Symmetric Distributions from Shallow Circuits

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

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

Kewen Wu
Caltech, Pasadena, CA, USA

Abstract

We characterize the symmetric distributions that can be (approximately) generated by shallow Boolean circuits. More precisely, let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a Boolean function where each output bit depends on at most d input bits. Suppose the output distribution of f evaluated on uniformly random input bits is close in total variation distance to a symmetric distribution $\mathcal{D}$ over $\{0,1\}^n$. Then $\mathcal{D}$ must be close to a mixture of the uniform distribution over n-bit strings of even Hamming weight, the uniform distribution over n-bit strings of odd Hamming weight, and γ-biased product distributions for γ an integer multiple of 2^{-d}. Moreover, the mixing weights are determined by low-degree, sparse $\mathbb{F}_2$-polynomials. This extends the previous classification for generating symmetric distributions that are also uniform over their support.

2012 ACM Subject Classification Theory of computation → Circuit complexity

Keywords and phrases Sampling, distributions, locality, circuit complexity

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.53

Category RANDOM

Related Version *Full Version*: `https://arxiv.org/abs/2511.14127` [32]

Funding *Daniel M. Kane*: Supported by NSF Medium Award CCF-2107547 and NSF CAREER Award CCF-1553288.

Acknowledgements We thank Farzan Byramji and anonymous reviewers for helpful comments on an earlier draft of this paper.

1 Introduction

One of the most celebrated results in complexity theory is that AC^0 circuits[1] require an exponential number of gates to compute the parity function [22, 1, 45, 26, 27]. Surprisingly, the weaker circuit class NC^0 suffices to perform a very similar task. In particular, mapping uniformly random bits $(x_1, x_2, \ldots, x_n)$ to $(x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_{n-1} \oplus x_n, x_n \oplus x_1)$ produces the uniform distribution over n-bit strings of even parity, or equivalently, over input-output pairs $(x, \mathrm{PARITY}(x))$ [5, 10]. This observation begs the question: what computational resources are required to (approximately) *generate* specific distributions, as opposed to the traditional task of *computing* specific functions?

A fundamental question in its own right, the complexity of sampling from distributions also has numerous applications. For example, results on the hardness of sampling can be translated to data structure lower bounds [40, 35, 6, 43, 11, 44, 47, 30, 2], provide input-independent quantum-classical separations [7, 44, 30, 24], and are key components

[1] Recall that these are (families of) Boolean circuits of constant depth and unbounded fan-in gates. They may be contrasted with NC^0 circuits, which are also of constant depth, but have bounded fan-in gates.

to the construction of explicit codes [37]. Moreover, techniques and intuition developed in this setting have successfully been applied to pseudorandom generators [40, 35, 6] and extractors [41, 17, 42, 13, 14].

While the general problem was considered in early work (see, e.g., [29]), a focus on generating distributions via shallow circuits was advocated for more recently by Viola [40]. Since then, the field has seen a number of exciting developments (see, e.g., the recent works [21, 44, 30, 37, 31, 2] and references therein). One notable takeaway from prior works is that NC^0 circuits can sample very few uniform symmetric distributions (i.e., uniform distributions over a symmetric support), even allowing for small errors. In particular, the line of work [40, 21, 44, 30, 31] recently culminated in the following classification result, which confirmed a conjecture of Filmus, Leigh, Riazanov, and Sokolov [21]. For a function $f\colon \{0,1\}^m \to \{0,1\}^n$, let $f(\mathcal{U}^m)$ be the distribution resulting from applying f to $x \sim \mathcal{U}^m$, the uniform distribution over $\{0,1\}^m$.

▶ **Theorem 1** ([31]). *Let $\varepsilon \in [0,1]$ be arbitrary. Assume $f\colon \{0,1\}^m \to \{0,1\}^n$ is computable by an NC^0 circuit of constant depth and $f(\mathcal{U}^m)$ is ε-close in total variation distance to a uniform symmetric distribution where n is sufficiently large. Then $f(\mathcal{U}^m)$ is $O(\varepsilon)$-close to one of the following six special uniform symmetric distributions:*

- *Point distribution on 0^n,*
- *Point distribution on 1^n,*
- *Uniform distribution over $\{0^n, 1^n\}$,*
- *Uniform distribution over n-bit strings with even Hamming weights,*
- *Uniform distribution over n-bit strings with odd Hamming weights,*
- *Uniform distribution over all n-bit strings.*

We emphasize that Theorem 1 works with *uniform* symmetric distributions, which does not capture, for example, the $(1/4)$-biased product distribution that is symmetric and easily sampleable by NC^0 circuits.

1.1 Our Result

In this paper, we take another step toward understanding the sampleability of distributions by extending Theorem 1 to handle *arbitrary* symmetric distributions. Let the *locality* of a Boolean function be the largest number of input bits that any output bit depends on. Note that NC^0 is precisely the class of sequences of functions with bounded locality.

▶ **Theorem 2** (Informal version of [32, Theorem 4.1]). *Let $d \geq 0$ be an integer. For any $\varepsilon \in (0,1]$, there exists some δ such that $\delta \to 0$ as $\varepsilon \to 0$ and the following holds.*

Suppose $f\colon \{0,1\}^m \to \{0,1\}^n$ is a d-local function and $f(\mathcal{U}^m)$ is ε-close in total variation distance to a symmetric distribution where n is sufficiently large in terms of d and ε. Then $f(\mathcal{U}^m)$ is δ-close to some mixture of

1. *The uniform distribution over n-bit strings of even Hamming weight,*
2. *The uniform distribution over n-bit strings of odd Hamming weight, and*
3. *γ-biased product distributions on n bits for γ an integer multiple of 2^{-d}.*

Moreover, the mixing weights are determined by degree-$O_d(1)$ $\mathbb{F}_2$-polynomials with $O_d(n)$ monomials.

The "moreover" conclusion implies this mixture can be exactly produced by $O_d(1)$-local functions (see [32, Remark 4.2]).

Learning Structured Distributions

As mentioned above, the study of sampling is rife with applications to other areas of theoretical computer science. Continuing this trend, we highlight one consequence of Theorem 2 to learning theory.

The reconstruction of an unknown probability density function based on observed data is a fundamental problem in both statistics and computer science. The typical setting is the PAC-learning model [38, 9, 33]: given access to independent samples of an unknown distribution $\mathcal{D}$, the goal is to output a hypothesis distribution $\mathcal{D}'$ close to $\mathcal{D}$. Much research has been carried out for Gaussian mixtures [16], log-concave distributions [20], monotone distributions [8], sums of independent integer random variables [15], junta distributions [3], mixtures of structured distributions [34], and more. The interested reader may wish to consult the survey [18] by Diakonikolas for additional background and references.

A black-box use of Theorem 2 in this setting is to learn symmetric distributions that are locally sampleable. Let $\mathcal{D}$ be a symmetric distribution over $\{0,1\}^n$. Assume $\mathcal{D}$ is produced by a d-local function, i.e., $\mathcal{D} = f(\mathcal{U}^m)$ for some d-local function $f\colon \{0,1\}^m \to \{0,1\}^n$. Then for any $\varepsilon > 0$ with n sufficiently large in terms of d and ε, Theorem 2 and the standard cover method [46] (see also [18, Theorem 1.5.1]) imply that we can efficiently learn $\mathcal{D}$ up to ε-error in total variation distance with $O_d(1/\varepsilon^2)$ samples with high probability. We believe this should hold for all $\varepsilon > 0$. Indeed, in Section 1.2 we propose Conjecture 4 for an exact classification of locally sampleable symmetric distributions, which, if true, would imply the desired learning theoretic result.

1.2 Open Problems

Recall that Theorem 2 does not work for the case of $\varepsilon = 0$, i.e., *exact* sampling. Indeed, we cannot conclude mixtures of the form given by Theorem 2 are the only distributions that can be exactly sampled by NC^0 circuits. We identify the following example, which illustrates that this is not simply a weakness in our analysis, but rather there are other symmetric distributions that can be sampled exactly.

▶ **Example 3.** Let $\mathcal{U}^n_{1/4}$ be the $(1/4)$-biased product distribution over $\{0,1\}^n$. Additionally, let $\texttt{evens}$ and $\texttt{odds}$ denote the uniform distribution over n-bit strings of even Hamming weight and odd Hamming weight, respectively. The distribution[2]

$$\mathcal{P} = \mathcal{U}^n_{1/4} + 2^{-n-1}\texttt{evens} - 2^{-n-1}\texttt{odds}$$

is not of the form given by Theorem 2, yet it can be sampled exactly by a bitwise AND of $\texttt{evens}$ and the uniform distribution over $\{0,1\}^n$, which is 3-local. The full details can be found in [32, Appendix A].

Additionally, we suspect that, similarly to the uniform symmetric case [31], one can take the upper bound in Theorem 2 to be linear in $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}}$ without any dependency on d. Combining with the previous discussion, we conjecture the following strengthening of Theorem 2 holds. Observe that Conjecture 4 captures Example 3.

▶ **Conjecture 4.** *For every $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, and n large enough in terms of d, if $f\colon \{0,1\}^m \to \{0,1\}^n$ is a d-local function and $f(\mathcal{U}^m)$ is ε-close in total variation distance to a symmetric distribution, then $f(\mathcal{U}^m)$ is $O(\varepsilon)$-close to some mixture*

$$\mathcal{M} = \sum_i \alpha_i \cdot g_i(\mathcal{D}^{(i)}_1, \ldots, \mathcal{D}^{(i)}_d),$$

[2] More formally, $\mathcal{P}$ is the distribution that assigns x probability $\mathcal{U}^n_{1/4}(x) + 2^{-n-1}\texttt{evens}(x) - 2^{-n-1}\texttt{odds}(x)$.

where each g_i is a bitwise function and each $\mathcal{D}_j^{(i)}$ is either the uniform distribution over n-bit strings of even Hamming weight or over n-bit strings of odd Hamming weight. Moreover, the mixing weights are determined by degree-$O_d(1)$ $\mathbb{F}_2$-polynomials with $O_d(n)$ monomials, as in Theorem 2.

Note the bitwise condition on the above g_i's guarantees the output distribution is symmetric.

Paper Organization

We provide an overview of the proof of Theorem 2 in Section 2, as well as a brief comparison of our techniques to those of prior literature. Our full results and proofs can be found in the arXiv version of the paper [32].

2 Proof Overview

In this section, we sketch the proof of Theorem 2 before discussing how the details compare to prior work.

2.1 Proof Overview of Theorem 2

We begin with a useful observation from [31]: the total variation distance between the distribution $f(\mathcal{U}^m)$ and any symmetric distribution $\mathcal{P}$ over $\{0,1\}^n$ is, up to constant factors, equal to the distance between the corresponding Hamming weight distributions of $f(\mathcal{U}^m)$ and $\mathcal{P}$ plus the distance between $f(\mathcal{U}^m)$ and its symmetrization (i.e., the distribution resulting from randomly permuting the coordinates of a string $x \sim f(\mathcal{U}^m)$). Expressed symbolically, we have

$$\|f(\mathcal{U}^m) - \mathcal{P}\|_{\mathsf{TV}} = \Theta(\||f(\mathcal{U}^m)| - |\mathcal{P}|\|_{\mathsf{TV}} + \|f(\mathcal{U}^m) - f(\mathcal{U}^m)_{\mathrm{sym}}\|_{\mathsf{TV}}). \tag{1}$$

The proof is straightforward, and it essentially follows from several applications of the triangle inequality which show that the two ways $f(\mathcal{U}^m)$ can be far from $\mathcal{P}$ are a weight mismatch or $f(\mathcal{U}^m)$ being far from symmetric itself (see [32, Lemma 3.5]).

By assumption, we know $f(\mathcal{U}^m)$ is ε-close to a symmetric distribution $\mathcal{D}$. Thus, Equation (1) implies $\|f(\mathcal{U}^m) - f(\mathcal{U}^m)_{\mathrm{sym}}\|_{\mathsf{TV}} = O(\varepsilon)$. Hence, it suffices to prove a weaker version of Theorem 2 that only compares the weight distributions ([32, Lemma 4.26]). More precisely, we want to show that for some δ tending to 0 with ε, the Hamming weight distribution of $f(\mathcal{U}^m)$, denoted $|f(\mathcal{U}^m)|$, is δ-close to a mixture $\mathcal{M}$ of

1. The binomial distribution $\mathsf{Bin}(n, \gamma)$ with n trials and success probability γ for γ an integer multiple of 2^{-d},
2. The binomial distribution $\mathsf{Bin}(n, 1/2)$ conditioned on the outcome being even, and
3. The binomial distribution $\mathsf{Bin}(n, 1/2)$ conditioned on the outcome being odd.

Moreover, we want the mixing weights to be determined by low-degree $\mathbb{F}_2$-polynomials with few monomials. (It will later become clear what "determined" means in this context.) We first prove a weaker version of this result that does not include mixing weight information ([32, Lemma 4.24]). Afterwards, we will discuss how to obtain the desired control over the weights. Before proceeding to the details, we first give a brief, high-level overview of the four main steps of the proof and how they fit together.

In the first step, we show that any fixing of the values of input bits which affect many output bits results in the bias of the output weight distribution of $f(\mathcal{U}^m)$ concentrating around a fixed dyadic rational[3] γ multiple of n. This allows us to represent the weight distribution $|f(\mathcal{U}^m)|$ as a mixture of distributions produced by the restricted functions. We then argue in step two that after grouping the parts of the mixture which concentrate around the same γ, each group assigns roughly the same amount of mass to any contiguous interval as the binomial distribution with success probability γ. The third step is to prove a continuity result for each grouped part of the mixture; namely, that the mass assigned to some weight w and to $w + \Delta$ for a small integer Δ are roughly equal. (There is a slight subtlety here in the case of $\gamma = 1/2$, but we will defer its discussion to later in the proof overview.) This allows us to conclude in the fourth and final step that most output weights are assigned comparable mass by the part of the mixture of $|f(\mathcal{U}^m)|$ concentrated around γ and the corresponding binomial distribution. In particular, our weight distribution is close in total variation distance to a mixture of binomial distributions, which is what we wanted to show.

2.1.1 Step 1: Removing Large Influences

In an extremely ideal setting, we might wish that all of f's output bits are roughly independent with bias around $\gamma = a/2^d$ for some integer $0 \le a \le 2^d$; in this case the output weight would resemble the binomial distribution $\mathsf{Bin}(n, \gamma)$. Of course, this is far too much to assume. In actuality, it may be that one specific input bit affects the value of *every* output bit, so none of them are independent.

To progress toward this dream scenario, we follow in the footsteps of many prior works (e.g., [40, 35, 6, 43, 44, 21, 30, 31, 24]) by strategically conditioning on certain input bits to express the distribution as a mixture of more structured sub-distributions. For example, suppose some input bit b affects many output bits. Since b takes value 0 and 1 with equal probability, we may express the distribution $f(\mathcal{U}^m)$ as the mixture

$$f(\mathcal{U}^m) = \frac{1}{2} \cdot f\left(\mathcal{U}^m \mid b = 0\right) + \frac{1}{2} \cdot f\left(\mathcal{U}^m \mid b = 1\right).$$

Observe that the large influence of b is no longer a problem in either of the restricted distributions, as its value is fixed.

In particular, we will condition on all "high degree" input bits that affect more than n/A output bits for some A to be chosen later. The goal is now to argue that the function resulting from each conditioning produces a distribution which is roughly of the form we originally sought. Note that this will not depend on the actual values that the input bits are set to in the conditioning.

We start with a result from [30]: let $\mathcal{D}_q$ be the uniform distribution over n-bit strings of Hamming weight q. If q/n is far from every integer multiple of 2^{-d}, then any d-local function must produce a distribution far (in total variation distance) from $\mathcal{D}_q$. This essentially follows from the fact that the input bits determining any fixed output bit can only be set in 2^d equally likely ways. Observe that symmetric distributions are simply mixtures of $\mathcal{D}_q$ for different q's. Thus we can show ([32, Lemma 4.8]) that if $f(\mathcal{U}^m)$ is close to a symmetric distribution $\mathcal{D}$, then for a typical $x \sim \mathcal{U}^m$, the normalized output weight $|f(x)|/n$ has distance at most $n^{-1/(800d)}$ from an integer multiple of 2^{-d}. Note that the closest multiple may be different for different inputs; however, after each conditioning $\rho \in \{0,1\}^S$ on the set of high degree input

[3] Recall a *dyadic rational* is a number that can be expressed as a fraction whose denominator is a power of two.

bits $S \subseteq [m] := \{1, 2, \ldots, m\}$, we can strengthen this result ([32, Lemma 4.9]) to say that for a typical $x \sim \mathcal{U}^{[m]\setminus S}$, the normalized output weight $|f(x, \rho)|/n$ is close to the *same* integer multiple of 2^{-d}. Here, we will only need "high degree" to correspond to affecting more than $n/O_d(1)$ many output bits (i.e., can take $A = O_d(1)$), although later we will obtain other constraints on how we must set A. Henceforth, we will shorthand $f(x, \rho)$ by $f_\rho(x)$ for clarity.

At a high level, the proof uses the second-moment method. By conditioning on the high degree input bits, we ensure that the variance of the resulting distribution is small, and thus we have concentration around a particular weight. Moreover, this weight must be close to an integer multiple of $n/2^d$, or else our previous insights imply $f(\mathcal{U}^m)$ cannot be close to $\mathcal{D}$. To be slightly more precise, a direct second-moment argument by itself is insufficient to rule out the case that some non-negligible fraction of the time the output weight is close to some other multiple of $n/2^d$. However, one can show ([32, Claim 4.14]) that if this occurs, $f_\rho(\mathcal{U}^{[m]\setminus S})$ would also have to assign decent probability mass to weights between these integer multiples, which would again imply $f(\mathcal{U}^m)$ cannot be close to $\mathcal{D}$.

Finally, we combine these deductions to prove that for each conditioning $\rho \in \{0, 1\}^S$ of the bits in S, there exists a set $T := T_\rho \subseteq [n]$ of size $|T| \leq O_{d,k}(1)$ such that every k-tuple of output bits in $[n] \setminus T$ has marginal distribution $\mathcal{U}^k_{\gamma_\rho}$, the γ_ρ-biased product distribution over $\{0, 1\}^k$, for γ_ρ an integer multiple of 2^{-d} (see [32, Proposition 4.5]). Here, k is some parameter at most $O_d(\log(1/\varepsilon))$. The proof operates by constructing a degree-k multilinear polynomial $P_i \colon \{0, 1\}^n \to \{0, 1\}$ for each k-tuple with the "wrong" distribution, where the expectation of P_i over $f_\rho(\mathcal{U}^{[m]\setminus S})$ is larger than over the γ-biased product distribution $\mathcal{U}^n_\gamma$ by at least an additive 2^{-kd} term. Note this follows from using our locality bound to view $P_i(f_\rho(\mathcal{U}^{[m]\setminus S}))$ as a polynomial of degree kd in the input bits. Summing the polynomials together into $P = \sum_i P_i$ magnifies the difference in expectations, and if the number of terms (i.e., number of bad k-tuples) is too large, we end up contradicting the assumption that $f(\mathcal{U}^m)$ is close to a symmetric distribution.

2.1.2 Step 2: Kolmogorov Distance

We now group the restricted functions according to their biases, defining $F_\gamma(\mathcal{U}^{[m]\setminus S}) = \mathbb{E}_{\rho:\gamma_\rho=\gamma}[f_\rho(\mathcal{U}^{[m]\setminus S})]$ for each γ. In this second step, we aim to show that every contiguous interval of output weights is assigned roughly the same amount of mass by $|F_\gamma(\mathcal{U}^{[m]\setminus S})|$ and $\mathsf{Bin}(n, \gamma)$. Since this difference in probability mass is convex over mixtures, it suffices to consider the individual distributions $|f_\rho(\mathcal{U}^{[m]\setminus S})|$. Moreover, it is enough to provide a bound on the *Kolmogorov distance*

$$\max_t \left| \Pr\left[|f_\rho(\mathcal{U}^{[m]\setminus S})| \geq t\right] - \Pr\left[\mathsf{Bin}(n, \gamma) \geq t\right] \right|.$$

In a sentence, the bound follows from combining the k-wise independence of most output coordinates with the fact that T is too small to have much of an effect. In the case of $\gamma = 1/2$ and $|T| = 0$, Diakonikolas, Gopalan, Jaiswal, Servedio, and Viola [19] gave an upper bound of roughly $1/\sqrt{k}$ using techniques from approximation theory. (In fact, their result holds for arbitrary threshold functions.) This was generalized by Gopalan, O'Donnell, Wu, and Zuckerman [23] to include the case of arbitrary $\gamma \in (0, 1)$. In our case, $|T|$ will likely not be 0, but it is small enough that a similar result ([32, Proposition 4.16]) still holds. In the edge cases of $\gamma \in \{0, 1\}$, we can no longer use [19, 23], nor would the size of T be negligible even if we could, but these special cases can be addressed later via simple arguments (see, e.g., the proof of [32, Lemma 4.23]).

For reasons that will become apparent in the subsequent step, we will also need a comparable bound in the case of $\gamma = 1/2$, even accounting for the parity of the output weight. That is, we wish to show

$$\Pr\left[|f_\rho(\mathcal{U}^{[m]\setminus S})| \geq t \text{ and } |f_\rho(\mathcal{U}^{[m]\setminus S})| \text{ is even}\right] \approx \Pr\left[\mathsf{Bin}(n,1/2) \geq t\right] \cdot \Pr\left[|f_\rho(\mathcal{U}^{[m]\setminus S})| \text{ is even}\right].$$

Our analysis here is similar to the previous case, but we now crucially rely on [12, Theorem 3.1]. In our context, it implies that there is a small set of input bits $R \subseteq [m] \setminus S$ such that re-randomizing over R typically re-randomizes the parity of f_ρ's output weight. Moreover, since R is small and we have already conditioned on input bits of large degree, the vast majority of output bits are unaffected by R. Thus, we can compare $\Pr\left[|f_\rho(\mathcal{U}^{[m]\setminus S})| \geq t\right]$ to $\Pr\left[\mathsf{Bin}(n,1/2) \geq t\right]$ using these unaffected output bits as a proxy for the entirety of the output bits. We briefly note that for the error bounds on the "parity Kolmogorov distance" to be meaningful given our choice of k, we need to take A (defined in our earlier high degree threshold) to be $O_{d,\varepsilon}(1)$ – larger than is necessary for Step 1.

2.1.3 Step 3: Approximate Continuity

Our third step is to argue that for each bias γ, the distribution $F_\gamma(\mathcal{U}^{[m]\setminus S})$ can be expressed as a mixture $\lambda \cdot E_\gamma + (1-\lambda) \cdot W_\gamma$, where λ is small and W_γ assigns similar probability mass to similar weights. Once again, it suffices to analyze the distributions produced by the individual restricted functions $f_\rho(\mathcal{U}^{[m]\setminus S})$. Here, we show that $f_\rho(\mathcal{U}^{[m]\setminus S})$ can be written as a mixture of distributions E and W, where E is extremely far from every symmetric distribution supported on strings of Hamming weight $\gamma n \pm n^{2/3}$, and the weight distribution of W satisfies a certain continuity property (see [32, Proposition 4.21]). More specifically, for positive integers w and Δ, the probability that W's output weight is w differs by no more than $O_d(\Delta/n)$ from the probability that W's output weight is $w + \Delta$. Since $f(\mathcal{U}^m)$ is close to $\mathcal{D}$, one can show that $F_\gamma(\mathcal{U}^{[m]\setminus S})$ is close to $\mathcal{D}$ conditioned on the output weight being $\gamma n \pm n^{2/3}$ (see [32, Claim 4.25]). Thus, proving the above result for $f_\rho(\mathcal{U}^{[m]\setminus S})$ implies minimal mass will typically be assigned to the E part of the mixtures, and so λ (in the mixture defining $F_\gamma(\mathcal{U}^{[m]\setminus S})$) will be small, as desired.

In our analysis, we will require a structural result about hypergraphs from [30]. Observe that we can associate any function $g\colon \{0,1\}^m \to \{0,1\}^n$ to a hypergraph on the vertex set $[n]$ with an edge for each input bit b containing all of the output bits that depend on b. In the case that g is d-local, this hypergraph has maximum degree at most d. We follow standard hypergraph terminology in defining the *neighborhood* of a vertex v to be the set of vertices sharing an edge with v. We additionally call two neighborhoods N_1, N_2 *connected* if there exist two adjacent (i.e., contained in the same edge) vertices $v_1 \in N_1, v_2 \in N_2$.

By [30, Corollary 4.11], we can find a collection of $r = \Omega_d(n)$ non-connected neighborhoods in $\mathcal{H}$ of size $O_d(1)$ by only removing $O_d(n)$ (with a small implicit constant) edges. Translating the result to f_ρ, we find that there exists a not-too-large set $B \subseteq [m]\setminus S$ of input bits such that any conditioning $\sigma \in \{0,1\}^B$ yields a sub-function $f_{\rho,\sigma}\colon \{0,1\}^{[m]\setminus(S\cup B)} \to \{0,1\}^n$ with many small, pairwise independent collections of output bits. We will define E and W according to the behavior of these neighborhoods, similarly to the arguments in [30, 31, 24].

Suppose for at least $r/2$ of the r non-connected neighborhoods $N_1, \ldots, N_r$, we have that $f_{\rho,\sigma}(\mathcal{U}^{[m]\setminus(S\cup B)})$ restricted to the neighborhood N_i has a marginal distribution which differs from the γ-biased product distribution over N_i, denoted $\mathcal{U}_\gamma^{N_i}$. Since $\mathcal{U}_\gamma^{N_i}$ is pointwise

close to the uniform distribution over strings of Hamming weight around γn, the marginal distributions of these neighborhoods (in $f_{\rho,\sigma}$) must also differ from the marginal distributions of the symmetric distribution $\mathcal{D}$. We can then accumulate these errors via concentration bounds to show that $f_{\rho,\sigma}(\mathcal{U}^{[m]\setminus(S\cup B)})$ is far from $\mathcal{D}$ (see [32, Lemma 3.3]).

We now set E to be the mixture over all conditions of the bits in B where most resulting neighborhoods differ from the corresponding γ-biased product distribution, and set W to be the mixture over the remaining conditionings. Since each sub-distribution in the mixture E is far from $\mathcal{D}$ by the previous paragraph, a union bound argument implies E, itself, must also be far from $\mathcal{D}$ (see [32, Lemma 3.4]). It remains to show the continuity property for W.

Suppose the r non-connected neighborhoods are generated by $v_1, v_2, \ldots, v_r \in [n]$ in the sense that all bits in the i-th neighborhood N_i are either v_i or in an edge that also contains v_i. We further condition on all input bits that do not affect any of $v_1, \ldots, v_r$, so that the value of every output bit outside of $N_1 \cup \cdots \cup N_r$ is fixed. In this way, the output weight of $f_{\rho,\sigma}(\mathcal{U}^{[m]\setminus(S\cup B)})$ becomes a fixed integer (corresponding to the fixed bits outside of the neighborhoods) plus the sum of the neighborhoods' output weights. Since we are in the case where most neighborhoods are extremely structured, we can show that for each $2 \leq \ell \leq t$, the output weight distribution modulo ℓ of N_i is not constant for a constant fraction of the N_i's with high probability (see [32, Claim 4.22]). Finally, we can obtain our desired continuity result through known density comparison theorems for sums of independent, non-constant integer random variables, such as [31, Theorem A.1], which relies on classical anticoncentration tools.

There is one important exception to the above analysis: the case of $\ell = 2$ and $\gamma = 1/2$. Here, we are not guaranteed to typically get many non-connected neighborhoods with variable output weight modulo 2 (even if most of them have the "correct" marginal distribution). To better understand where our reasoning breaks down, consider one of the most surprising examples in this area of work. Define $h\colon \{0,1\}^n \to \{0,1\}^n$ to be

$$h(x_1, \ldots, x_n) = (x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_{n-1} \oplus x_n, x_n \oplus x_1),$$

so that $h(\mathcal{U}^n)$ is the uniform distribution over n-bit strings of even Hamming weight. Observe that h is extremely simple, only requiring two bits of locality. Additionally, the marginal distribution onto any $k \leq n-1$ coordinates is exactly the product distribution $\mathcal{U}_{1/2}^k$.

Let us focus on a particular output bit $y_i = x_i \oplus x_{i+1}$. (For simplicity, assume i is not too close to 1 or n.) Its neighborhood is $y_{i-1} = x_{i-1} \oplus x_i$, y_i, and $y_{i+1} = x_{i+1} \oplus x_{i+2}$, so its Hamming weight modulo 2 is

$$(x_{i-1} \oplus x_i) \oplus (x_i \oplus x_{i+1}) \oplus (x_{i+1} \oplus x_{i+2}) = x_{i-1} \oplus x_{i+2}.$$

If we follow our earlier analysis and fix the value of the input bits that do not affect y_i (i.e., x_{i-1} and x_{i+2}), the neighborhood's Hamming weight modulo two is fixed, regardless of the values of x_i and x_{i+1}.

At a high level, the reason for this exceptional case is that it is the only setting of ℓ and γ for which the binomial distribution $\mathsf{Bin}(t,\gamma)$ mod ℓ can equal $\mathsf{Bin}(t-1,\gamma)$ mod ℓ. In other words, the weight distribution modulo ℓ of $\mathcal{U}_\gamma^k$ conditioned on the first bit being 0 is different than that distribution conditioned on the first bit being 1, unless $\ell = 2$ and $\gamma = 1/2$. This difference means that in the case of $\ell = 2$ and $\gamma = 1/2$, we can only derive a continuity result for weights that are an even distance apart. Hence, we must be cognizant of the output weight's parity throughout much of our analysis, which explains the required parity version of our Kolmogorov distance result mentioned earlier in the proof overview.

2.1.4 Step 4: Putting It Together

At this point, all that remains is to combine the pieces. Recall we have already conditioned on high degree (i.e., larger than $n/O_{d,\varepsilon}(1)$) input bits to find sub-functions $\{f_\rho\}_\rho$, each producing a distribution with some approximate bias γ_ρ. Additionally, the mixtures $F_\gamma(\mathcal{U}^{[m]\setminus S})$ obtained by grouping the sub-functions around their biases each satisfy a Kolmogorov distance bound and approximate continuity.

We partition $\{0, 1, \ldots, n\}$ into consecutive intervals of length $c\sqrt{n}$ for some small $c > 0$, and restrict our attention to the $O(\log(1/\alpha))$ of them that contain all but $O(\alpha)$ of the mass. By our Kolmogorov distance result ([32, Lemma 4.15]), we have that $|F_\gamma(\mathcal{U}^{[m]\setminus S})|$ and the binomial distribution $\mathsf{Bin}(n, \gamma)$ assign similar mass to each of these intervals. Moreover, our continuity result ([32, Proposition 4.21]) implies the mass in a fixed interval is almost uniformly distributed. Thus for most weights w, $|F_\gamma(\mathcal{U}^{[m]\setminus S})|$ assigns roughly the same mass to w as $\mathsf{Bin}(n, \gamma)$ does. Summing over all weights provides an upper bound on the total variation distance between the two weight distributions. We remark that there is a slight complication in the case of $\gamma = 1/2$, as there [32, Proposition 4.21] only lets us compare weights that are an even distance apart. However, by further subdividing each interval we consider into its even and odd parts, we are able to carry out a similar argument as before.

Finally, we recall that $f(\mathcal{U}^m)$ is a mixture of the distributions $F_\gamma(\mathcal{U}^{[m]\setminus S})$. Since the weight distribution of each $F_\gamma(\mathcal{U}^{[m]\setminus S})$ is close to the weight distribution corresponding to one from Theorem 2, their mixture $|f(\mathcal{U}^m)|$ naturally is close to a mixture of those weight distributions (see [32, Lemma 4.24]). We conclude by recalling that Equation (1) implies it was sufficient to classify the output weight distribution.

2.1.5 Mixing Weights

The above argument shows that $f(\mathcal{U}^m)$ is close to a mixture $\mathcal{M}$ of the form specified in Theorem 2, but without the additional control on the mixing weights. The reason for this shortcoming is that the threshold A at which we consider an input bit "high degree" depends not just on d, but also on ε, and so the mixing weights in the obtained mixture also depend on ε. If we instead chose a threshold that only depended on d, we would not have been able to obtain an effective error bound in the Kolmogorov distance step (i.e., [32, Proposition 4.16]) when $\gamma = 1/2$.

The key observation is that the arguments of Step 1 only require conditioning on input bits of degree larger than $n/O_d(1)$. In other words, if we perform some different conditioning κ on the set of input bits $S' \subseteq [m]$ above the weaker threshold $n/O_d(1)$, the distributions generated by the restricted functions will still have output weights which strongly concentrate around some integer multiple of $n/2^d$. Moreover, the mixing weights in this setting *are* integer multiples of $2^{-O_d(1)}$.

It remains to argue that the mixing weights on $\mathcal{M}$, the mixture we proved $f(\mathcal{U}^m)$ is close to by the stronger conditioning ρ, correspond to the mixing weights derived from the weaker conditioning κ (see [32, Lemma 4.26]). Since both conditionings produce mixtures of the same distribution, we have

$$\sum_\gamma \Pr_\rho[\gamma_\rho = \gamma] \cdot \mathop{\mathbb{E}}_{\rho:\gamma_\rho=\gamma} f_\rho(\mathcal{U}^{[m]\setminus S}) = f(\mathcal{U}^m) = \sum_\gamma \Pr_\kappa[\gamma_\kappa = \gamma] \cdot \mathop{\mathbb{E}}_{\kappa:\gamma_\kappa=\gamma} f_\kappa(\mathcal{U}^{[m]\setminus S'}).$$

By standard concentration bounds, the probability that $x \sim f(\mathcal{U}^m)$ has Hamming weight close to γn is almost entirely determined by the mass assigned to the distributions whose output weight concentrates around γ. In other words, we know $\Pr_\rho[\gamma_\rho = \gamma] \approx \Pr_\kappa[\gamma_\kappa = \gamma]$

for each γ. By our earlier analysis and the assumption that $f(\mathcal{U}^m)$ is close to $\mathcal{D}$, it must be that the mixing weight on $\mathsf{Bin}(n,\gamma)$ in $\mathcal{M}$ is roughly $\Pr_\rho[\gamma_\rho = \gamma]$, so it must also be close to $\Pr_\kappa[\gamma_\kappa = \gamma]$, which has the form we sought.

The analysis in the case of $\gamma = 1/2$ is similar, but as always, we need to keep track of the output weight's parity. Here, the analogous equivalence is

$$\Pr_\rho[\gamma_\rho = 1/2 \wedge |f(\mathcal{U}^m)| \text{ is even}] \approx \frac{1}{2^{|S'|}} \sum_{\kappa:\gamma_\kappa = 1/2} \Pr_{x \sim \mathcal{U}^{[m]\setminus S'}}[|f_\kappa(x)| \text{ is even}]. \tag{2}$$

Since f is d-local with n output bits, the parity of $|f_\kappa(x)|$ can be represented as a degree-d $\mathbb{F}_2$-polynomial with $O_d(n)$ monomials. Thus, the mixing weight for the uniform distribution over n-bit strings of even Hamming weight in $\mathcal{M}$ must be close to the right-hand side of Equation (2), which is the precise sense in which the mixing weights are "determined by low-degree $\mathbb{F}_2$-polynomials". The odd parity case is essentially identical. This concludes the proof overview of Theorem 2.

2.2 Technical Comparison to Prior Works

We now briefly survey the approaches of related prior works. To avoid an overly verbose digression, we attempt to focus only on the most relevant papers and do not hope to be exhaustive. Similarly, we restrict our attention within the mentioned works to the results and techniques most pertinent to our own; most, if not all, of the papers discussed contain a number of other interesting results.

The study of the complexity of sampling goes back to at least the 1980s in the influential work of Jerrum, Valiant, and Vazirani [29]. In the context of shallow circuits, several clever sampling constructions (including the one in the introduction) were devised in [5, 10, 28], while the first serious treatment of the complexity of sampling (with shallow circuits) appeared in [40]. There, Viola proved an assortment of sampling-related results, but we will confine our attention to two on the hardness of approximately generating $\mathcal{D}_{n/2}$, the uniform distribution over n-bit strings of Hamming weight $n/2$. (This specific choice is purely for simplicity, and both results also apply to the setting of $\mathcal{D}_{\alpha n}$ for $\alpha \in (0,1)$.)

The first result we will mention [40, Theorem 1.6] is an unconditional lower bound of $2^{-O(d)} - O(1/n)$ on the distance between $f(\mathcal{U}^m)$, where $f\colon \{0,1\}^m \to \{0,1\}^n$ is a d-local[4] function, and $\mathcal{D}_{n/2}$. The proof, much like our own (see Step 2), relies on a Kolmogorov distance bound obtained from a k-wise independence assumption [19, 23]. If the output distribution $f(\mathcal{U}^m)$ is k-wise independent for some large integer k, then [23, Theorem I.5] implies it has the wrong Hamming weight with constant probability, and we are done. Otherwise, there exists a k-tuple $T \subseteq [n]$ of output bits that are not uniformly distributed over $\{0,1\}^k$. In this case, one observes that the probability assigned to any element of $\{0,1\}^k$ is an integer multiple of 2^{-kd}, so the marginal distribution over T must be at least 2^{-kd}-far from uniform. The proof concludes by noting that the marginal distribution of $\mathcal{D}_{n/2}$ onto T is very close to uniform. Note that we use similar granularity ideas in the proof of [32, Proposition 4.5] (in Step 1).

The second result [40, Theorem 1.3] provides a much stronger lower bound of $1 - 1/\mathsf{poly}(n)$ for the distance between $f(\mathcal{U}^m)$ and $\mathcal{D}_{n/2}$ for any $O(\log n)$-local function f, but is *conditional* on the input size m not substantially exceeding the information-theoretic minimum required to generate the target distribution. The proof begins by partitioning the input bits u of f as $u = (x, y)$ and expressing

[4] In fact, the result is proven for an adaptive version of locality.

$$f(u) = f(x, y) = h(y) \circ g_1(x_1, y) \circ \cdots \circ g_s(x_s, y),$$

where each g_i may depend arbitrarily on y but only on the single bit x_i of x. Additionally, each $g_i(x_i, y) \in \{0,1\}^{O(d)}$. By a greedy approach, this can be accomplished with $s \geq \Omega(n/d^2) = n/\mathsf{polylog}(n)$. The argument proceeds by conditioning on certain input bits (in this case y) and splitting into two scenarios depending on the amount of concentration; this strategy has been employed by several later works, including this one.

In the "concentrated" case where at least $\sqrt{n}$ many g_i's become fixed, the output distribution of $f(x, y)$ has small support. Even taking the union over all choices of such y, one finds that many elements in the support of $\mathcal{D}_{n/2}$ cannot be obtained. Note that this is where the input length restriction originates. Alternatively, at least $s - \sqrt{n} = n/\mathsf{polylog}(n)$ of the g_i's take multiple values. One would like to apply standard anticoncentration results to argue the output weight is too spread out, but as noted earlier in the proof overview (see Step 3), it may be that the Hamming weight of a g_i is fixed, even if the output is not. Viola circumvents this issue by adding an additional "test" to determine whether at most $2\sqrt{n}$ many g_i's can output the all zeros string; in this case, taking two or more values implies the weight cannot be fixed, and thus anticoncentration inequalities can be applied.

We now turn to the recent work of Filmus, Leigh, Riazanov, and Sokolov [21], who addressed the case of $\mathcal{D}_k$ for $k = o(n)$. They proved [21, Theorem 1.2] that $\widetilde{\Omega}(\log(n/k))$ bits of locality (even adaptively chosen) are required to generate $\mathcal{D}_k$ to constant error. Moreover, the same result holds for the uniform distribution over strings whose Hamming weight lies in a set S, where $\max_{s \in S} s = k$. Their proof first reduces to the case of considering $\mathcal{D}_1$, and then proceeds via a hitting set vs independent set dichotomy. This can again be viewed as fitting into the concentration vs anticoncentration paradigm. If there are $O(d2^d)$ input bits that together affect every output bit, then one can afford to condition on them and induct on the remaining $(d-1)$-local sub-distributions. Bounds from these distributions can eventually be recombined to deduce bounds on the full distribution using some version of a union bound, such as [32, Lemma 3.4]. Otherwise, there are $\Omega(2^d)$ independent output bits, and one can argue it is very likely that two of these bits evaluate to 1, since each (nonzero) output bit is 1 with probability at least 2^{-d}. Interestingly, the authors are able to use the *robust sunflower lemma* [36, 4] from extremal combinatorics to improve their quantitative bounds; we refer the reader to their original paper for details.

Also recently, strong unconditional locality lower bounds for sampling $\mathcal{D}_{\alpha n}$ where α is a non-dyadic rational were proven independently in [44, 30]. At a very high level, both proofs follow from the observation (recorded in [39]) that the marginal distribution on any output bit of a d-local function f must have probabilities that are integer multiples of 2^{-d}, which cannot accurately approximate (say) $1/3$, corresponding to the marginal distribution for $\mathcal{D}_{n/3}$. We begin by describing the approach of Viola [44, Theorem 25], continuing to focus on the case of $\alpha = 1/3$ for simplicity. First assume by contradiction $\left\| f(\mathcal{U}^m) - \mathcal{D}_{n/3} \right\|_{\mathsf{TV}} \leq 1 - \varepsilon$ for some ε to be determined. The proof proceeds by arguing that there must exist a large subset $R \subseteq \{0,1\}^m$ of the inputs, such that $f(\mathcal{U}(R))$, f evaluated on inputs drawn uniformly at random from R, lands entirely in the support of $\mathcal{D}_{n/3}$ and has almost full min-entropy.[5] The so-called *fixed-set lemma* of Grinberg, Shaltiel, and Viola [25] then implies that R can be further restricted to a large subset $R' \subseteq R$ whose uniform distribution is nearly indistinguishable by d-local functions from a product distribution where each bit is either

[5] Recall the *min-entropy* of a random variable X is $\log(1/\max_{x \in \mathsf{supp}(X)} \Pr[X = x])$.

fixed or uniform. Then the large min-entropy of $f(\mathcal{U}(R))$ (and thus $f(\mathcal{U}(R'))$) combined with the fact that the distribution is contained in the support of $\mathcal{D}_{n/3}$ implies there must be at least one output bit whose marginal distribution is extremely close to 1/3. This, however, contradicts the fact that 1/3 cannot accurately be approximated by an integer multiple of 2^{-d}.

The proof in [30] instead operates very similarly to the details of Step 3. Through a graph-theoretic lemma [30, Corollary 4.8], they show that there must be a set of at most $r/2^d$ many input bits upon whose conditioning results in r many independent output bits for $r \approx n/2^{d^2}$. As previously noted, each of these output bits inevitably incurs some error from mismatched marginal distributions, and these errors can be aggregated via standard concentration inequalities (as in [32, Lemma 3.3]). Bounds from the individual sub-distributions can once again be recombined via a union bound result like [32, Lemma 3.4].

The work [30] also contains locality bounds for $\mathcal{D}_{\alpha n}$ when α is a dyadic rational. The proof has strong similarities to the non-dyadic case, except that now the marginal distributions on output bits of f do not necessarily disagree with those of the target distribution. The authors overcome this issue by turning to the familiar concentration vs. anticoncentration paradigm. The details have almost entirely been spelled out already in Step 3, as our analysis is nearly identical at that part. We briefly note that the quantitative dependencies here are much worse than in the non-dyadic case, essentially for the reason that one needs to obtain a much richer structure than simply independent output bits, and so the analogous graph-theoretic lemma in this case has poor bounds.[6]

Finally, let us mention the predecessor of this work [31], which classifies what uniform symmetric distributions can be sampled by functions of bounded locality, confirming a conjecture of [21]. In broad strokes, it follows similarly to the proof overview, so we will be brief. Through a strategic conditioning of input bits, one may reduce to the case where no input bits have large degree and almost all the output bits are k-wise independent. Then one can show the resulting Hamming weight distribution is close in Kolmogorov distance to the binomial distribution, and it satisfies a continuity-type property. Combining these steps together, one classifies the Hamming weight distribution, which is enough to classify the actual distribution, since we assume $f(\mathcal{U}^m)$ is close to a uniform symmetric distribution. Much of the present work is an extension of the ideas in [31] to the general symmetric case, although there are several parts (e.g., much of the analysis in Step 1 and the finer control on the mixing weights) that require substantially new ideas.

We conclude by noting that several works [35, 6, 42, 43, 11] prove hardness results against the more powerful classes of AC^0 circuits or read-once branching programs. There is a reasonable overlap in the techniques used with those discussed above, but the proofs of all these results rely on the special properties of certain pseudorandom objects like good codes [35, 6, 11] or extractors [41, 43].

References

1 M. 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.

[6] Moreover, the bounds in this graph-theoretic lemma are essentially best possible; see [30, Appendix A].

3 Maryam Aliakbarpour, Eric Blais, and Ronitt Rubinfeld. Learning and testing junta distributions. In *Conference on Learning Theory*, pages 19–46. PMLR, 2016. URL: http://proceedings.mlr.press/v49/aliakbarpour16.html.

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

5 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.

6 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.

7 Adam Bene Watts and Natalie Parham. Unconditional Quantum Advantage for Sampling with Shallow Circuits. In Shubhangi Saraf, editor, *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, volume 362 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:12, Dagstuhl, Germany, 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2026.17.

8 Lucien Birgé. Estimating a density under order restrictions: Nonasymptotic minimax risk. *The Annals of Statistics*, pages 995–1012, 1987.

9 Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4):929–965, 1989. doi:10.1145/76359.76371.

10 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.

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

12 Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, Shachar Lovett, and David Zuckerman. XOR lemmas for resilient functions against polynomials. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 234–246, 2020. doi:10.1145/3357713.3384242.

13 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. doi:10.1145/2897518.2897528.

14 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. doi:10.1109/FOCS.2016.27.

15 Constantinos Daskalakis, Ilias Diakonikolas, Ryan O'Donnell, Rocco A Servedio, and Li-Yang Tan. Learning sums of independent integer random variables. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 217–226. IEEE, 2013. doi:10.1109/FOCS.2013.31.

16 Constantinos Daskalakis and Gautam Kamath. Faster and sample near-optimal algorithms for proper learning mixtures of Gaussians. In *Conference on Learning Theory*, pages 1183–1213. PMLR, 2014. URL: http://proceedings.mlr.press/v35/daskalakis14.html.

17 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.

18 Ilias Diakonikolas. Learning structured distributions. *Handbook of Big Data*, 267:10–1201, 2016.

19 Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010. doi:10.1137/100783030.

20 Lutz Dümbgen and Kaspar Rufibach. Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, pages 40–68, 2009.

21 Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. Sampling and certifying symmetric functions. In *Approximation, Randomization, and Combinatorial Optimization. (APPROX/RANDOM)*, 2023.

22 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.

23 Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *2010 IEEE 25th Annual Conference on Computational Complexity*, pages 223–234. IEEE, 2010. doi:10.1109/CCC.2010.29.

24 Daniel Grier, Daniel M. Kane, Jackson Morris, Anthony Ostuni, and Kewen Wu. Quantum Advantage from Sampling Shallow Circuits: Beyond Hardness of Marginals. In Shubhangi Saraf, editor, *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, volume 362 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 73:1–73:14, Dagstuhl, Germany, 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2026.73.

25 Aryeh Grinberg, Ronen Shaltiel, and Emanuele Viola. Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 956–966. IEEE, 2018. doi:10.1109/FOCS.2018.00094.

26 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.

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

28 Russell Impagliazzo and Moni Naor. Efficient cryptographic schemes provably as secure as subset sum. *Journal of cryptology*, 9(4):199–216, 1996. doi:10.1007/BF00189260.

29 Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical computer science*, 43:169–188, 1986. doi:10.1016/0304-3975(86)90174-X.

30 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.

31 Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locally sampleable uniform symmetric distributions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1807–1816, 2025. doi:10.1145/3717823.3718243.

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

33 Michael Kearns, Yishay Mansour, Dana Ron, Ronitt Rubinfeld, Robert E Schapire, and Linda Sellie. On the learnability of discrete distributions. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 273–282, 1994. doi:10.1145/195058.195155.

34 Bruce G Lindsay. Mixture models: Theory, geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, pages i–163. JSTOR, 1995.

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

37 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.

38 Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984. doi:10.1145/1968.1972.

39 Emanuele Viola. Bit-probe lower bounds for succinct data structures. *SIAM Journal on Computing*, 41(6):1593, 2012.

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

41 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.

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

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

44 Emanuele Viola. New sampling lower bounds via the separator. In *38th Computational Complexity Conference (CCC 2023)*, 2023. doi:10.4230/LIPIcs.CCC.2023.26.

45 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.

46 Yannis G Yatracos. Rates of convergence of minimum distance estimators and kolmogorov's entropy. *The Annals of Statistics*, 13(2):768–774, 1985.

47 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.