) = |x| \coloneqq \{i \in [n] \mid x_i = 1\}$. We denote $e_i = 0^{i-1}10^{n-i} \in \{0,1\}^n$. For a string $x \in \{0,1\}^n$ and a set $T \subseteq [n]$, we write $x_T \coloneqq (x_i)_{i \in T} \in \{0,1\}^T$. We write $\boldsymbol{U}_k^n$ for the uniform distribution of Hamming weight k vectors.

For two distributions S and T over the same domain $\mathcal{X}$, the *statistical distance* is defined as

$$\Delta(S,T) \coloneqq \max_{A \subseteq \mathcal{X}} \left| \Pr_{\boldsymbol{x} \sim S}[\boldsymbol{x} \in A] - \Pr_{\boldsymbol{x} \sim T}[\boldsymbol{x} \in A] \right| = \frac{1}{2} \sum_{a \in \mathcal{X}} \left| \Pr_{\boldsymbol{x} \sim S}[\boldsymbol{x} = a] - \Pr_{\boldsymbol{x} \sim T}[\boldsymbol{x} = a] \right|.$$

The statistical distance between two random variables is the statistical distance between their distributions. We say that the distribution S is ε-*close* from the distribution T if $\Delta(S,T) < \varepsilon$. Otherwise, the distributions are ε-*far*.

We say that an input bit $i \in [m]$ *affects* an output bit $j \in [n]$, or equivalently that the output bit j *depends* on the input bit i, if there exist inputs $x, x' \in \{0,1\}^m$, differing only in the ith bit, such that $f(x)_j \neq f(x')_j$. A function $f \colon \{0,1\}^m \to \{0,1\}^n$ is d-*local* if each of its output bits depends only on d input bits. A function $f \colon \{0,1\}^m \to \{0,1\}^n$ has *decision depth* d if each of its output bits can be computed as a depth-d decision tree, i.e. decided with at most d adaptive input bit queries. If a function is d-local, then it has decision depth at most d.

1.2 Sampling Slices

The k-*slice* is the set of all n-bit strings of Hamming weight k. We denote the uniform distribution over the k-slice by $\boldsymbol{U}_k^n$. A simple computation (see Section 3.5) shows that $\boldsymbol{U}_S^n$ and $\boldsymbol{U}_{\max S}^n$ are close in statistical distance whenever $\max S = o(n)$. This means that in order to show the hardness of sampling from symmetric distributions over sublinear-Hamming-weight strings it is sufficient to study $\boldsymbol{U}_k^n$ for $k = o(n)$.

Although in the context of Conjecture 1 it is sufficient to lower bound the locality of a sampler, in the context of slices the more natural complexity measure is *decision depth*. A function $f \colon \{0,1\}^m \to \{0,1\}^n$ is computable by a *decision forest* of depth d if every output bit of f can be computed by a decision tree of depth d, i.e. with at most d *adaptive* queries to the input (in contrast, a d-local function is computed by d *non-adaptive* queries). Viola [26, Lemma 6.4] shows that a decision depth d sampler for $\boldsymbol{U}_k^n$ can be obtained from a depth-d *switching network*. Czumaj [11, Theorem 3.7] proves the existence of such switching networks with $d = O(\log n)$. The following theorem shows that for $k = o(n)$, this construction is almost tight.

▶ **Theorem 2.** *Suppose that $\boldsymbol{U}_k^n$ can be sampled with decision depth d and error η in variation distance.*

1. *For every $\varepsilon > 0$ there exists a constant τ such that $d \le \tau \log(n/k)/\log\log(n/k)$ implies $\eta \ge 1 - \varepsilon$.*
2. *There exists a constant τ for which the following holds. For every $\varepsilon \in (0,1)$, if $k \in [\log_2 n, 2^{\log^{1-\varepsilon} n}]$ and $d \le \tau \log^{\varepsilon} n$, then $\eta = 1 - n^{-\Omega(k)}$. The same bound holds for $k \in [1, \log_2 n)$ and $d \le \tau \log^{\varepsilon} n/\log\log n$.*

Moreover, Item 1 holds for any $\boldsymbol{U}_S^n$ with $\max_{x\in S} x = k$ or $\min_{x\in S} x = n - k$.

The first key observation in our proof of Theorem 2 is that it is sufficient to prove it for $k = 1$, namely:

▶ **Theorem 3.** *There exists a constant $\tau > 0$ such that any distribution sampled with decision depth $\tau \log n/\log\log n$ is $(1 - n^{-\Omega(1)})$-far from $\boldsymbol{U}_1^n$.*

To see why this implies Item 1, observe that the marginal distribution of the first n/k bits of $\boldsymbol{U}_k^n$ is $(1 - 1/e + o(1))$-close to $\boldsymbol{U}_1^{n/k}$. Theorem 3 then implies the distance lower bound $1/e - o(1)$ for sampling $\boldsymbol{U}_k^n$ with depth $\tau \log(n/k)/\log\log(n/k)$. The $1 - \varepsilon$ distance lower bound for every ε is achieved by generalizing Theorem 3 so it applies to distributions of the form "first $\Theta(n/k)$ bits of $\boldsymbol{U}_k^n$" directly.

Item 2 is implied by another reduction from Theorem 3. Suppose we have a depth-d sampler for $\boldsymbol{U}_k^n$. Using this sampler, we can construct a depth-kd sampler for $\boldsymbol{U}_1^{\binom{n}{k}}$ as follows. Identify each output bit with a unique subset of $[n]$ of size k, and assign to it the conjunction of the corresponding k bits in the output of the sampler for $\boldsymbol{U}_k^n$. It is easy to see that the resulting distribution has the same distance to $\boldsymbol{U}_1^{\binom{n}{k}}$ as the initial sampler has to $\boldsymbol{U}_k^n$.

Technique for Proving Theorem 3

The locality of a sampler is the maximum number of input bits that an output bit depends on. The locality is always bounded from above by the decision depth. As a warm-up, let us discuss an $\Omega(\log\log n)$ locality lower bound for sampling $\boldsymbol{U}_1^n$ which is close in spirit to [30, Theorem 3].

The main idea in the locality lower bound is a *hitting set* versus *independent set* dichotomy: for a d-local source, it is easy to see that either there are $\tau 2^d$ independent output bits, or there is a hitting set of input bits of size $\tau d 2^d$ such that every output bit depends on one of the bits in this set. In the former case, we can show that it is very likely that at least two of the independent bits evaluate to 1, since for each output bit its probability to be 1 is at least 2^{-d}.[2] In the latter case, by fixing the hitting set bits in every possible way, we observe that our source is a mixture (convex combination) of $2^{\tau d 2^d}$ many $(d-1)$-local sources. If we show that $(d-1)$-local sources must be $(1-\varepsilon)$-far from the target distribution, then our source is $(1 - \varepsilon 2^{\tau d 2^d})$-far by [30, Corollary 18]. Picking $d = \delta \log\log n$ for small enough δ yields a $1 - o(1)$ lower bound on the distance to the target distribution.

In order to improve this lower bound from $\log\log n$ to $\log n/\log\log n$, we introduce several new ideas.

[2] There is a caveat that some bits might be identically zero, but it is not a real issue, since there cannot be too many of them.

Monotonization. We observe that it is in some sense sufficient to deal with sources where each bit is a *monotone term* in the input bits. The intuitive reason is that the expected number of output bits evaluating to 1 is $2^{-d} \cdot n$ (again, this is not always true, since there are identically zero outputs), so there exists an assignment where that many bits evaluate to 1. By focusing on those bits and replacing them with terms corresponding to the satisfying assignments, we show that it is likely that at least two of these terms evaluate to 1.

Sunflowers. Let us pretend that all of the output bits are monotone terms. For $i \in [n]$, let N_i be the set of inputs mentioned by the term of the ith output bit. We find a sunflower $\mathcal{S} \subseteq \{N_1, \ldots, N_n\}$, i.e. there exists a kernel K such that the intersection of any pair of sets in $\mathcal{S}$ is K. If the kernel is fixed to 1, the output bits in the sunflower become independent, so if the sunflower is large enough, it is likely that at least two of them evaluate to 1. For some small enough $d = \Omega(\sqrt{\log n})$, a large sunflower always exists among any n sets. Moreover, we can cover all but $o(n)$ output bits with sunflowers. Then we have the following dichotomy: either all kernels evaluate to 0, so the source is likely to be identically zero, or at least one kernel evaluates to 1, which makes it very likely that at least two output bits from the corresponding sunflower evaluate to 1.

Using *robust sunflowers* instead of classical sunflowers, we obtain a lower bound of $\Omega(\log n / \log\log n)$ on the decision depth.

1.3 Local Proof Systems

Local proof systems, introduced in [5] and further studied in [20], are defined as follows: a local proof system for a language L is an $\mathbf{NC}^0$-circuit family C_n such that $L \cap \{0,1\}^n$ is exactly the set of all possible outputs of C_n. A language L has a *d-local* proof system if for each n there exists a d-local function whose image is $L \cap \{0,1\}^n$. In relation to the sampling, it can be viewed as follows: the "sampler" needs to match the support of the given distribution exactly, but we do not care about matching the actual probabilities.

The hardness landscape in this setting is different from the setting of sampling distributions. The most notable difference is that the language of strings with Hamming weight divisible by p always has an $O(1)$-local proof system, while we conjecture that sampling from the uniform distribution over this language requires a super-constant locality, unless $p = 1$ or $p = 2$.

Our main contribution is in improving the locality lower bound for another symmetric language: $\textsc{Maj}^{-1}(1) \coloneqq \{x \in \{0,1\}^n \mid |x| \geq n/2\}$. The previous best locality lower bound for proof systems for this language was $\Omega(\log^* n)$ [20], with a very complex proof which we expose in Appendix B. We simplify this proof and improve the locality lower bound to $\Omega(\sqrt{\log n})$, which is only polynomially smaller than the current best upper bound of $O(\log^2 n)$. The key idea is to consider proof systems with bounded locality of both input and output bits. For such proof systems it is easy to derive very strong requirements on locality. These can then be used to count the number of input-output bit pairs where the input affects the output, which yields the output locality lower bounds.

1.4 Switching Networks

The technique behind Theorem 2 breaks down for linear slices $\boldsymbol{U}^n_{\alpha n}$. Does it mean there is a low-depth sampler for such distributions? The only construction of a sampler we have is based on switching networks [11, Theorem 3.7].

A switching network of depth d is a layered graph with d layers and n nodes in each layer. The edges of the switching network do not cross between layers, and in each layer, the edges form a partial matching. A switching network defines a process over permutations of $[n]$: we

start with the identity permutation, and then, for each layer, we toss a coin for each edge in the matching of this layer, and if the coin comes up heads, we transpose the endpoints of the edge.

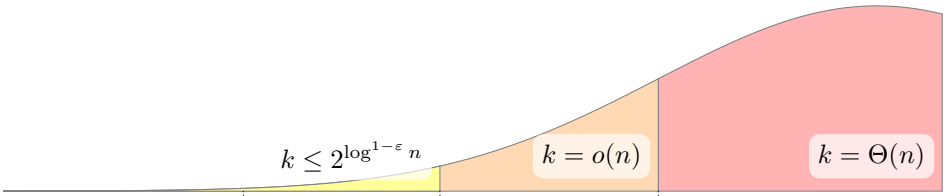
Currently, the best upper bound on the depth of a switching network which produces a distribution close to the uniform distribution over all permutations is $O(\log^2 n)$ [11]. The situation is better if the switching network only needs to shuffle sequences of zeroes and ones: the input is now $1^k 0^{n-k}$. [11, Theorem 3.7] gives a $O(\log n)$ depth upper bound for this case (the construction is randomized), [13] gives an *explicit* lower bound for generating $\boldsymbol{U}_k^n$ for $k \leq \sqrt{n}$.

It is almost immediate that a switching network that samples $\boldsymbol{U}_k^n$ within a non-trivial distance must have depth $\Omega(\log(n/k))$: each input bit of a network of depth d has at most 2^d potential positions that it can take in the output, so if $d = o(\log(n/k))$, the switching network produces a distribution where only $o(n)$ bits have a non-zero probability to have value 1, which is $(1 - o(1))$-far from $\boldsymbol{U}_k^n$.

In Appendix A we show that switching networks that produce a distribution close to $\boldsymbol{U}_{\alpha n}^n$ must have depth $\Omega(\log \log n)$. We use the following properties of samplers that are constructed from switching networks of depth d: the first is that each input bit of such a sampler affects at most 2^d output bits, the second is that the error is one-sided, i.e. such a sampler never outputs a string outside the domain of $\boldsymbol{U}_{\alpha n}^n$. The second property highlights the similarity with local certificates, and indeed our lower bound proof uses similar ideas.

1.5 Further Research

Our results on sampling slices with decision forests taken together with results of Viola [26] are summarized in Figure 1.



Theorem 2	$d = \tilde{\Omega}(\log^{\varepsilon} n)$ $\Delta = 1 - n^{-\Omega(k)}$	$d = \tilde{\Omega}(\log(n/k))$ $\Delta = 1 - \delta$	
[26, Thm 1.6]			$\Delta = 2^{-O(d)} - O(1/n)$

Figure 1 The table above depicts the implications of Theorem 2 and [26, Theorem 1.6] for sampling $\boldsymbol{U}_k^n$ for different k. The plot above it illustrates the size of the corresponding set of bitstrings in the boolean cube.

Here are some important challenges that are left open:

- Give any non-trivial decision depth (or even locality!) lower bound for linear Hamming weight in the constant-error regime. For a non-dyadic α[3] an $\omega(1)$ decision depth lower bound for sampling $\boldsymbol{U}_{\alpha n}^n$ follows (in a not completely straightforward way) from the new separator theorem in [31]: the key idea is to use the fact that biases of all bits of a distribution generated by a decision forest all have form $a/2^d$, so they are $\Theta(2^{-d})$-off from any non-dyadic number. The challenge is to show any non-trivial lower bound for, say $\boldsymbol{U}_{n/2}^n$ or $\boldsymbol{U}_{n/4}^n$, where the bit biases can be matched exactly by a decision forest.

[3] That is, not representable in the form $a/2^t$ for integers a and t, e.g. $1/3$.

- Tighten up the decision depth lower bound for $\boldsymbol{U}_1^n$ to $\Omega(\log n)$. This would immediately yield tight (up to a constant) decision depth lower bounds for all polynomial k.
- Give any locality lower bound for $(\boldsymbol{X}, f(\boldsymbol{X}))$, where $\boldsymbol{X} \sim \{0,1\}^n$ and $f = [(x_1 + \cdots + x_n) \bmod p \geq p/2] \oplus x_1 \oplus \cdots \oplus x_n$, as this would separate quantum and classical $\mathbf{NC}^0$ circuits for sampling, improving on the partial separation of [32].
- Find any non-trivial lower bounds for sampling a uniform vector with Hamming weight divisible by k, for any $k > 2$. This is likely to give insights on the $\mathbf{QNC}^0$ versus $\mathbf{NC}^0$ separation for sampling.
- Determine the optimal depth of a switching network that samples a uniform permutation or a uniform vector in a slice.

2 Tools

In this section, we describe two off-the-shelf tools we use in our proof.

2.1 FKG inequality

▶ **Theorem 4** ([16, 12]). *Suppose that X is a product distribution over $\{0,1\}^n$ (that is, $\Pr[X = x] = \prod_{i\in[n]} \Pr[X_i = x_i]$). Let $A, B \subseteq \{0,1\}^n$ be two monotone events (if $x \in A$ and $x_i \leq y_i$ for all $i \in [n]$ then $y \in A$, and similarly for B). Then*

$$\Pr_{\boldsymbol{x}\sim X}[\boldsymbol{x} \in A \cap B] \geq \Pr_{\boldsymbol{x}\sim X}[\boldsymbol{x} \in A] \cdot \Pr_{\boldsymbol{x}\sim X}[\boldsymbol{x} \in B].$$

2.2 Robust Sunflowers

In this section, we discuss robust sunflowers.

▶ **Definition 5** (Robust sunflower). *Let $0 < \alpha, \beta < 1$ be parameters, let $\mathcal{F}$ be a set system over a finite universe, and let $K \coloneqq \bigcap_{S\in\mathcal{F}} S$ be the intersection of all sets in $\mathcal{F}$, which we refer to as the* kernel. *The family $\mathcal{F}$ is an* (α, β)-robust sunflower *if*

1. *$K \notin \mathcal{F}$;*
2. *$\Pr_{\boldsymbol{R}}[\exists S \in \mathcal{F} \colon S \subseteq \boldsymbol{R} \cup K] \geq 1 - \beta$, where each element of the universe appears in $\boldsymbol{R}$ with probability α independently.*

We can write this condition in the equivalent form $\Pr_{\mathbf{R}}[\exists S \in \mathcal{F} \colon \mathbf{R} \supseteq S \mid \mathbf{R} \supseteq K] \geq 1 - \beta$. A set system is called (α, β)-satisfying *if it is an (α, β)-robust sunflower with an empty kernel.*

Large enough set systems always contain a robust sunflower, as proved by Rossman [24] and improved by later authors.

▶ **Theorem 6** ([1, 4, 23]). *There exists a constant $B > 0$ such that the following holds for all $p, \varepsilon \in (0, 1/2]$. Let $\mathcal{F}$ be a family of sets of size exactly d such that $|\mathcal{F}| \geq (B \log(d/\varepsilon)/p)^d$. Then $\mathcal{F}$ contains a (p, ε)-robust sunflower.*

▶ **Corollary 7.** *There exists a constant $B > 0$ such that the following holds for all $p, \varepsilon \in (0, 1/2]$. Let $\mathcal{F}$ be a family of non-empty sets of size* at most *d such that $|\mathcal{F}| \geq d \cdot (B \log(d/\varepsilon)/p)^d$. Then $\mathcal{F}$ contains a (p, ε)-robust sunflower.*

Proof. Let $d_0 \in [d]$ be the most common size of sets in $\mathcal{F}$. Then the number of sets of size d_0 is at least $(B \log(d/\varepsilon)/p)^d$, which allows us to apply Theorem 6 to these sets. ◀

If we remove a single petal from a robust sunflower, then it remains a robust sunflower (with slightly worse parameters).

▶ **Lemma 8.** *Suppose that $N_1, \dots, N_k \subseteq X$ is a (p, ε)-robust sunflower with kernel K. Then for every i, the sets $N_1, \dots, N_{i-1}, N_{i+1}, \dots, N_k$ form a $(2p, 2\varepsilon)$-robust sunflower with kernel K.*

Proof. Let $\boldsymbol{\tau}$ be distributed over $[3]^{X \setminus K}$ such that $\Pr[\boldsymbol{\tau}_i = 1] = \Pr[\boldsymbol{\tau}_i = 2] = p$, $\Pr[\boldsymbol{\tau}_i = 3] = 1 - 2p$, and the coordinates of $\boldsymbol{\tau}$ are independent. For $\ell \in [3]$, let $\boldsymbol{\tau}^\ell = \{j \in X \mid \boldsymbol{\tau}_j = \ell\} \cup K$. The definition of (p, ε)-robust sunflower implies that for every $\ell \in [2]$ we have $\Pr[\exists j \in [k] \colon \boldsymbol{\tau}^\ell \supseteq N_j] \geq 1 - \varepsilon$. An application of the union bound implies that

$$\Pr\left[\exists j \in [k] \colon \boldsymbol{\tau}^1 \supseteq N_j \wedge \exists j' \in [k] \colon \boldsymbol{\tau}^2 \supseteq N_{j'}\right] \geq 1 - 2\varepsilon.$$

If $j = j'$, then since $\boldsymbol{\tau}^1 \cap \boldsymbol{\tau}^2 = K$, we have $N_j = K$, which is impossible by the definition of a sunflower. Thus $j \neq j'$ whenever the event happens. Let $\boldsymbol{R}$ be a distribution of subsets of X where each element appears in $\boldsymbol{R}$ independently with probability $2p$. Then since $(\boldsymbol{R} | \boldsymbol{R} \supseteq K)$ has the same distribution as $\boldsymbol{\tau}^1 \cup \boldsymbol{\tau}^2$, we have

$$\Pr_{\boldsymbol{R}}\left[\exists j \neq j' \in [k] \colon \boldsymbol{R} \supseteq N_j \wedge \boldsymbol{R} \supseteq N_{j'} \mid \boldsymbol{R} \supseteq K\right] \geq 1 - 2\varepsilon.$$

In particular, for every $i \in [k]$ we have

$$\Pr_{\boldsymbol{R}}\left[\exists j \neq i \in [k] \colon \boldsymbol{R} \supseteq N_j \mid \boldsymbol{R} \supseteq K\right] \geq 1 - 2\varepsilon,$$

and so the sets $N_1, \dots, N_{i-1}, N_{i+1}, N_k$ form a $(2p, 2\varepsilon)$-robust sunflower with kernel K. ◀

Another lemma we use is very similar to the standard connection between robust sunflowers and the classical ones (see e.g. Lemma 1.6 in [1]):

▶ **Lemma 9.** *Suppose that $N_1, \dots, N_m \subseteq X$ is a $(1/(2k), \varepsilon)$-robust sunflower with a kernel K. Then*

$$\Pr_{\boldsymbol{R} \sim 2^X}\left[\exists I \in \binom{[m]}{k} \forall i \in I \colon \boldsymbol{R} \supseteq N_i \middle| \boldsymbol{R} \supseteq K\right] \geq 1 - \varepsilon k.$$

Proof. Let $\tau \sim [2k]^{X \setminus K}$, and $\tau^i \coloneqq \{j \in X \setminus K \mid \tau_j = i\} \cup K$ for $i \in [2k]$. Then $\boldsymbol{\tau}^1 \cup \dots \cup \boldsymbol{\tau}^k$ is distributed equivalently to $(\boldsymbol{R} \mid \boldsymbol{R} \supseteq K)$ where $\boldsymbol{R} \sim 2^X$. On the other hand, by the definition of the $(1/(2k), \varepsilon)$-robust sunflower, for each $i \in [2k]$ we get

$$\Pr[\exists j \in [m] \colon \boldsymbol{\tau}^i \supseteq N_j] \geq 1 - \varepsilon.$$

Since $\boldsymbol{\tau}^i \cap \boldsymbol{\tau}^{i'} = K$ for any $i \neq i' \in [2k]$, the lemma follows by the union bound over $i \in [k]$. ◀

3 Sampling Uniform Hamming Weight k Distributions

In this section we prove the following results mentioned in the introduction, which we restate here for convenience.

▶ **Theorem 2.** *Suppose that $\boldsymbol{U}_k^n$ can be sampled with decision depth d and error η in variation distance.*

1. *For every $\varepsilon > 0$ there exists a constant τ such that $d \leq \tau \log(n/k)/\log\log(n/k)$ implies $\eta \geq 1 - \varepsilon$.*

2. *There exists a constant τ for which the following holds. For every $\varepsilon \in (0,1)$, if $k \in [\log_2 n, 2^{\log^{1-\varepsilon} n}]$ and $d \leq \tau \log^{\varepsilon} n$, then $\eta = 1 - n^{-\Omega(k)}$. The same bound holds for $k \in [1, \log_2 n)$ and $d \leq \tau \log^{\varepsilon} n / \log\log n$.*

Moreover, Item 1 holds for any $\boldsymbol{U}_S^n$ with $\max_{x \in S} x = k$ or $\min_{x \in S} x = n - k$.

▶ **Theorem 3.** *There exists a constant $\tau > 0$ such that any distribution sampled with decision depth $\tau \log n / \log\log n$ is $(1 - n^{-\Omega(1)})$-far from $\boldsymbol{U}_1^n$.*

We first prove Theorem 3, in Section 3.1. We then prove Item 2 of Theorem 2 in Section 3.3, and Item 1 of the Theorem in Section 3.4. We prove the "moreover" part in Section 3.5.

3.1 Proof of Theorem 3

We prove a more general result which immediately yields Theorem 3.

First let us sketch a proof of Theorem 3 for d-local functions. Suppose that $\Delta(\boldsymbol{U}_1^n, \boldsymbol{X}) \leq 1 - \eta$. Call a coordinate i *good* if $\Pr[\boldsymbol{X} = e_i] \geq 1/n^2$. Since $\Pr[U_1 = e_i] = 1/n$, many coordinates are good: at least $\Omega(\eta n)$.

Let $\boldsymbol{Y} \sim \{0,1\}^m$ denote the random input bits. Each X_i depends on some subset $N_i \subseteq [m]$ of coordinates of size at most $d = \tau \log n / \log\log n$, say $\boldsymbol{X}_i = f_i(\boldsymbol{Y}_{N_i})$.

For each good coordinate i, we choose an assignment $\alpha_i \in f_i^{-1}(1)$ which maximizes the conditional probability $\Pr[\boldsymbol{X} = e_i \mid \boldsymbol{Y}_{N_i} = \alpha_i]$, that is, the probability that if $\boldsymbol{Y}_{N_i} = \alpha_i$ then all other output bits are 0. This probability is at least $1/(2^d n^2) = \Omega(1/n^3)$.

The assignments α_i do not necessarily agree with each other. However, a random assignment ρ to $\boldsymbol{Y}$ agrees with at least $\Omega(\eta n / 2^d) = \Omega(\eta n^{1-o(1)})$ of them. Let T consists of the domains of the assignments α_i which agree with ρ. These domains are distinct since $N_i = N_j$ implies $\alpha_i = \alpha_j$ and hence that $\Pr[\boldsymbol{X} = e_i \mid \boldsymbol{Y}_{N_i} = \alpha_i] = 0$. The choice of d guarantees that T supports a $(1/4, \varepsilon)$-robust sunflower $\mathcal{S}$, for any ε which is inverse-polynomial in n. Let K be the kernel of $\mathcal{S}$.

If we remove any single petal i from $\mathcal{S}$ then by Lemma 8 the result is a $(1/2, 2\varepsilon)$-robust sunflower, and so given that $\boldsymbol{Y}_K$ agrees with ρ, the probability that $\boldsymbol{X}_j = 1$ for some $j \neq i$ is at least $1 - 2\varepsilon$. If we replace the condition with "$\boldsymbol{Y}_{N_i}$ agrees with ρ" (and so with α_i), then intuitively, the probability can only increase, and this can be formalized using the FKG inequality (Theorem 4). By definition of α_i, this means that $\Pr[\boldsymbol{X} = e_i] \leq |f_i^{-1}(1)| 2\varepsilon \leq 2^{d+1}\varepsilon$. Choosing $\varepsilon = 1/2^{d+1}n^2$ shows that i is not good, and we reach a contradiction.

We move on to prove the generalization of Theorem 3.

▶ **Theorem 10.** *Let $\boldsymbol{Y} \sim \{0,1\}^m$ be the input bits of the n-bit source $\boldsymbol{X}$. Suppose that every bit of $\boldsymbol{X}$ is computed as a DNF of bits of $\boldsymbol{Y}$ of size at most s and width at most d. For every $\kappa \in \mathbb{R}$ there exists a constant τ such that for $d = \tau \log n / \log\log n$ and $s \leq \kappa n^{\kappa}$, we have $\Delta(\boldsymbol{X}, \boldsymbol{U}_1^n) = 1 - \eta = 1 - n^{-\Omega(1)}$.*

This implies Theorem 3 since the output of a decision tree of depth d can be represented as a DNF of size at most 2^d and width at most d. In our case $d = o(\log n)$ and so $2^d \leq n$ (for large enough n).

Proof. We say that an output bit $i \in [n]$ is *good* if $\Pr[\boldsymbol{X} = e_i] \geq 1/n^2$. Let $G \subseteq [n]$ be the set of all good bits, and let $\overline{G} = [n] \setminus G$. Let us estimate the size of G: $\Pr[\boldsymbol{X} \in \{e_i \mid i \notin G\}] \leq |\overline{G}|/n^2$, but $\Pr[\boldsymbol{U}_1^n \in \{e_i \mid i \notin G\}] = |\overline{G}|/n$), so $|\overline{G}| \cdot (1/n - 1/n^2) \leq 1 - \eta$, which yields $|G| = \Omega(\eta n)$.

For each $i \in G$, since each bit of $\boldsymbol{X}$ is represented as a DNF we have $\boldsymbol{X}_i = \bigvee_{j \in [s_i]} [\boldsymbol{Y}_{N_i^j} = \alpha_i^j]$, where $N_i^1, \ldots, N_i^{s_i} \subseteq [m]$ are sets, $\alpha_i^j \in \{0,1\}^{N_i^j}$ are truth assignments, and $s_i \leq s$. By the law of total probability we have

$$\Pr[\boldsymbol{X} = e_i] \leq \sum_{j \in [s_i]} \Pr[\boldsymbol{X} = e_i \wedge \boldsymbol{Y}_{N_i^j} = \alpha_i^j] \leq s \Pr[\boldsymbol{X} = e_i \wedge \boldsymbol{Y}_{N_i^{\max}} = \alpha_i^{\max}],$$

where $N_i^{\max}$ and $\alpha_i^{\max}$ correspond to the term in the DNF maximizing the probability $\Pr[\boldsymbol{X} = e_i \wedge \boldsymbol{Y}_{N_i^j} = \alpha_i^j]$.

Consider the expected number of good output bits such that $\boldsymbol{Y}_{N_i^{\max}} = \alpha_i^{\max}$:

$$\mathbb{E}\left[\sum_{i \in G} [\boldsymbol{Y}_{N_i^{\max}} = \alpha_i^{\max}]\right] \geq |G| 2^{-d}.$$

Hence there exists an assignment ρ to the input bits such that for at least $|G|2^{-d}$ good output bits, we have $\rho_{N_i^{\max}} = \alpha_i^{\max}$. Let $T \subseteq G$ be the set of those output bits. If $i, j \in T$ then $N_i^{\max} \neq N_j^{\max}$, since otherwise $\boldsymbol{Y}_{N_i^{\max}} = \alpha_i^{\max}$ implies that also $\boldsymbol{Y}_{N_j^{\max}} = \alpha_j^{\max}$ and so $\boldsymbol{X} \neq e_i$, and so $\Pr[\boldsymbol{X} = e_i] = 0$, contradicting $i \in G$. Observe moreover that none of the sets N_i for $i \in T$ is empty, since otherwise $|T| = 1$ and we get an immediate contradiction with the size of G for any $d = o(\log n)$.

Case 1. $|T| < d(4B \log(d/\varepsilon))^d$. In this case, we immediately get the lower bound on δ. Indeed, the inequality $|G|2^{-d} \leq |T| < d(4B \log(d/\varepsilon))^d$ implies $|G| \leq d(8B \log(d/\varepsilon))^d$, which together with $|G| = \Omega(\eta n)$ yields $\eta \leq d(8B \log(d/\varepsilon))^d / n$. If ε is inverse polynomial in n, then for small enough τ we get $\eta = n^{-\Omega(1)}$ with $d = \tau \log n / \log \log n$.

Case 2. $|T| \geq d(4B \log(d/\varepsilon))^d$. Then by Corollary 7 there exists a $(1/4, \varepsilon)$-robust sunflower formed by the sets $N_{t_1}^{\max}, \ldots, N_{t_k}^{\max}$ for $\{t_1, \ldots, t_k\} \subseteq T$ (recall the sets $N_i^{\max}$ for $i \in T$ are all distinct, and none of them is empty). Let K denote the kernel of this sunflower. Consider an arbitrary petal t_i of this sunflower. By Lemma 8 we have that $\{N_i^{\max}\}_{i \in T \setminus \{t_i\}}$ is a $(1/2, 2\varepsilon)$-robust sunflower. Let $\boldsymbol{U}$ be the set of indices such that $\boldsymbol{Y}_k = \rho_k$. Then

$$\begin{aligned}
&\Pr[\boldsymbol{X}_j = 1 \text{ for some } j \in T \setminus \{t_i\} \mid \boldsymbol{Y}_{N_{t_i}^{\max}} = \rho_{N_{t_i}^{\max}}] \geq \\
&\Pr[\boldsymbol{U} \supseteq N_j^{\max} \text{ for some } j \in T \setminus \{t_i\} \mid \boldsymbol{U} \supseteq N_{t_i}^{\max}] = \\
&\frac{\Pr[\boldsymbol{U} \supseteq N_{t_i}^{\max} \text{ and } \boldsymbol{U} \supseteq N_j^{\max} \text{ for some } j \in T \setminus \{t_i\} \mid \boldsymbol{U} \supseteq K]}{\Pr[\boldsymbol{U} \supseteq N_{t_i}^{\max} \mid \boldsymbol{U} \supseteq K]} \geq \qquad \text{Theorem 4} \\
&\Pr[\boldsymbol{U} \supseteq N_j^{\max} \text{ for some } j \in T \setminus \{t_i\} \mid \boldsymbol{U} \supseteq K] \geq \\
&1 - 2\varepsilon,
\end{aligned}$$

where the last inequality is due to the definition of a $(1/2, 2\varepsilon)$-robust sunflower. Recall that by the choice of T, we have $\rho_{N_{t_i}^{\max}} = \alpha_{t_i}^{\max}$. Therefore

$$\Pr[\boldsymbol{X} \neq e_{t_i} \mid \boldsymbol{Y}_{N_{t_i}^{\max}} = \alpha_{t_i}^{\max}] = \Pr[\boldsymbol{X}_j = 1 \text{ for some } j \in T \setminus \{t_i\} \mid \boldsymbol{Y}_{N_{t_i}^{\max}} = \rho_{N_{t_i}^{\max}}] \geq 1 - 2\varepsilon.$$

Thus $\Pr[\boldsymbol{X} = e_{t_i} \wedge \boldsymbol{Y}^{N_{t_i}^{\max}} = \alpha_{t_i}^{\max}] \leq 2\varepsilon$. By the choice of $\alpha_{t_i}^{\max}$, $\Pr[\boldsymbol{X} = e_{t_i}] \leq s \cdot 2\varepsilon$. Picking $\varepsilon < 1/(2sn^2)$, which is inverse polynomial in n as required in Case 1, we get that $\Pr[\boldsymbol{X} = e_{t_i}] < 1/n^2$, so t_i is bad, which contradicts the choice of T. ◀

3.2 A Generalized Version of Theorem 10

In this section, we generalize Theorem 10 so it can be used to prove Item 1 of Theorem 2. The proof follows the same path as the proof of Theorem 10, we decided to include both proofs for simplicity.

▶ **Theorem 11.** *Let $\boldsymbol{Y} \sim \{0,1\}^m$ be the input bits of the n-bit source $\boldsymbol{X}$. Suppose that every bit of $\boldsymbol{X}$ is computed as a DNF of bits of $\boldsymbol{Y}$ of size at most s and width at most d. Let $t \in [n]$ be a parameter, $\alpha(n)$ be a function and let $\boldsymbol{F}$ be distributed over $\{0,1\}^n$ and have the following properties:*

- *For every set $T \subseteq [n]$ such that $T \geq n/2$ we have $\Pr[\boldsymbol{F}_T = 0^T] \leq \alpha(n)$;*
- $\Pr[|\boldsymbol{F}| > t] \leq \alpha(n)$.

If $2d \cdot t \cdot (40Bt \log n)^d \leq n$ then $\Delta(\boldsymbol{X}, \boldsymbol{F}) \geq 1 - 2\alpha(n) - 1/2n$. Here B is the constant from Corollary 7.

Proof. We say that an output bit $i \in [n]$ is *good* if $\Pr[\boldsymbol{X}_i = 1 \wedge |\boldsymbol{X}| \leq t] \geq 1/n^2$. Let G be the set of good bits and let $\overline{G} \coloneqq [n] \setminus G$. Suppose that $|G| \leq n/2$. Then by the conditions on $\boldsymbol{F}$ we have $\Pr[\boldsymbol{F}_{\overline{G}} = 0^{\overline{G}}] \leq \alpha(n)$. On the other hand $\Pr[\boldsymbol{X}_{\overline{G}} \neq 0^{\overline{G}} \wedge |\boldsymbol{X}| \leq t] < |\overline{G}|/n^2 < 1/2n$. Then $\Delta(\boldsymbol{X}, \boldsymbol{F}) \geq 1 - 2\alpha(n) - 1/2n$, as required. In the rest of the proof, we derive a contradiction with $|G| \geq n/2$.

As in the proof of Theorem 10, we pick the likeliest term in the DNF representation of each of the output bits. For each $i \in G$, since each bit of $\boldsymbol{X}$ is represented as a DNF, we have $\boldsymbol{X}_i = \bigvee_{j \in [s_i]} [\boldsymbol{Y}_{N_i^j} = \alpha_i^j]$, where $N_i^1, \ldots, N_i^{s_i} \subseteq [m]$ are sets, $\alpha_i^j \in \{0,1\}^{N_i^j}$ are truth assignments, and $s_i \leq s$. By the law of total probability, we have

$$\Pr[\boldsymbol{X}_i = 1 \wedge |\boldsymbol{X}| \leq t] \leq \sum_{j \in [s_i]} \Pr[|\boldsymbol{X}| \leq t \wedge \boldsymbol{Y}_{N_i^j} = \alpha_i^j] \leq s \Pr[|\boldsymbol{X}| \leq t \wedge \boldsymbol{Y}_{N_i^{\max}} = \alpha_i^{\max}],$$

where $N_i^{\max}$ and $\alpha_i^{\max}$ correspond to the term in the DNF maximizing the probability $\Pr[|\boldsymbol{X}| \leq t \wedge \boldsymbol{Y}_{N_i^j} = \alpha_i^j]$. The expected number of good output bits with $\boldsymbol{Y}_{N_i^{\max}} = \alpha_i^{\max}$ is at least $2^{-d}|G|$, so there exists an assignment ρ to the input bits such that $\rho_{N_i^{\max}} = \alpha_i^{\max}$ for at least $|G|2^{-d}$ good output bits. Let $T \subseteq G$ be the set of these output bits.

Let us estimate how many distinct elements are in the set $\mathcal{N} \coloneqq \{N_i^{\max} \mid i \in T\}$. Suppose there exist $i_1, \ldots, i_{t+1} \in T$ such that $N_{i_1}^{\max} = \cdots = N_{i_{t+1}}^{\max}$. Then, by the definition of ρ, we have $\alpha_{i_1}^{\max} = \cdots = \alpha_{i_{t+1}}^{\max}$ as well. Thus $\boldsymbol{Y}_{N_{i_1}^{\max}} = \alpha_{i_1}^{\max}$ implies that for every $j \in [t+1]$ we have $\boldsymbol{Y}_{N_{i_j}^{\max}} = \alpha_{i_j}^{\max}$, which in turn implies that $|\boldsymbol{X}| \geq t+1$, and so $\Pr[\boldsymbol{Y}_{N_{i_1}^{\max}} = \alpha_{i_1}^{\max} \wedge |\boldsymbol{X}| \leq t] = 0$, which contradicts that i_1 is good. Hence $|\mathcal{N}| \geq |T|/t \geq |G|2^{-d}/t \geq 2^{-d}n/2t$.

Let $\varepsilon > n^{-5}$ be a parameter to be chosen later. By the condition on n we have $|\mathcal{N}| \geq 2^{-d}n/2t \geq d \cdot (4Bt \log(d/\varepsilon))^d$, and so $\mathcal{N}$ contains a $(1/(4t), \varepsilon)$-robust sunflower $\mathcal{S}$. Let K denote the kernel of this sunflower.

Fix an arbitrary petal $p \in \mathcal{N}$. Then $\mathcal{N} \setminus \{p\}$ is a $(1/(2t), 2\varepsilon)$-robust sunflower by Lemma 8. Now by Lemma 9 we have

$$\Pr_{\boldsymbol{R} \sim 2^{[m]}} [\text{There are } t \text{ distinct petals of } \mathcal{N} \setminus \{p\} \text{ contained in } \boldsymbol{R} \mid \boldsymbol{R} \supseteq K] \geq 1 - 2t\varepsilon.$$

Let $P \subseteq T$ be the indices of the output bits corresponding to the elements of $\mathcal{N} \setminus \{p\}$, let i be the index of the output bit corresponding to the petal p, and let $\boldsymbol{U}$ be the set of indices of input bits such that $\boldsymbol{Y}_t = \rho_t$. Then

$$\begin{aligned}
&\Pr[|\boldsymbol{X}| > t \mid \boldsymbol{Y}_{N_i^{\max}} = \rho_{N_i^{\max}}] = \\
&\Pr[|\boldsymbol{X}_{[n]\setminus\{i\}}| \geq t \mid \boldsymbol{Y}_{N_i^{\max}} = \rho_{N_i^{\max}}] \geq \\
&\Pr\left[\sum_{j\in P}[\boldsymbol{U} \supseteq N_j^{\max}] \geq t \,\middle|\, \boldsymbol{U} \supseteq N_i^{\max}\right] = \\
&\frac{\Pr[\sum_{j\in P\cup\{i\}}[\boldsymbol{U} \supseteq N_j^{\max}] \geq t \mid \boldsymbol{U} \supseteq K]}{\Pr[\boldsymbol{U} \supseteq N_i^{\max} \mid \boldsymbol{U} \supseteq K]} \geq \text{ Theorem 4} \\
&\Pr\left[\sum_{j\in P}[\boldsymbol{U} \supseteq N_j^{\max}] \geq t \,\middle|\, \boldsymbol{U} \supseteq K\right] \geq \\
&1 - 2t\varepsilon.
\end{aligned}$$

Recall that by the choice of T we have $\rho_{N_i^{\max}} = \alpha_i^{\max}$, hence $\Pr[|\boldsymbol{X}| > t \mid \boldsymbol{Y}_{N_i^{\max}} = \alpha_i^{\max}] \geq 1 - 2t\varepsilon$. Thus

$$\Pr[|\boldsymbol{X}| \leq t \wedge \boldsymbol{X}_i = 1] \leq s \cdot \Pr[|\boldsymbol{X}| \leq t \wedge \boldsymbol{Y}_{N_i^{\max}} = \alpha_i^{\max}] \leq 2st \cdot \varepsilon.$$

Picking $\varepsilon = 1/(4stn^2) \geq n^{-5}$, we get a contradiction with i being good. ◀

3.3 Subpolynomial Weights

Although Theorem 11 implies Item 2 of Theorem 2, we give a simpler proof via a reduction from $\boldsymbol{U}_1^n$.

▶ **Lemma 12.** *Let $S \subseteq \{0,1\}^n$ and let $\boldsymbol{S} \sim S$. Suppose that $\boldsymbol{S}$ can be sampled with a depth-d decision forest with error η. Assume furthermore that for each $s \in S$ there exists a decision tree T_s of depth k that accepts s and does not accept any of $S \setminus \{s\}$. Then there exists a decision depth-kd sampler for $\boldsymbol{U}_1^{|S|}$ with error η.*

Proof. Let $\boldsymbol{Y}$ be the distribution sampled by the sampler for $\boldsymbol{S}$. For each output bit of our sampler for $\boldsymbol{U}_1^{|S|}$ we take a unique element $s \in S$ and implement each of the queries of T_s via the query to the bits of $\boldsymbol{Y}$ (which makes at most d queries to the input bits). This results in a kd-deep decision tree T'_s. Let $\boldsymbol{X}$ be the sampled distribution. Then

$$\begin{aligned}
\Delta(\boldsymbol{X}, \boldsymbol{U}_1^{|S|}) &= \frac{1}{2}\left(\Pr[w(\boldsymbol{X}) \neq 1] + \sum_{s\in S}\left|\Pr[\boldsymbol{X} = e_s] - 1/|S|\right|\right) \\
&= \frac{1}{2}\left(\Pr[\boldsymbol{Y} \notin S] + \sum_{s\in S}\left|\Pr[\boldsymbol{Y} = s] - 1/|S|\right|\right) = \Delta(\boldsymbol{Y}, \boldsymbol{S}) = \eta.
\end{aligned}$$ ◀

▶ **Corollary 13.** *For some constant $\tau' > 0$ and every $\varepsilon \in (0,1)$ every $(\tau' \log^{\varepsilon} n)$-decision depth sampler outputs a distribution $(1 - n^{-\Omega(k)})$-far from $\boldsymbol{U}_k^n$ for $k \in [\log n, 2^{\log^{1-\varepsilon} n}]$. If $k < \log_2 n$, this holds for every $(\tau' \log^{\varepsilon} n / \log\log n)$-local sampler.*

Proof. The decision tree that queries all the elements of a k-size set and accepts iff all of them are 1 satisfies the condition in Lemma 12. If we had a $(1-\delta)$-error sampler for $\boldsymbol{U}_k^n$, we would get a $(1-\delta)$-error sampler for $\boldsymbol{U}_1^{\binom{n}{k}}$ with decision depth kd, which by Theorem 3 yields that $\delta = \binom{n}{k}^{-\Omega(1)} = n^{-\Omega(k)}$ whenever $kd \leq \tau \log\binom{n}{k} / \log\log\binom{n}{k}$. Since $\log\binom{n}{k} = \Theta(k\log(n/k))$, we get $d = \Omega(\log(n/k)/(\log k + \log\log n))$. ◀

3.4 Sublinear Weights

In this section, we prove Item 1 of Theorem 2.

▶ **Lemma 14.** *Suppose $\boldsymbol{X}$ is sampled with decision depth d. Then for every small enough $\varepsilon > 0$ there exists a constant τ such that if $d \leq \tau \log(n/k)/\log\log(n/k)$ then $\Delta(\boldsymbol{X}, \boldsymbol{U}_k^n) \geq 1 - 2\varepsilon - \frac{1}{2n}$.*

Proof. Consider the first $\ell := \varepsilon^{-1} \cdot n/k$ bits of the sampler: $\boldsymbol{X}_{\leq \ell} := \boldsymbol{X}_1, \ldots, \boldsymbol{X}_\ell$. Let $\boldsymbol{Y}$ be the first ℓ bits of the distribution $\boldsymbol{U}_k^n$. We show that $\boldsymbol{Y}$ satisfies the conditions of Theorem 11 for $t = \varepsilon^{-2}$ and $\alpha(n) = \varepsilon$. First, we have

$$\Pr[|\boldsymbol{Y}| > t] = \Pr\left[|\boldsymbol{Y}| > \varepsilon^{-1}\mathbf{E}[|\boldsymbol{Y}|]\right] < \varepsilon.$$

Now let T be any subset of $[\ell]$ of size at least $\ell/2$. Then

$$\Pr[\boldsymbol{Y}_T = 0^T] = \binom{n-\ell/2}{k}\binom{n}{k}^{-1} = \prod_{i=0}^{k-1} \frac{n-i-\ell/2}{n-i} \leq \left(1 - \frac{\ell}{2n}\right)^k = \left(1 - \frac{1}{2\varepsilon k}\right)^k \leq e^{-2\varepsilon^{-1}} < \varepsilon.$$

Here we assumed $k < n/2$, since otherwise the lemma is trivially true.

Applying Theorem 11, for small enough τ (which depends on ε) we have $\Delta(\boldsymbol{X}_{\leq \ell}, \boldsymbol{Y}) = 1 - 2\varepsilon - 1/2n$. To finish the proof, observe that $\Delta(\boldsymbol{X}, \boldsymbol{U}_k^n) \geq \Delta(\boldsymbol{X}_{\leq \ell}, \boldsymbol{Y})$, since the random variables on the RHS are the marginals of the variables on the LHS. ◀

3.5 Unions of Slices

In this section, we prove the "moreover" part of Theorem 2 by observing that the distribution $\boldsymbol{U}_S^n$ is close to $\boldsymbol{U}_{\max_{x\in S} x}^n$ as long as $\max_{x\in S} x = o(n)$.

▶ **Proposition 15.** *Let $k = o(n)$ and suppose that $S \subseteq \{0, 1, \ldots, k\}$ with $k \in S$. Then we have $\Delta(\boldsymbol{U}_k^n, \boldsymbol{U}_S^n) = o(1)$.*

Proof. We use the notation $\binom{n}{S} := \sum_{i\in S}\binom{n}{i}$.

$$\begin{aligned}
\Delta(\boldsymbol{U}_k^n, \boldsymbol{U}_S^n) &= \frac{1}{2}\binom{n}{S \setminus \{k\}}\binom{n}{S}^{-1} + \frac{1}{2}\left(1 - \binom{n}{k}\binom{n}{S}^{-1}\right) \\
&= \binom{n}{S \setminus \{k\}}\binom{n}{S}^{-1} \\
&\leq \binom{n}{k}^{-1} \sum_{i \in S\setminus\{k\}} \binom{n}{i} \\
&\leq \sum_{i=0}^{k-1} \left(\frac{k}{n-i}\right)^{i-k} = \Theta(k/n).
\end{aligned}$$

◀

The case of S with $n - \min_{x\in S} x = o(n)$ reduced to the case where $\max_{x\in S} x = o(n)$ by observing that flipping all output bits can be done with no increase in the decision depth of the sampler.

4 Local Certificates

In this section, we explore the power of local proof systems. Section 4.1 gives an example of a language that requires locality $\Omega(n)$, which is inspired by a similar lower bound in the context of sampling [22]. Section 4.2 then gives our main result, a lower bound on the locality of proof systems for $\textsc{Maj}^{-1}(1)$.

4.1 Error-correcting codes

In this section, we show that a good error-correcting code requires a proof system of a linear locality. This showcases the simple counting technique that we also use for our majority lower bound.

▶ **Proposition 16.** *Let $C \subseteq \{0,1\}^n$ be a good code, that is, $|C| \geq 2^{\alpha n}$ and for every $x \neq y \in C$, the Hamming distance between x and y is at least βn, where α and β are constants in $(0,1)$.*

If $f\colon \{0,1\}^m \to \{0,1\}^n$ is a d-local function and $f(\{0,1\}^m) = C$ then $d \geq \alpha\beta n$

Proof. We may assume w.l.o.g. that all input bits of f affect some output bits. Take an arbitrary input bit $i \in [m]$ and an output bit $j \in [n]$ that depends on i. Then take $x, x' \in \{0,1\}^m$ that differ only in the ith coordinate and such that $f(x)_j \neq f(x')_j$. Since $f(x)$ and $f(x')$ are two distinct codewords of C, they must be at Hamming distance at least βn, hence i must affect at least βn output bits, as x and x' only differ in i. Thus every bit affects at least βn output bits.

There are at least αn input bits since $|C| = 2^{\alpha n}$. Therefore there are $\alpha\beta n^2$ input-output pairs in which the input bit affects the output bit. On the other hand, there are at most dn such pairs, hence $d \geq \alpha\beta n$. ◀

4.2 Majority

Let $\textsc{Maj}_n$ be the the set $\{x \in \{0,1\}^n \mid |x| \geq n/2\}$. First, let us give a simple upper bound on locality which is implicit in [20].

▶ **Proposition 17** (essentially Theorem 3.9 and Corollary 3.10 in [20])**.** *There exists an $O(\log^2 n)$-local function $f\colon \{0,1\}^* \to \{0,1\}^n$ such that $f(\{0,1\}^*) = \textsc{Maj}_n^{-1}(1)$.*

Proof. For simplicity, suppose that n is odd. Construct a binary tree whose root is the interval $[1,n]$, whose leaves are the singletons $\{1\},\dots,\{n\}$, and in which each internal node $[\ell,r]$ has two children $[\ell,c],[c+1,r]$, where $c = \lfloor(\ell+r)/2\rfloor$. We can construct such a tree whose depth is $O(\log n)$. For each interval $[\ell,r]$ in the tree we will have a label $w(\ell,r)$ whose value ranges from 0 to $r-\ell+1$, which is supposed to indicate $x_\ell + \cdots + x_r$ (where $x_1,\dots,x_n$ is the output). We implement the variables using $O(\log n)$ input bits.

An internal node $[\ell,r]$ with children $[\ell,c],[c+1,r]$ is *consistent* if $w(\ell,r) = w(\ell,c) + w(c+1,r)$. In addition, if the internal node is the root $[1,n]$, we require $w(1,n) \geq (n+1)/2$. Each position $i \in \{1,\dots,n\}$ corresponds to the leaf $w(i,i)$, which has $O(\log n)$ ancestors. We say that position i is *good* if all its non-leaf ancestors are consistent. The i'th output is $w(i,i)$ if i is good, and 1 otherwise. Since each $w(\ell,r)$ is encoded using $O(\log n)$ bits, this system has locality $O(\log^2 n)$.

Every vector $x_1,\dots,x_n$ of weight at least $(n+1)/2$ can be generated using this system by taking $w(\ell,r) = x_\ell + \cdots + x_r$. In the other direction, consider any assignment of weights to the tree. If the root is inconsistent, then the output is $1,\dots,1$, so we can assume that the root is consistent. Prune the tree by removing all children of inconsistent nodes. If $[\ell,r]$ is any node in the pruned tree then either $\ell = r$ and $x_\ell = w(\ell,r)$, or $\ell < r$ and $x_\ell + \cdots + x_r = r - \ell + 1 \geq w(\ell,r)$. It follows that $x_1 + \cdots + x_n \geq w(1,n) \geq (n+1)/2$. ◀

The $\Omega(\log^* n)$ locality lower bound in [5] is inspired by the following observation:

▶ **Proposition 18.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be such that every output bit is a function of at most c input bits, and every input bit affects at most d output bits. Suppose that $cd \leq (n+1)/2$. Then $f(\{0,1\}^m) \neq \textsc{Maj}^{-1}(1)$.*

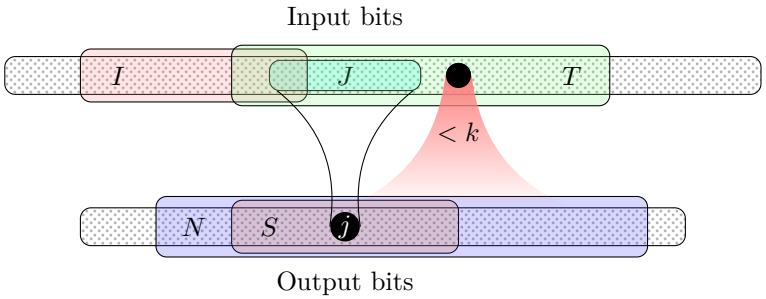


Figure 2 I is the set of k-influential inputs bits, and S is the given set. The set T consists of all inputs bits affecting S, and the set N consists of all bits influenced by $T \setminus I$.

Proof. Let $i \in [n]$ be an arbitrary output bit. There are at most cd many output bits in the "neighborhood" $N(i)$ of i, which is the set of outputs that share an input bit with i. If $|N(i)| \leq (n+1)/2$ then we can find an output y of weight $(n+1)/2$ such that $y_{N(i)} = 1^{N(i)}$. Suppose that y is generated by the input x. There must be some setting to the inputs of i which sets it to zero (since there is a valid output z with $z_i = 0$). If we modify x using this setting then the new output z agrees with y outside of $N(i)$, and furthermore $z_i = 0$. Since $y_{N(i)} = 1^{N(i)}$, it follows that $|z| < |y|$, which is impossible, since y had the smallest possible weight. ◀

One of the steps in our proof (namely, Lemma 20) is essentially an adaptation of the proof of Proposition 18 for the sources with influential input bits.

We give a simplified exposition of their proof in Appendix B. Our own lower bound is contained in the following theorem.

▶ **Theorem 19.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function such that $f(\{0,1\}^m) = MAJ_n^{-1}(1)$, where n is odd. Then $d = \Omega(\sqrt{\log n})$.*

Proof. We say that an input bit $i \in [m]$ is *k-influential* if at least k output bits of f depend on it. We are going to show that there are $\Omega(n/(dk))$ many k-influential bits for every k. Then the number of input-output bit pairs where the output depends on the input is at least $\sum_{k \in [n]} cn/(dk) = \Omega(n \log n/d)$. On the other hand, there are at most nd such pairs since f is d-local. Therefore $d = \Omega(\sqrt{\log n})$.

It remains to show the lower bound on the number of k-influential bits. This is done by combining the following two lemmas.

▶ **Lemma 20.** *Let I be the set of all k-influential input bits. Then for every set of output bits S of size at most $n/(4kd)$ there exists an assignment ρ to I such that*

1. *All bits in S are fixed to 1 by ρ, i.e. for every total extension ρ' of ρ we have $f(\rho')_S = 1^S$.*
2. *ρ fixes to 1 at most $(n+1)/2$ output bits.*

Proof. Fix a set S of size at most $n/(4kd)$.

Let T be the set of input bits that affect S, so $|T| \leq d|S|$. Let N be the set of all bits influenced by $T \setminus I$. See Figure 2 for a pictorial representation of these definitions.

Since I contains all k-influential input bits, $|N| \leq kd|S|$. Let ρ' be a total assignment such that $f(\rho')_N = 1^N$ and $|f(\rho')| = (n+1)/2$. This is possible since $|N| \leq n/4$ by the statement of the lemma. Let $\rho := \rho'_I$.

Since $|f(\rho')| = (n+1)/2$, in particular ρ fixes to 1 at most $(n+1)/2$ bits. We claim that ρ fixes all bits in S to 1. Suppose for the sake of contradiction that it doesn't fix to 1 the bit $j \in S$. Let J be the set of input bits affecting j. Let ρ'' be a total assignment consistent with ρ' everywhere except $J \setminus I$ such that $f(\rho'')_j = 0$. Observe that $f(\rho'')_{[n]\setminus N} = f(\rho')_{[n]\setminus N}$, hence $f(\rho'') \leq f(\rho')$ coordinate-wise. Since $f(\rho'')_j = 0$ and $f(\rho')_j = 1$, we have $|f(\rho'')| < n/2$, which contradicts the fact that the image of f is $\textsc{Maj}_n^{-1}(1)$. ◀

▶ **Lemma 21.** *Let I be an arbitrary set of input bits of size at most $n/20$. Then there exists a set S of output bits such that $|S| = O(|I|)$ and for every assignment ρ to I that fixes to 1 at most $(n+1)/2$ output bits, there exists a bit in S that is not fixed to 1 by ρ.*

Proof. Let $\rho_1, \ldots, \rho_K$ be all assignments to I that fix at most $(n+1)/2$ output bits. Denote by $U_1, \ldots, U_K \subseteq [n]$ the sets of bits that are not fixed by $\rho_1, \ldots, \rho_K$, respectively, so that $|U_1|, \ldots, |U_K| \geq (n-1)/2 \geq n/3$. Then

$$K\frac{n}{3} \leq \sum_{i=1}^{K} |U_i| = \sum_{j\in[n]} |\{i \in [K] \mid U_i \ni j\}|.$$

Hence there exists j such that $|\{i \in [K] \mid U_i \ni j\}| \geq K/3$. Let $S_1 \coloneqq \{j\}$, and continue this process for the set of bits $[n] \setminus \{j\}$ and the set of assignments $\{\rho_i \colon U_i \not\ni j\}$. Suppose the previous iteration yields a set $S_k \subseteq [n]$ of size k and a set of indices $T_k \subseteq [K]$. Then let $U_i' \coloneqq U_i \setminus S_k$ for $i \in T_k$. Then $|U_1'|, \ldots, |U_K'| \geq (n-1)/2 - k \geq n/3$, where the last inequality is true if $k < n/6$. As before, there exists $j \notin S_k$ such that $|\{i \in T_k \mid U_i' \ni j\}| \geq |T_k|/3$. We then let $S_{k+1} \coloneqq S_k \cup \{j\}$ and $T_{k+1} \coloneqq \{i \in T_k \mid U_i' \not\ni j\}$.

Clearly $|T_k| \leq K \cdot (2/3)^{k-1}$, hence in $\tau = \lceil \log_{3/2} K \rceil \leq 2\log_2 K \leq 2|I| \leq n/10$ steps we eliminate all assignments from the set, i.e. S_τ satisfies that for every assignment ρ to I that fixes at most $(n+1)/2$ output bits, there exists $j \in S_\tau$ that is not fixed by ρ. (The bound $2|I| \leq n/10$ guarantees that the condition $k < n/6$ holds). ◀

Let I be the set of all k-influential bits. If $|I| \geq n/20$ we get the desired lower bound immediately, so assume otherwise. Then let S be the set given by Lemma 21, $|S| = O(|I|)$. By Lemma 20 we get that $|S| = \Omega(n/(dk))$, so $|I| = \Omega(n/(dk))$ as well. ◀

References

1 Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795–815, 2021. doi:10.4007/annals.2021.194.3.5.

2 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.

3 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. doi:10.1109/FOCS.2012.82.

4 Tolson Bell, Suchakree Chueluecha, and Lutz Warnke. Note on sunflowers. *Discrete Mathematics*, 344(7):112367, 2021. doi:10.1016/j.disc.2021.112367.

5 Olaf Beyersdorff, Samir Datta, Andreas Krebs, Meena Mahajan, Gido Scharfenberger-Fabian, Karteek Sreenivasaiah, Michael Thomas, and Heribert Vollmer. Verifying proofs in constant depth. *ACM Trans. Comput. Theory*, 5(1):Art. 2, 23, 2013. doi:10.1145/2462896.2462898.

6 Andrej Bogdanov, Krishnamoorthy Dinesh, Yuval Filmus, Yuval Ishai, Avi Kaplan, and Akshayaram Srinivasan. Bounded indistinguishability for simple sources. In Mark Braverman, editor, *13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 – February 3, 2022, Berkeley, CA, USA*, volume 215 of *LIPIcs*, pages 26:1–26:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.ITCS.2022.26.

7 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*, pages 593–618. Springer, 2016.

8 Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The Space Complexity of Sampling. In Mark Braverman, editor, *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, volume 215 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 40:1–40:23, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2022.40.

9 Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 670–683, 2016.

10 Gil Cohen and Leonard J Schulman. Extractors for near logarithmic min-entropy. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 178–187. IEEE, 2016.

11 Artur Czumaj. Random permutations using switching networks. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 703–712, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2746539.2746629.

12 C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22:89–103, 1971. doi:10.1007/bf01651330.

13 Efraim Gelman and Amnon Ta-Shma. The Benes Network is $q(q-1)/2n$-Almost q-set-wise Independent. In Venkatesh Raman and S. P. Suresh, editors, *34th International Conference on Foundation of Software Technology and Theoretical Computer Science (FSTTCS 2014)*, volume 29 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 327–338, Dagstuhl, Germany, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.FSTTCS.2014.327.

14 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.

15 Torben Hagerup. Fast parallel generation of random permutations. In *Automata, Languages and Programming: 18th International Colloquium Madrid, Spain, July 8–12, 1991 Proceedings 18*, pages 405–416. Springer, 1991.

16 T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

17 Johan Håstad. Computational limitations of small-depth circuits. *MIT Press*, 1987.

18 Russell Impagliazzo and Moni Naor. Efficient cryptographic schemes provably as secure as subset sum. *Journal of cryptology*, 9(4):199–216, 1996.

19 Joe Kilian. Founding crytpography on oblivious transfer. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, STOC '88, pages 20–31, New York, NY, USA, 1988. Association for Computing Machinery. doi:10.1145/62212.62215.

20 Andreas Krebs, Nutan Limaye, Meena Mahajan, and Karteek Sreenivasaiah. Small depth proof systems. *ACM Trans. Comput. Theory*, 9(1):Art. 2, 26, 2016. doi:10.1145/2956229.

21 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. IEEE Computer Society, 2011. doi:10.1109/CCC.2011.11.

22 Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. *Comput. Complexity*, 21(2):245–266, 2012. doi:10.1007/s00037-012-0039-3.

23 Anup Rao. Coding for sunflowers, 2019. doi:10.48550/ARXIV.1909.04774.

24 Benjamin Rossman. The monotone complexity of k-clique on random graphs. *SIAM J. Comput.*, 43(1):256–279, 2014. doi:10.1137/110839059.

25 R. 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*, STOC '87, pages 77–82, New York, NY, USA, 1987. Association for Computing Machinery. doi:10.1145/28395.28404.

26 Emanuele Viola. The complexity of distributions. *SIAM J. Comput.*, 41(1):191–218, 2012. doi:10.1137/100814998.

27 Emanuele Viola. Extractors for Turing-machine sources. In Anupam Gupta, Klaus Jansen, José D. P. Rolim, and Rocco A. Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques – 15th International Workshop, APPROX 2012, and 16th International Workshop, RANDOM 2012, Cambridge, MA, USA, August 15–17, 2012. Proceedings*, volume 7408 of *Lecture Notes in Computer Science*, pages 663–671. Springer, 2012. doi:10.1007/978-3-642-32512-0_56.

28 Emanuele Viola. Extractors for circuit sources. *SIAM J. Comput.*, 43(2):655–672, 2014. doi:10.1137/11085983X.

29 Emanuele Viola. Quadratic maps are hard to sample. *ACM Trans. Comput. Theory*, 8(4):18:1–18:4, 2016. doi:10.1145/2934308.

30 Emanuele Viola. Sampling lower bounds: Boolean average-case and permutations. *SIAM J. Comput.*, 49(1):119–137, 2020. doi:10.1137/18M1198405.

31 Emanuele Viola. New sampling lower bounds via the separator, 2021.

32 Adam Bene Watts and Natalie Parham. Unconditional quantum advantage for sampling with shallow circuits, 2023. arXiv:2301.00995.

A Switching Networks for Sampling Linear Slices

In this section, we discuss the limitations of switching networks for constructing decision tree samplers.

▶ **Definition 22.** *A switching network of depth d is a sequence of d matchings $M_1, \dots, M_d$. Each M_i is a set of $n/2$ disjoint pairs of elements from $[n]$. The distribution generated by a switching network over a slice $\binom{[n]}{\ell}$ is defined as follows:*

- *Initialize the string as $1^\ell 0^{n-\ell}$;*
- *For each $i \in [d]$: for every pair in $(a, b) \in M_i$, we toss a fair coin, and if it comes up heads, we swap the ath and the bth bits of the current sequence.*

▶ **Lemma 23** (A variation of [26])**.** *Suppose there exists a switching network of depth d that generates the variable $\boldsymbol{X}$ over $\binom{[n]}{\ell}$. Then there exists a decision depth d sampler for $\boldsymbol{X}$ such that each input bit affects at most 2^d output bits. Moreover, the support of $\boldsymbol{X}$ is a subset of $\binom{[n]}{\ell}$.*

Proof. The input bits correspond to the coin tosses in the switching network. Each output bit is computed by tracing back its initial position: first, we query the coin corresponding to the pair in M_d containing the bit, then we query the coin corresponding to the pair in M_{d-1} and so on until we compute the location of the bit in the initial sequence. Then if the location is in $[\ell]$ we output 1 and otherwise output 0. It is easy to see that the described sampler has the required properties. ◀

▶ **Lemma 24.** *Let $\alpha \in (0, 1)$ be a constant. Suppose $\boldsymbol{X}$ is samplable with a d-local sampler such that each input bit affects at most c output bits, the support of $\boldsymbol{X}$ is within $\binom{[n]}{\alpha n}$, and $n/(cd)^2 = \omega(2^{cd})$. Then $\Delta(\boldsymbol{X}, \boldsymbol{U}^n_{\alpha n}) = 1 - o(1)$.*

Proof. For each output bit $i \in [n]$ of $\boldsymbol{X}$, let $N(i) \subseteq [n]$, the *neighborhood* of i, be the set of output bits that share an affecting input bit with i. By assumption, $|N(i)| \leq cd$. Let us greedily choose a set of output bits with disjoint neighborhoods: $t_1 = 1$, and for $j > 1$,

$t_j \in [n]$ is a bit such that $N(t_j) \cap (N(t_1) \cup \cdots \cup N(t_{j-1})) = \emptyset$. For each output bit i there are at most $(cd)^2$ output bits j for which $N(i) \cap N(j) \neq \emptyset$, so the greedy process yields $\ell \geq n/(cd)^2$ bits.

Suppose that there is a bit $i \in \{t_1, \ldots, t_\ell\}$ such that $\Pr[\boldsymbol{X}_i = 0] > 0$ and $\Pr[\boldsymbol{X}_{N(i)} = 1^{N(i)}] > 0$. Then we use the approach from Proposition 18: let $x \in \binom{[n]}{\alpha n}$ be a string in the support of $\boldsymbol{X}$ such that $x_{N(i)} = 1^{N(i)}$, and let ρ be the input bits that yield x. Since $\Pr[\boldsymbol{X}_i = 0] > 0$, we can change the bits of ρ affecting the ith output bit such that its value switches to 0. Denote the resulting input by x'. Observe then that $x_{[n] \setminus N(i)} = x'_{[n] \setminus N(i)}$, since $N(i)$ is the set of output bits that are affected by the input bits affecting the ith bit. Then $|x'| < |x|$, hence $x' \notin \binom{[n]}{\alpha n}$, so it does not lie in the support of $\boldsymbol{X}$, which is a contradiction.

Now let us analyze the case when there are no bits satisfying the condition. Let $I \subseteq \{t_1, \ldots, t_\ell\}$ consist of those output bits for which $\Pr[\boldsymbol{X}_i = 0] = 0$. If $|I| \geq \ell/2$ then

$$\Delta(\boldsymbol{X}, \boldsymbol{U}^n_{\alpha n}) \geq \left|\Pr[\boldsymbol{X}_I = 1^I] - \Pr[(\boldsymbol{U}^n_{\alpha n})_I = 1^I]\right| = 1 - \binom{n - |I|}{\alpha n}\binom{n}{\alpha n}^{-1}$$
$$= 1 - \prod_{j=0}^{\alpha n - 1} \frac{n - |I| - j}{n - j} \geq 1 - \left(1 - \frac{|I|}{n - \alpha n + 1}\right)^{\alpha n} \geq 1 - 2^{-\Omega(\frac{\alpha}{1-\alpha}|I|)} = 1 - 2^{-\Omega(\ell)}.$$

Since $\ell = \Omega(n/(cd)^2) = \omega(1)$, in this case $\Delta(\boldsymbol{X}, \boldsymbol{U}^n_{\alpha n}) = 1 - o(1)$.

Now suppose that $|I| \leq \ell/2$, and let $J = \{t_1, \ldots, t_\ell\} \setminus I$. Assume that all bits in J satisfy $\Pr[\boldsymbol{X}_{N(i)} \neq 1^{N(i)}] = 1$. Let us compute the probability of this event for $\boldsymbol{U}^n_{\alpha n}$:

$$\Pr[(\boldsymbol{U}^n_{\alpha n})_{N(i)} \neq 1^{N(i)}] = 1 - \binom{n - |N(i)|}{\alpha n}\binom{n}{\alpha n}^{-1} = 1 - \prod_{j=0}^{\alpha n - 1} \frac{n - |N(i)| - j}{n - j}$$
$$\leq 1 - \left(1 - \frac{|N(i)|}{n}\right)^{\alpha n} \leq 1 - 2^{-\Omega(\alpha |N(i)|)} = 1 - 2^{-\Omega(cd)}.$$

Therefore

$$\Pr[(\boldsymbol{U}^n_{\alpha n})_{N(i)} \neq 1^{N(i)} \text{ for all } i \in J] = \prod_{i \in J}(1 - 2^{-\Omega(cd)}) \leq (1 - 2^{-\Omega(cd)})^{\ell/2}.$$

Since $\ell/2 = n/(2(cd)^2) = \omega(2^{cd})$, we get the desired lower bound on the statistical distance in this case as well. ◀

▶ **Corollary 25.** *Let $\alpha \in (0, 1)$ be a constant. Any switching network that generates a distribution that is $1 - \Omega(1)$ close to $\boldsymbol{U}^n_{\alpha n}$ has depth $\Omega(\log \log n)$.*

Proof. Consider a switching network of depth d that generates a distribution that is $1 - \Omega(1)$ close to $\boldsymbol{U}^n_{\alpha n}$. Lemma 23 translates it to a decision depth d sampler for $\boldsymbol{U}^n_{\alpha n}$ such that each input bit affects at most $c = 2^d$ output bits. The sampler is $1 - \Omega(1)$ close to $\boldsymbol{U}^n_{\alpha n}$, and supported within $\binom{[n]}{\alpha n}$. The result now follows from Lemma 24. ◀

B Exposition of the $\Omega(\log^* n)$ lower bound for Majority

The following is an exposition of the proof of [5, Theorem 5.1].

Suppose that n is odd, and consider a locality c proof system for the vectors containing more 1s than 0s, that is, having Hamming weight at least $(n+1)/2$. We can assume that $c \geq 2$.[4]

The proof system has *inputs* and *outputs*. The number of outputs is n, and each one depends on at most c input bits. An input bit is *d-influential* if at least d output bits depend on it. There are at most cn/d many d-influential input bits.

Let $1 = B(0) < B(1) < \cdots < B(c+1)$ be a sequence of constants (depending on c but not on n), and let $d(\ell) = cn/B(\ell)$, so that $cn = d(0) > d(1) > \cdots > d(c+1) = \Omega(n)$. Thus there are at most $B(\ell)$ many input bits which are $d(\ell)$-influential.

We will construct a sequence of sets $\emptyset = R_0 \subseteq R_1 \subseteq \cdots \subseteq R_c \subseteq [n]$ with the following property: *If ρ is a truth assignment to the $d(\ell)$-influential variables which extends to a complete truth assignment setting all coordinates in R_ℓ to zero, then for each coordinate $i \notin R_\ell$, ρ also extends to a complete truth assignment setting coordinate i to zero.*[5]

Given $R_{\ell-1}$, here is how we construct R_ℓ. We start with $R \coloneqq R_{\ell-1}$, and will potentially add more output coordinates to R. At any point, suppose that there is a truth assignment ρ to the $d(\ell)$-influential variables which (i) extends to a complete truth assignment setting all coordinates in R to zero, and (ii) for some coordinate $i \notin R$, any complete truth assignment extending ρ sets coordinate i to one. If that happens, then we add i to R. Henceforth, ρ will not come up again, since no complete truth assignment extending ρ sets i to one, and i belongs to R. Eventually, there is no such "bad" truth assignment, and we set $R_\ell \coloneqq R$.

If $i \in R_\ell$ then there exists some $r \leq \ell$, some $R \subseteq R_r$, and some truth assignment ρ_i to the $d(r)$-influential variables, such that ρ_i extends to a complete truth assignment setting all coordinates in R to zero, and any complete truth assignment extending ρ_i sets i to one. If $j \in R_\ell$ was added after i then the truth assignment ρ_j extends to a complete truth assignment which sets all coordinates in $R \cup \{i\}$ to zero. In particular, ρ_j doesn't extend ρ_i (as a special case, $\rho_j \neq \rho_i$). It follows that if we extend each ρ_i arbitrarily to a truth assignment to the $d(\ell)$-influential variables, then the resulting assignments will all be different. Consequently,

$$|R_\ell| \leq 2^{B(\ell)}.$$

For large enough n, this will be at most $\frac{n-1}{2}$.

We show below that for a proper choice of parameters, there is an output coordinate i which satisfies the following, for all $\ell \in \{0, \ldots, c\}$: *$i \notin R_\ell$, and all inputs to i which are also inputs to R_ℓ are $d(\ell+1)$-influential.* We will show that for each $\ell \in \{0, \ldots, c\}$, i has an input which is $d(\ell+1)$-influential but not $d(\ell)$-influential. For different ℓ these inputs are different (since an input which is not $d(\ell)$-influential is also not $d(r)$-influential for all $r < \ell$), and so i depends on $c+1$ inputs, which is impossible.

Let $\ell \in \{0, \ldots, c\}$. Let us show that i has an input which is $d(\ell+1)$-influential but not $d(\ell)$-influential. We do this by contradiction: suppose that all $d(\ell+1)$-influential inputs of i are $d(\ell)$-influential. By assumption, $i \notin R_\ell$ and all inputs to i which are also inputs to R_ℓ are $d(\ell)$-influential. Let $N(i)$ consist of i together with all other output bits which share some non-$d(\ell)$-influential bit with i. All of these shared input bits are in fact non-$d(\ell+1)$-influential, and so

$$|N(i)| \leq 1 + cd(\ell+1) \leq 1 + cd(1) = 1 + \frac{c^2 n}{B(1)}.$$

[4] Alternatively, change $B(0)$ or redefine d-influential as having *more than* d output bits depending on it.
[5] In the paper, ρ extends to a complete truth assignment setting both R_ℓ and i to zero.

If $B(1) > 2c^2$ then $|N(i)| < 1 + \frac{n}{2}$ and so $|N(i)| \leq (n+1)/2$. Since $|R_c| \leq \frac{n-1}{2}$, the proof system generates some vector v of weight $(n+1)/2$ in which all coordinates of R_c are zero and all coordinates of $N(i)$ are one. Consider an arbitrary complete truth assignment α which generates v, and let ρ be its restriction to the $d(\ell)$-influential coordinates. Since $i \notin R_\ell$, by construction, we know that ρ extends to some complete truth assignment β which sets i to zero. Now consider the following complete truth assignment:

$$\gamma(j) = \begin{cases} \rho(j) & \text{if } j \text{ is } d(\ell)\text{-influential,} \\ \beta(j) & \text{if } j \text{ is not } d(\ell)\text{-influential and influences } N(i), \\ \alpha(j) & \text{if } j \text{ is not } d(\ell)\text{-influential and doesn't influence } N(i). \end{cases}$$

Since α, β both extend ρ, γ agrees with them on the $d(\ell)$-influential variables. Therefore the output generated by γ agrees with that generated by α except for the coordinates in $N(i)$, which could change from one to zero. Moreover, the i'th output of γ agrees with the i'th output of β, namely, it is zero. Therefore the output generated by γ has Hamming weight strictly less than $(n+1)/2$, which is impossible.

It remains to show that there exists an output bit i such that $i \notin R_c$, and for all $\ell \in \{0, \ldots, c\}$, all inputs to i which are also inputs to R_ℓ are $d(\ell+1)$-influential. We do this by giving an upper bound on the number of bad output bits. An output bit is bad if it either belongs to R_c, or for some $\ell \in \{0, \ldots, c\}$, there is a joint input of i and R_ℓ which is not $d(\ell+1)$-influential. If i is bad due to some ℓ, then there must be some non-$d(\ell+1)$-influential input of R_ℓ which is an input of i. Therefore the number of bad inputs is at most

$$|R_c| + \sum_{\ell=0}^{c} |R_\ell| \cdot c \cdot d(\ell+1) \leq 2^{B(c)} + n \cdot c^2 \sum_{\ell=0}^{c} \frac{2^{B(\ell)}}{B(\ell+1)}.$$

For a judicious choice of the sequence $B(\ell)$, the coefficient of n will be strictly less than 1, and so for large enough n, there are fewer than n bad inputs.

One choice for the sequence $B(\ell)$ is

$$B(\ell+1) = 2^{B(\ell)} \cdot 2c^2(c+1).$$

In particular, $B(1) = 4c^2(c+1) > 2c^2$, which was needed above. For this choice of $B(0), \ldots, B(c)$, the number of bad inputs is at most

$$2^{B(c)} + \frac{n}{2},$$

which less than n if $2^{B(c)} \leq \frac{n-1}{2}$, a condition which was required at a different step of this proof.

Roughly speaking, $B(\ell+1) \approx 2c^3 \cdot 2^{B(\ell)}$, and so $B(c) \approx 2 \uparrow\uparrow c$. Therefore $2^{B(c)} \leq \frac{n-1}{2}$, and so the argument works, for $c \leq \kappa \log^* n$, for an appropriate constant κ.