Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 180 (2024)

Locally Sampleable Uniform Symmetric Distributions

Daniel M. Kane* Anthony Ostuni† Kewen Wu‡

Abstract

We characterize the power of constant-depth Boolean circuits in generating uniform symmetric distributions. Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a Boolean function where each output bit of f depends only on $O(1)$ input bits. Assume the output distribution of f on uniform input bits is close to a uniform distribution $\mathcal{D}$ with a symmetric support. We show that $\mathcal{D}$ is essentially one of the following six possibilities: (1) point distribution on 0^n, (2) point distribution on 1^n, (3) uniform over $\{0^n, 1^n\}$, (4) uniform over strings with even Hamming weights, (5) uniform over strings with odd Hamming weights, and (6) uniform over all strings. This confirms a conjecture of Filmus, Leigh, Riazanov, and Sokolov (RANDOM 2023).

1 Introduction

Despite being one of the simplest models of computation, NC^0 circuits (i.e., Boolean circuits of constant depth and bounded fan-in) elude a comprehensive understanding. Even very recently, the model has been the subject of active research on the range avoidance problem [RSW22, GLW22, GGNS23], quantum advantages [BGK18, WKST19, BGKT20, WP23, KOW24], proof verification [GGH+07, BDK+13, KLMS16], and more.

Pertinent to this paper is the study of the sampling power of NC^0 circuits. While the general problem was considered at least as early as [JVV86], interest in the NC^0 setting has seen a strong uptick lately [Vio12b, LV11, BIL12, DW12, Vio16, Vio20, GW20, CGZ22, Vio23, FLRS23, KOW24, SS24]. At a high level, it considers what distributions can be (approximately) produced by simple functions on random inputs. More formally, let $f(\mathcal{U}^m)$ denote the distribution resulting from applying an NC^0 function $f\colon \{0,1\}^m \to \{0,1\}^n$ to a random string drawn from $\mathcal{U}^m$, the uniform distribution over $\{0,1\}^m$. Typically, m is viewed as being arbitrarily large and n is the parameter of interest. Then the goal is to analyze the distance between $f(\mathcal{U}^m)$ and some specific distribution. Aside from being inherently interesting, this question has played a crucial role in applications ranging from data structure lower bounds [Vio12b, LV11, BIL12, Vio20, CGZ22, Vio23, KOW24] to pseudorandom generators [Vio12a, LV11, BIL12] to extractors [Vio12c, DW12, Vio14, CZ16, CS16] to coding theory [SS24].

One recurring class of distributions in this line of work is uniform symmetric distributions (i.e., uniform distributions over a symmetric support). Indeed, these are precisely the distributions that arise in an elegant connection to succinct data structures (see [Vio12b, Claim 1.8]), for example. Moreover, this seemingly simple class is already rich enough to allow surprisingly powerful results. For instance, NC^0 circuits can sample the uniform distribution over the preimage $\mathrm{PARITY}^{-1}(0)$

[*] University of California, San Diego. Email: dakane@ucsd.edu. Supported by NSF Medium Award CCF-2107547 and NSF CAREER Award CCF-1553288.

[†] University of California, San Diego. Email: aostuni@ucsd.edu.

[‡] University of California, Berkeley. Email: shlw_kevin@hotmail.com. Supported by a Sloan Research Fellowship and NSF CAREER Award CCF-2145474.

ISSN 1433-8092

(and PARITY$^{-1}(1)$), despite a celebrated result of Håstad [Hås86] proving that more powerful AC^0 circuits require an exponential number of gates to *compute* PARITY. Perhaps more surprisingly, the strategy to sample a uniform random string with even Hamming weight is extremely simple: map the uniform random bits $x_1, \dots, x_n$ to $x_1 \oplus x_2, x_2 \oplus x_3, \dots, x_n \oplus x_1$ [Bab87, BL87].

A number of notable prior results already rule out specific distributions from being accurately sampled by NC^0 circuits. For example, let $\mathcal{D}_{\{k\}}$ denote the uniform distribution over all n-bit strings of Hamming weight k. The influential early paper of [Vio12b] showed that such shallow circuits could not accurately sample $\mathcal{D}_{\{k\}}$ for $k = \Theta(n)$ under certain assumptions about the input length or accuracy tolerance; recent works [FLRS23, Vio23, KOW24] have eliminated the need for these assumptions. Additionally, a number of results are known for uniform symmetric distributions over multiple Hamming weights, such as the case of exclusively tail weights [FLRS23], all weights divisible by q for fixed $3 \le q \ll \sqrt{n}$ [KOW24], and all weights above $n/2$ [GGH$^+$07, Vio12b, FLRS23] (see also [WP23]).

Despite much effort, the previous body of work proceeds in a somewhat ad-hoc fashion, with techniques tailored to rule out specific cases. However, an exciting recent work by Filmus, Leigh, Riazanov, and Sokolov [FLRS23] gave the following bold conjecture about the capabilities of NC^0 circuits for sampling distributions, unifying prior results.

Conjecture 1.1 ([FLRS23, Conjecture 1.1])**.** *For every $d \in \mathbb{N}, \varepsilon \in (0,1)$, and large enough n, if $f\colon \{0,1\}^m \to \{0,1\}^n$ is computable by an NC^0 circuit of depth at most d and $f(\mathcal{U}^m)$ is ε-close (in total variation distance) to a uniform symmetric distribution, then $f(\mathcal{U}^m)$ is $O(\varepsilon)$-close to one of the following six distributions:*

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

All six distributions can be sampled (exactly) by functions whose output bits each depend on at most two input bits. Hence one may informally view the conjecture as asserting that more input dependencies do not substantially increase the ability of NC^0 circuits to generate uniform symmetric distributions.

In this work, we confirm the conjecture of [FLRS23].

Theorem 1.2 (Consequence of Theorem 4.1)**.** *Conjecture 1.1 is true.*

We emphasize that the implicit constant in the distance $O(\varepsilon)$ in our result has *no* dependence on the depth d. Additionally, note that this result is optimal up to that implicit constant. We include a more thorough discussion of our result's tightness in Section 4, where we present a quantitative version of Theorem 1.2 parametrized by the locality (i.e., number of input bits each output bit depends on) of f. The following corollary is immediate.

Corollary 1.3. *For sufficiently large n, the only uniform symmetric distributions over $\{0,1\}^n$ exactly sampleable by NC^0 functions are the six distributions in Conjecture 1.1.*

As a contrasting example to the limitation given by Theorem 1.2, consider the next simplest class of circuits commonly studied: AC^0. Up to some exponentially small error, they are able to sample the uniform distribution over permutations of $[n]$ [MV91, Hag91]. Thus by sampling $1^w 0^{n-w}$ for the appropriate distribution over weights w accepted (or rejected) by a symmetric function f, one can apply a randomly sampled permutation to output the uniform distribution over $f^{-1}(1)$ (or $f^{-1}(0)$) [Vio12b, Lemma 4.3].

Paper Organization. We provide a proof overview of Theorem 1.2 in Section 2. Preliminary definitions and results are given in Section 3. The full proof of our main result is in Section 4, with some technical proofs deferred to Section 5 and Appendix A.

2 Proof Overview

Our starting point is similar to many past works [Vio12b, Vio20, FLRS23, Vio23, KOW24]: we reduce an arbitrary function (computable by an NC^0 circuit) to a collection of structured functions which are more amenable to analysis. Our results then follow by lifting insights from these structured functions to our original function.

It will be convenient to work with the abstraction of *locality*. We say a function $f\colon \{0,1\}^m \to \{0,1\}^n$ is *d-local* if every output bit depends on at most d input bits. Observe that the class of d-local functions captures functions computable by Boolean circuits of depth $O(\log d)$ and bounded fan-in. In particular, constant locality functions are equivalent to those computable by NC^0 circuits. Henceforth, let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function. For simplicity, we hide minor factors in the following discussion.

2.1 A Structured Decomposition

We will use the "graph elimination" reduction strategy of [KOW24], rephrased slightly more naturally in the language of hypergraphs. Let G be the hypergraph on the output bits $[n]$ with an edge for each input bit connecting all of the output bits that depend on it. By assumption, each vertex is contained in at most d edges (i.e. G has maximum degree d). Using standard hypergraph terminology, we define the *neighborhood* of a vertex v to be the set of all vertices sharing an edge with v. Furthermore, we call two neighborhoods N_1, N_2 *connected* if there exist vertices $v_1 \in N_1, v_2 \in N_2$ contained in the same edge.

By [KOW24, Proposition 5.20], there exists a set of $o(n)$ edges whose deletion results in a graph with $\Omega_d(n)$ non-connected neighborhoods of size $O_d(1)$. In other words, there always exists a choice of a few input bits whose conditioning upon decomposes f into a mixture of subfunctions with substantially independent output bits.

This independence is crucial in ruling out the sampleability of various distributions by these structured subfunctions. For example, [KOW24] used the following win-win argument to prove strong bounds on the distance between any distribution sampleable by a local function and the uniform distribution over n-bit strings of Hamming weight $k = \Theta(n)$, denoted $\mathcal{D}_{\{k\}}$. If the marginal distributions of most independent neighborhoods noticeably differ from the corresponding marginals of $\mathcal{D}_{\{k\}}$, then the errors can be combined together via a straightforward concentration bound argument [KOW24, Lemma 4.2].

Otherwise, the marginal distributions of most independent neighborhoods $N(v_1), \dots, N(v_r)$ closely match the marginals of $\mathcal{D}_{\{k\}}$. Hence by conditioning on all the input bits that do not affect the output bits $v_1, \dots, v_r$, the weight of the output becomes a sum of well-behaved independent integer random variables. From this property, one can show ([KOW24, Claims 5.16 & 5.23]) that with high probability many of these random variables are not constant (or even constant modulo

q for $q \geq 3$), in which case anticoncentration inequalities (e.g., [Ush86, Theorem 3] or [KOW24, Lemma 3.7]) imply no specific output weight can be obtained with good probability. Hence the subfunctions cannot accurately sample $\mathcal{D}_{\{k\}}$, so (by a union bound argument) neither can their mixture.

Note that the distribution $\mathcal{D}_{\{k\}}$ is a special kind of uniform symmetric distribution (i.e., uniform distribution over a symmetric support). In this work, we need to handle more general ones; however, many of the same ideas will drive our analysis.

2.2 Classification of Locally Sampleable Uniform Symmetric Distributions

Now we show how to handle a general uniform symmetric distribution and obtain our classification result. For convenience, we use a non-empty set $\Psi \subseteq \{0, 1, \ldots, n\}$ to denote the acceptable Hamming weights and use $\mathcal{D}_\Psi$ to denote the uniform distribution over strings of Hamming weights in Ψ. Then our goal is to show that local functions cannot approximate $\mathcal{D}_\Psi$ unless Ψ is $\{0\}$ (the point distribution on 0^n), $\{n\}$ (the point distribution on 1^n), $\{0, n\}$ (uniform over $\{0^n, 1^n\}$), $\{0, 2, 4, \ldots\}$ (uniform over strings with even parity), $\{1, 3, 5, \ldots\}$ (uniform over strings with odd parity), or $\{0, 1, \ldots, n\}$ (uniform over all strings). We will often refer to the corresponding $\mathcal{D}_\Psi$ as the six special distributions.

Let $s \in \Psi$ be an element closest to the middle weight $n/2$. Note the majority of the mass of $\mathcal{D}_\Psi$ is supported on strings roughly as close to $n/2$ as s is. Informally, we view $\mathcal{D}_\Psi$ as either $\mathcal{D}_{\{s\}}$ (uniform over the Hamming slice of weight s) or $\frac{1}{2}\mathcal{D}_{\{s\}} + \frac{1}{2}\mathcal{D}_{\{n-s\}}$ (uniform over the Hamming slices of weight s and $n - s$). Then the above six locally sampleable distributions can be classified by s: either it is the endpoint (i.e., s equals 0 or n) or it is the middle point (i.e., s roughly equals $n/2$). Our proof follows this intuition. If $|s - n/2| > n^{2/3}$, we will show that it must be the case of $s \in \{0, n\}$. Otherwise $|s - n/2| \leq n^{2/3}$, and we will show that it must be the case that Ψ is effectively all-even, all-odd, or everything.

The $|s - n/2| > n^{2/3}$ Case. This far in the tails the binomial distribution decays rapidly, so it is not difficult to show (Claim 4.5) that half the mass of $\mathcal{D}_\Psi$ is supported on $O(n^{1/3})$ different weights. However, [FLRS23, Theorem 1.2] and [KOW24, Theorem 5.10] show that any distribution $\mathcal{D}_{\{k\}}$ other than $k = 0$ or n has total variation distance $1 - O_d(n^{-1/2})$ from any d-locally sampleable uniform symmetric distribution. Therefore, by a union bound argument (Lemma 3.3), $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} > 1/3$ unless Ψ is one of $\{0\}, \{n\}, \{0, n\}$. We now turn to the more challenging case.

The $|s - n/2| \leq n^{2/3}$ Case. In this case we note that $\mathcal{D}_\Psi$ is reasonably close to the uniform distribution. (In particular, its restriction to any constant number of output bits will be quite close.) Thus, the framework explained in Subsection 2.1 is now applicable. That is, we remove a modest number of edges from the corresponding hypergraph to obtain one with $\Omega_d(n)$ many non-adjacent neighborhoods of size $O_d(1)$. This means that after conditioning on the removed inputs, we will have many sets of output bits that are independent of each other.

From here we again split into cases. If many of these neighborhoods are far from uniform, our distribution must be far from the uniform-ish $\mathcal{D}_\Psi$. This part of the mixture will contribute nearly equally to the distance from $\mathcal{D}_\Psi$ and from the closest special distribution. Otherwise, we have many roughly unbiased neighborhoods $N(v_1), \ldots, N(v_r)$. As sketched in Subsection 2.1, if we further condition on all the inputs that do not affect $v_r, \ldots, v_r$, then the total weight of the output becomes a sum of independent random variables. Using the fact that the overall distribution on these neighborhoods is uniform, we can show that with high probability many of these random variables are not constant, and in fact for each $q > 2$, many are not constant modulo q. This allows us to show (Proposition 4.9) that our distribution over weights must be continuous. In particular, the probability that we have weight x and the probability that we have weight $x+\Delta$ for even $\Delta \in \mathbb{Z}$ must differ by at most $O_d(|\Delta|/n)$ (Theorem 5.10). This implies that if $f(\mathcal{U}^m)$ is close to $\mathcal{D}_\Psi$, then Ψ cannot have many instances where it contains x but not $x + \Delta$ or vice versa. By pairing up the present x's with missing ones, we attempt to show that the distance between $f(\mathcal{U}^m)$ with $\mathcal{D}_\Psi$ is comparable to the distance between $\mathcal{D}_\Psi$ and the nearest special distribution.

Unfortunately, there are two issues with this argument. The first is that our bounds on the difference between probabilities of x and $x + \Delta$ have error terms that are inverse polynomial in n.[1] This makes them useless if Ψ contains all of the elements of $[n]$ within $\Omega_d(n)$ of $n/2$. In this case, however, the coordinates of $f(\mathcal{U}^m)$ must match moments with the uniform distribution (Claim 4.12). Thus, a careful argument (Proposition 4.13) can show that $f(\mathcal{U}^m)$ cannot be much closer to $\mathcal{D}_\Psi$ than the relevant special distribution, as a decent fraction of any mass removed from the first "missing" Hamming slice $\mathcal{D}_{\{s\}}$ of $\mathcal{D}_\Psi$ must be placed at even more extreme weights rather than closer to the middle.

The other issue is simply the d-dependence of our continuity bounds. A naive application would only be enough to show that $f(\mathcal{U}^m)$ is $O_d(\varepsilon)$-close to the nearest special distribution. If, for example, $\Psi = \{0, 1, \dots, n/2\}$, we would only be able to find $\Omega_d(\sqrt{n})$ many pairs of elements to compare and could only show that $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} = \Omega_d(1)$. To fix this, we use the fact that our output distribution must match moments with the uniform distribution to show (Proposition 5.1) that most of $f(\mathcal{U}^m)$'s mass has a weight distribution close in Kolmogorov distance[2] to the binomial distribution, even accounting for parity.

More specifically, we use an inductive argument (Lemma 5.5) to decompose the input space $\{0,1\}^m$ into subcubes. Aside from some negligible mass, each subcube falls into two types. A Type-I subcube C satisfies that the output bits of f on the uniform distribution over C are nearly k-wise independent (for an appropriately chosen parameter k) with no input bit affecting too many output bits. Such cubes are useful, as we can leverage known results about k-wise independence fooling threshold functions (Theorem 5.3) and the structure of low-degree $\mathbb{F}_2$-polynomials (Theorem 5.4) to prove our desired Kolmogorov property in this special case (Lemma 5.2).

Alternatively, Type-II subcubes satisfy that f applied to the uniform distribution over the union of all Type-II subcubes is far from $\mathcal{D}_\Psi$. These correspond to the aforementioned neighborhoods whose distributions are far from uniform, and whose contribution to the distance between $f(\mathcal{U}^m)$ and $\mathcal{D}_\Psi$ roughly matches that of the contribution to the distance between $f(\mathcal{U}^m)$ and the nearest special distribution. This is the small fraction of $f(\mathcal{U}^m)$'s mass that does not satisfy the desired Kolmogorov distance property.

Combining with our continuity result, we obtain a local limit theorem (Theorem 4.7) stating that $f(\mathcal{U}^m)$ must be close to some mixture $\mathcal{M}$ of the uniform distribution over strings with even Hamming weight and the uniform distribution over strings with odd Hamming weight. Furthermore, observe that for any two strings, $\mathcal{D}_\Psi$ either assigns them the same mass, or assigns one of them zero mass. Since $f(\mathcal{U}^m)$ is close to both $\mathcal{D}_\Psi$ and $\mathcal{M}$, the triangle inequality guarantees $\mathcal{D}_\Psi$ is close to $\mathcal{M}$. Then by the previous observation, $\mathcal{M}$ must be close to either the uniform distribution over strings with even Hamming weight, the uniform distribution over strings with odd Hamming weight, or the uniform distribution over all strings (Corollary 4.10). In other words, we can extract a similar conclusion to our limit limit theorem for the nearest special distribution $\mathcal{D}$: $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) + \delta$ for any $\delta > 0$, so long as n is sufficiently large in terms of δ and d.

[1] The following issue persists even if we sharpen the bounds to have inverse exponential error (as was done in a previous version https://arxiv.org/pdf/2411.08183v1).

[2] Recall the *Kolmogorov distance* between two distributions $\mathcal{P}$, $\mathcal{Q}$ is given by $\sup_{t\in\mathbb{R}} |\mathbf{Pr}_{x\in\mathcal{P}}[x > t] - \mathbf{Pr}_{y\in\mathcal{Q}}[y > t]|$.

At this point, it is tempting to apply the above conclusion with $\delta = O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}})$ to conclude the proof. However, if δ is particularly small, we may not be able to guarantee n is large enough to apply the result. Hence, we can only assume $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \le O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) + \delta$ for some sufficiently small constant $\delta > 0$.

It remains to handle the special case of $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}}$ less than some sufficiently small constant. Here, we can take advantage of strong structural properties, such as matching many moments with the uniform distribution (Claim 4.12) and having the support of $f(\mathcal{U}^m)$ entirely contained in the support of $\mathcal{D}$ (Claim 4.11). These additional properties allow us to deduce that unless our desired result already holds, the weight distribution $|f(\mathcal{U}^m)|$ has noticeably too little mass on weights around $n/2$. We can now run the previously discussed pairing argument, where we find nearby $x \in \Psi$ and $x + \Delta \notin \Psi$ for even $\Delta \in \mathbb{Z}$, except now restricted to a small interval around $n/2$. Crucially, the mass $\mathcal{D}_\Psi$ assigns to strings with weight in that interval exceeds the error from the continuity bound, removing the dependence on d.

Previous Version. A previously posted version of this paper provided a weaker form of Theorem 1.2 where $f(\mathcal{U}^m)$ is only $O_d(\varepsilon)$-close to one of the six special distributions. The approach taken was a more technically involved form of the original pairing argument described above. In particular, the key improvement in this version that allows us to remove the dependence on d from the distance bound is the local limit theorem (Theorem 4.7).

3 Preliminaries

For a positive integer n, we use $[n]$ to denote the set $\{1, 2, \dots, n\}$. We use $\mathbb{R}$ to denote the set of real numbers, use $\mathbb{N} = \{0, 1, 2, \dots\}$ to denote the set of natural numbers, and use $\mathbb{Z}$ to denote the set of integers. For a binary string x, we use $|x|$ to denote its Hamming weight. We use $\log(x)$ and $\ln(x)$ to denote the logarithm with base 2 and $e \approx 2.71828\dots$ respectively.

Asymptotics. We use the standard $O(\cdot), \Omega(\cdot), \Theta(\cdot)$ notation, and emphasize that in this paper they only hide universal positive constants that do not depend on any parameter. Occasionally we will use subscripts to suppress a dependence on particular variable (e.g., $O_d(1)$).

Probability. Let $\mathcal{P}$ be a (discrete) distribution. We use $x \sim \mathcal{P}$ to denote a random sample x drawn from the distribution $\mathcal{P}$. If $\mathcal{P}$ is a distribution over a product space, then we say $\mathcal{P}$ is a product distribution if its coordinates are independent. In addition, let S be a non-empty set. If S indexes $\mathcal{P}$, we use $\mathcal{P}[S]$ to denote the marginal distribution of $\mathcal{P}$ on coordinates in S. We reserve $\mathcal{U}$ to denote the uniform distribution over $\{0, 1\}$ and $\mathcal{U}(S)$ to denote the uniform distribution over S.

For a deterministic function f, we use $f(\mathcal{P})$ to denote the output distribution of $f(x)$ given a random $x \sim \mathcal{P}$. For every event $\mathcal{E}$, we define $\mathcal{P}(\mathcal{E})$ to be the probability that $\mathcal{E}$ happens under distribution $\mathcal{P}$. In addition, we use $\mathcal{P}(x)$ to denote the probability mass of x under $\mathcal{P}$, and use $\mathsf{supp}(\mathcal{P}) = \{x : \mathcal{P}(x) > 0\}$ to denote the support of $\mathcal{P}$. If $\mathcal{P}$ is a distribution over $\{0, 1\}^n$, we use $|\mathcal{P}|$ to denote the distribution over weights. That is, $|\mathcal{P}|(w) = \sum_{x:|x|=w} \mathcal{P}(x)$. We additionally define the symmetrized distribution $\mathcal{P}_{\mathrm{sym}}$ to be the distribution resulting from randomly permuting strings $x \sim \mathcal{P}$.

Let $\mathcal{Q}$ be a distribution. We use $\|\mathcal{P}-\mathcal{Q}\|_{\mathsf{TV}} = \frac{1}{2}\sum_x |\mathcal{P}(x)-\mathcal{Q}(x)|$ to denote their total variation distance.[3] We say $\mathcal{P}$ is ε-close to $\mathcal{Q}$ if $\|\mathcal{P}(x)-\mathcal{Q}(x)\|_{\mathsf{TV}} \le \varepsilon$, and ε-far otherwise.

Fact 3.1. *Total variation distance has the following equivalent characterizations:*

$$\|\mathcal{P}-\mathcal{Q}\|_{\mathsf{TV}} = \max_{\text{event } \mathcal{E}} \mathcal{P}(\mathcal{E}) - \mathcal{Q}(\mathcal{E}) = \min_{\substack{\text{random variable } (X,Y) \\ X \text{ has marginal } \mathcal{P} \text{ and } Y \text{ has marginal } \mathcal{Q}}} \mathbf{Pr}\left[X \ne Y\right].$$

Let $\mathcal{P}_1,\ldots,\mathcal{P}_t$ be distributions. Then $\mathcal{P}_1 \times \cdots \times \mathcal{P}_t$ is a distribution denoting the product of $\mathcal{P}_1,\ldots,\mathcal{P}_t$. We also use $\mathcal{P}^t$ to denote $\mathcal{P}_1 \times \cdots \times \mathcal{P}_t$ if each $\mathcal{P}_i$ is the same as $\mathcal{P}$. For a finite set $S \subseteq [t]$, we use $\mathcal{P}^S$ to denote the distribution $\mathcal{P}^t$ restricted to the coordinates of S. We say distribution $\mathcal{P}$ is a convex combination (or mixture) of $\mathcal{P}_1,\ldots,\mathcal{P}_t$ if there exist $\alpha_1,\ldots,\alpha_t \in [0,1]$ such that $\sum_{i\in[t]} \alpha_i = 1$ and $\mathcal{P}(x) = \sum_{i\in[t]} \alpha_i \cdot \mathcal{P}_i(x)$ for all x in the sample space. When it is clear from context, we will occasionally write mixtures more simply as $\mathcal{P} = \sum_{i\in[t]} \alpha_i \cdot \mathcal{P}_i$.

We will require two inequalities about total variation distance. The first allows us to argue that two product distributions must be far apart if their marginals do not match.

Lemma 3.2 ([KOW24, Lemma 4.2])**.** *Let $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{W}$ be distributions over an n-dimensional product space, and let $S \subseteq [n]$ be a non-empty set of size s. Assume*

- *$\mathcal{P}[S]$ and $\mathcal{W}[S]$ are two product distributions,*
- *$\|\mathcal{P}[\{i\}] - \mathcal{W}[\{i\}]\|_{\mathsf{TV}} \ge \varepsilon$ holds for all $i \in S$, and*
- *$\mathcal{W}(x) \ge \eta \cdot \mathcal{Q}(x)$ holds for some $\eta > 0$ and all x.*

Then

$$\|\mathcal{P}-\mathcal{Q}\|_{\mathsf{TV}} \ge 1 - 2 \cdot e^{-\varepsilon^2 s/2}/\eta.$$

The second states that the distance between a distribution $\mathcal{D}$ and a mixture of distributions must be large if the distance between $\mathcal{D}$ and each individual distribution in the mixture is also large.

Lemma 3.3 ([Vio20, Section 4.1], [KOW24, Lemma 4.3])**.** *Let $\mathcal{P}_1,\ldots,\mathcal{P}_t$ and $\mathcal{Q}$ be distributions. Assume there exists a value ε such that $\|\mathcal{P}_i - \mathcal{Q}\|_{\mathsf{TV}} \ge 1-\varepsilon$ for all $i \in [t]$. Then for any distribution $\mathcal{P}$ as a convex combination of $\mathcal{P}_1,\ldots,\mathcal{P}_t$, we have*

$$\|\mathcal{P}-\mathcal{Q}\|_{\mathsf{TV}} \ge 1 - (t+1)\cdot\varepsilon.$$

Occasionally, we will use a different distance measure between distributions $\mathcal{P}$ and $\mathcal{Q}$; namely, the Kolmogorov distance

$$\max_t \left| \mathop{\mathbf{Pr}}_{x\sim\mathcal{P}}\left[x > t\right] - \mathop{\mathbf{Pr}}_{y\sim\mathcal{Q}}\left[y > t\right] \right|.$$

Locality and Hypergraphs. Let $f\colon \{0,1\}^m \to \{0,1\}^n$. We say f is a d-local function if each output bit $i \in [n]$ depends on at most d output bits. If unspecified, we will always assume n, m, d are positive integers.

We sometimes take an alternative view, using hypergraphs to model the dependency relations in f. Let $G = (V, E)$ be an (undirected) hypergraph. For each $i \in V$, we use $I_G(i) \subseteq V$ to denote the set of vertices that share an edge with i. We say G has maximum degree d if $|I_G(i)| \le d$ holds for all $i \in V$. Define $N_G(i) = \{i' \in V : I_G(i) \cap I_G(i') \ne \emptyset\}$ to be the neighborhood of i. We visualize the function $f\colon \{0,1\}^m \to \{0,1\}^n$ as a hypergraph on the output bits $[n]$ with an edge for each input bit containing all of the output bits that depend on it.

[3] To evaluate total variation distance, we need two distributions to have the same sample space. This will be clear throughout the paper and thus we omit it for simplicity.

Concentration. We will need the following standard concentration bounds.

Fact 3.4 (Hoeffding's Inequality)**.** *Assume $X_1, \dots, X_n$ are independent random variables such that $a \le X_i \le b$ holds for all $i \in [n]$. Then for all $\delta \ge 0$, we have*

$$\max\left\{\mathbf{Pr}\left[\frac{1}{n}\sum_{i\in[n]}(X_i - \mathbb{E}[X_i]) \ge \delta\right], \mathbf{Pr}\left[\frac{1}{n}\sum_{i\in[n]}(X_i - \mathbb{E}[X_i]) \le -\delta\right]\right\} \le \exp\left\{-\frac{2n\delta^2}{(b-a)^2}\right\}.$$

Fact 3.5 (Chernoff's Inequality)**.** *Assume $X_1, \dots, X_n$ are independent random variables such that $X_i \in [0,1]$ holds for all $i \in [n]$. Let $\mu = \sum_{i\in[n]} \mathbb{E}[X_i]$. Then for all $\delta \in [0,1]$, we have*

$$\mathbf{Pr}\left[\sum_{i\in[n]} X_i \le (1-\delta)\mu\right] \le \exp\left\{-\frac{\delta^2\mu}{2}\right\}.$$

Norms and Hypercontractivity. For a function $g\colon \{0,1\}^n \to \{0,1\}$ and integer $q \ge 1$, define the norm $\|g\|_q = \left(\mathbb{E}_{x\sim\{0,1\}^n}[|g(x)|^q]\right)^{1/q}$. We will use the following *hypercontractive inequality* in several locations to control higher norms.

Lemma 3.6 ([Bon70])**.** *For a degree-d polynomial $p\colon \{0,1\}^n \to \mathbb{R}$ and an integer $q \ge 2$, we have*

$$\|p\|_q \le \sqrt{q-1}^{\,d} \cdot \|p\|_2.$$

Binomials and Entropy. Let $\mathcal{H}(x) = x \cdot \log\left(\frac{1}{x}\right) + (1-x)\cdot\log\left(\frac{1}{1-x}\right)$ be the binary entropy function. We will frequently use the following estimates regarding binomial coefficients and the entropy function.

Fact 3.7 (See e.g., [CT06, Lemma 17.5.1])**.** *For $1 \le k \le n-1$, we have*

$$\frac{2^{n\cdot\mathcal{H}(k/n)}}{\sqrt{8k(1-k/n)}} \le \binom{n}{k} \le \frac{2^{n\cdot\mathcal{H}(k/n)}}{\sqrt{\pi k(1-k/n)}}.$$

Fact 3.8 (See e.g., [Wik23a])**.** *For any $x \in [-1,1]$, we have*

$$1 - x^2 \le \mathcal{H}\left(\frac{1+x}{2}\right) = 1 - \frac{1}{2\ln(2)}\sum_{n=1}^{+\infty}\frac{x^{2n}}{n\cdot(2n-1)} \le 1 - \frac{x^2}{2\ln(2)}.$$

Fact 3.9 (See e.g., [Wik23b, Lug17])**.** *For $1 \le k \le n/2$, we have*

$$\sum_{i=0}^{k}\binom{n}{i} \le \min\left\{2^{n\cdot\mathcal{H}(k/n)}, \binom{n}{k}\cdot\frac{n-k+1}{n-2k+1}\right\}.$$

4 Classification of Locally Sampleable Distributions

In this section, we will prove a general classification result for uniform distributions with symmetric support that can be sampled by local functions. Let $\Psi \subseteq \{0,1,\dots,n\}$ be a non-empty set. We define $\mathcal{D}_\Psi$ to be the uniform distribution over $x \in \{0,1\}^n$ conditioned on $|x| \in \Psi$.

We will show that if the output distribution of a local function is close to $\mathcal{D}_\Psi$, then it is in fact close to one of the following six specific symmetric distributions: `zeros`, `ones`, `zerones`, `evens`, `odds`, and `all`, where

- $\texttt{zeros} = \mathcal{D}_{\{0\}}$, $\texttt{ones} = \mathcal{D}_{\{n\}}$, and $\texttt{zerones} = \mathcal{D}_{\{0,n\}}$,
- $\texttt{evens} = \mathcal{D}_{\{\text{even numbers in } \{0,1,\ldots,n\}\}}$ and $\texttt{odds} = \mathcal{D}_{\{\text{odd numbers in } \{0,1,\ldots,n\}\}}$,
- $\texttt{all} = \mathcal{D}_{\{0,1,\ldots,n\}}$.

Theorem 4.1. *Let $d \in \mathbb{N}$ and $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of d). Assume $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \le \varepsilon$ for some $\Psi \subseteq \{0,1,\ldots,n\}$. Then*

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \le O(\varepsilon)$$

for some $\mathcal{D} \in \{\texttt{zeros}, \texttt{ones}, \texttt{zerones}, \texttt{evens}, \texttt{odds}, \texttt{all}\}$.

Remark 4.2. We show the qualitative tightness of Theorem 4.1 from different angles.

- The six special distributions admit local sampling schemes: $\texttt{zeros}$ and $\texttt{ones}$ can be sampled by a 0-local function; $\texttt{all}$ and $\texttt{zerones}$ can be sampled by a 1-local function; $\texttt{evens}$ and $\texttt{odds}$ can be sampled by a 2-local function.
- The lower bound on n is necessarily depending on d. If $n \le d$, then one can sample the uniform distribution over any support $S \subseteq \{0,1\}^n$ of size $|S|$ dividing 2^d. This can be achieved by fixing a regular mapping $\pi\colon \{0,1\}^d \to S$ and using the d input bits to compute it. Also if n is a power of two and $d = \log(n)$, then one can directly sample a uniform string of Hamming weight one, which is uniform symmetric.
- The unspecified distance assumption ε cannot be replaced by a constant, i.e., local functions are indeed able to *arbitrarily* closely approximate uniform symmetric distributions.

  Starting from $\texttt{evens}$, we randomly flip the first $c \in [n]$ output bits with probability $1/4$. This distribution is 4-local since both $\texttt{evens}$ and the $1/4$-biased flipping are 2-local. It is easy to see that this distribution is at distance $2^{-\Theta(c)}$ to $\texttt{all}$ and $\texttt{evens}$, and is much farther from other uniform symmetric distributions. This shows that ε can be arbitrarily small even when d is a fixed constant.
- The distance blowup from $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}$ to $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}}$ is qualitatively necessary, i.e., the local distribution can be closer to a uniform symmetric distribution than to one of the six special ones. In particular, we identify the following example which rules out a bound of the form $(1 + o(1))\varepsilon$.

  Consider the distribution $\mathcal{D}$ that with probability $3/4$ is $\texttt{evens}$ and with probability $1/4$ is $\texttt{odds}$. Observe $\mathcal{D}$ can be sampled by a 3-local function via a similar strategy to that for $\texttt{evens}$. The uniform distribution over n-bit strings of Hamming weight $0, 1, 2$, or $4 \bmod 6$ is approximately $(1/6)$-close to $\mathcal{D}$; however, all six special distributions are at least $(1/4)$-far from $\mathcal{D}$. Thus, the implicit constant in Theorem 4.1 must be at least $3/2$.

To prove Theorem 4.1, we will classify Ψ into two cases and handle them separately. To this end, define $\iota(\Psi) \in \Psi$ to be the Hamming weight in Ψ that is closest to the middle:

$$\iota(\Psi) = \operatorname*{arg\,min}_{s \in \Psi} |s - n/2|$$

where we break ties arbitrarily. Intuitively, since $\iota(\Psi)$ is the dominating Hamming weight under the binomial distribution, $\mathcal{D}_\Psi$ is close to either $\mathcal{D}_{\{\iota(\Psi)\}}$ or $\frac{1}{2}\left(\mathcal{D}_{\{\iota(\Psi)\}} + \mathcal{D}_{\{n-\iota(\Psi)\}}\right)$. Based on this intuition, we divide Ψ into the following cases:

- Tail Regime: $\iota(\Psi) < n/2 - n^{2/3}$ or $\iota(\Psi) > n/2 + n^{2/3}$.

- Central Regime: $n/2 - n^{2/3} \le \iota(\Psi) \le n/2 + n^{2/3}$.

In the tail regime, we wish to show that $\mathcal{D}_\Psi$ is essentially `zeros`, `ones`, or `zerones`. The following result will be proved in Subsection 4.1.

Theorem 4.3. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of d). If $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \le 1/3$ for some Ψ in the tail regime, then $\mathcal{D}_\Psi \in \{\texttt{zeros}, \texttt{ones}, \texttt{zerones}\}$.*

In the central regime, we aim to show that $\mathcal{D}_\Psi$ is essentially `evens`, `odds`, or `all` by Theorem 4.4, which will be proved in Subsection 4.2.

Theorem 4.4. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of d). If $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \le \varepsilon$ for some Ψ in the central regime, then $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \le O(\varepsilon)$ for some $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}, \texttt{all}\}$.*

We can now easily establish Theorem 4.1.

Proof of Theorem 4.1. We assume $d \ge 1$, as otherwise $f(\mathcal{U}^m)$ is $\Omega(1)$-far from every uniform symmetric distribution other than `zeros` or `ones`, in which case

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \le 1 \le O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}).$$

Similarly, we may assume $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \le 1/3$. Applying Theorem 4.3 or Theorem 4.4 depending on which regime Ψ is in yields the result. □

4.1 Tail Regime

In this section, we deal with the regime where strings from $\mathcal{D}_\Psi$ are spread out in the tail layers, but no weight is extremely close to the center, i.e., $\iota(\Psi) < n/2 - n^{2/3}$ or $\iota(\Psi) > n/2 + n^{2/3}$ is the Hamming weight in Ψ closest to $n/2$.

Theorem (Theorem 4.3 Restated)**.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of d). If $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \le 1/3$ for some Ψ in the tail regime, then $\mathcal{D}_\Psi \in \{\texttt{zeros}, \texttt{ones}, \texttt{zerones}\}$.*

The main idea is that in this regime, the output must be concentrated on few weights. Thus, we can apply known error bounds on the individual weights, and use Lemma 3.3 to combine the errors. We formalize this idea below.

Claim 4.5. Let $\Psi \subseteq \{0,1,\dots,n\}$ be a non-empty set satisfying $|\iota(\Psi) - n/2| > n^{2/3}$. If n is sufficiently large, then there exists a non-empty set $\overline{\Psi} \subseteq \Psi$ of size $|\overline{\Psi}| \le O(n^{1/3})$ such that

$$\left\|\mathcal{D}_\Psi - \mathcal{D}_{\overline{\Psi}}\right\|_{\mathsf{TV}} \le 1/2.$$

Proof. Let $\Psi = \{s_1, \dots, s_k\}$, where the weights are ordered nonincreasing in their distance from $n/2$. Additionally, let $\ell = Cn^{1/3}$ for a sufficiently large constant C. If $k < \ell$, simply set $\overline{\Psi} = \Psi$. Otherwise we will show it suffices to set $\overline{\Psi} = \{s_1, \dots, s_\ell\}$.

Consider an arbitrary index $i \in [k]$, and assume without loss of generality that $s_i \le n/2$. Note that $|s_i - n/2| > n^{2/3}$ by assumption, so we in fact have $s_i < \frac{n}{2} - n^{2/3}$. Then,

$$\Pr_{x \sim \mathcal{D}_\Psi}\left[|x| = s_i \;\middle|\; |x| \notin \{s_1, s_2, \dots, s_{i-1}\}\right] \ge \frac{\binom{n}{s_i}}{\sum_{\substack{j \le s_i \\ j \ge n - s_i}} \binom{n}{j}} = \frac{\binom{n}{s_i}}{2\sum_{j \le s_i} \binom{n}{j}}$$

$$
\begin{aligned}
&\geq \frac{\binom{n}{s_i}}{2\binom{n}{s_i}\frac{n-s_i+1}{n-2s_i+1}} && \text{(by Fact 3.9)}\\
&\geq \frac{n-2\left(\frac{n}{2}-n^{2/3}\right)}{2n} = \Theta(n^{-1/3}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\Pr_{x\sim\mathcal{D}_\Psi}\left[|x|\notin\{s_1,\ldots,s_\ell\}\right] &= \prod_{i=1}^{\ell}\Pr_{x\sim\mathcal{D}_\Psi}\left[|x|\neq s_i \;\middle|\; |x|\notin\{s_1,s_2,\ldots,s_{i-1}\}\right]\\
&\leq \left(1-\Theta(n^{-1/3})\right)^{\ell} \leq \exp\left\{-\ell\cdot\Theta(n^{-1/3})\right\} \leq 1/2.
\end{aligned}
$$

In other words, setting $\overline{\Psi}=\{s_1,\ldots,s_\ell\}$ yields $\left\|\mathcal{D}_\Psi-\mathcal{D}_{\overline{\Psi}}\right\|_{\mathsf{TV}}\leq 1/2$. □

Theorem 4.6 (Combination of [FLRS23, Theorem 1.2] and [KOW24, Theorem 5.10])**.** *Let $1\leq k\leq n-1$ be an integer, and let $f\colon\{0,1\}^m\to\{0,1\}^n$ be a d-local function. Then*

$$
\left\|f(\mathcal{U}^m)-\mathcal{D}_{\{k\}}\right\|_{\mathsf{TV}}\geq 1-O_d(n^{-1/2}).
$$

We now proceed to the proof of the section's main result.

Proof of Theorem 4.3. We will prove the contrapositive. Suppose $\Psi\not\subseteq\{0,n\}$. If $|\Psi|\leq n^{1/3}$, let $\overline{\Psi}=\Psi$; otherwise, let $\overline{\Psi}\subseteq\Psi$ as guaranteed by Claim 4.5. We further prune $\overline{\Psi}$ to ensure no individual weight is sampleable by defining $\Psi^\dagger=\overline{\Psi}\setminus\{0,n\}$. Observe that our previous assumptions imply $\Psi^\dagger$ is non-empty. Thus, this removal changes the distribution minimally, as the support of $\mathcal{D}_{\overline{\Psi}}$ has size at least n, and we removed at most two elements. In particular,

$$
\left\|\mathcal{D}_{\Psi^\dagger}-\mathcal{D}_{\overline{\Psi}}\right\|_{\mathsf{TV}}\leq\frac{2}{n}.
$$

For each $s\in\Psi^\dagger$, we apply Theorem 4.6 to obtain $\left\|f(\mathcal{U}^m)-\mathcal{D}_{\{s\}}\right\|_{\mathsf{TV}}\geq 1-O_d(n^{-1/2})$. Since $\mathcal{D}_{\Psi^\dagger}$ is the convex combination of $\mathcal{D}_{\{s\}}$ for $s\in\Psi^\dagger$, Lemma 3.3 implies

$$
\|f(\mathcal{U}^m)-\mathcal{D}_{\Psi^\dagger}\|_{\mathsf{TV}}\geq 1-O_d\left(\frac{|\Psi^\dagger|}{\sqrt{n}}\right)\geq 1-O_d\left(n^{-1/6}\right).
$$

We conclude by applying the triangle inequality to deduce

$$
\begin{aligned}
\|f(\mathcal{U}^m)-\mathcal{D}_\Psi\|_{\mathsf{TV}} &\geq \|f(\mathcal{U}^m)-\mathcal{D}_{\Psi^\dagger}\|_{\mathsf{TV}}-\left\|\mathcal{D}_{\Psi^\dagger}-\mathcal{D}_{\overline{\Psi}}\right\|_{\mathsf{TV}}-\left\|\mathcal{D}_{\overline{\Psi}}-\mathcal{D}_\Psi\right\|_{\mathsf{TV}}\\
&\geq 1-O_d\left(n^{-1/6}\right)-\frac{2}{n}-\frac{1}{2}>\frac{1}{3}.
\end{aligned}
$$

□

4.2 Central Regime

In this section, we handle the regime where some Hamming weight is very close to the center, i.e., $n/2-n^{2/3}\leq\iota(\Psi)\leq n/2+n^{2/3}$.

Theorem (Theorem 4.4 Restated)**.** *Let $f\colon\{0,1\}^m\to\{0,1\}^n$ be a d-local function with n sufficiently large (in terms of d). If $\|f(\mathcal{U}^m)-\mathcal{D}_\Psi\|_{\mathsf{TV}}\leq\varepsilon$ for some Ψ in the central regime, then $\|f(\mathcal{U}^m)-\mathcal{D}\|_{\mathsf{TV}}\leq O(\varepsilon)$ for some $\mathcal{D}\in\{\texttt{evens},\texttt{odds},\texttt{all}\}$.*

A crucial structural result in our analysis of this regime is the following local limit theorem, which says that any locally sampleable distribution close to a roughly centered uniform symmetric distribution must also be near to a mixture of `evens` and `odds`. The proof is somewhat involved, so we defer it to its own section (Section 5) for clarity.

Theorem 4.7. *Let $\delta > 0$ and $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of δ and d). Let $\Psi \subseteq \{0,1,2,\dots,n\}$ be a set containing some element $n(1/2 \pm c(d,\delta))$ for some $c(d,\delta) > 0$ a small enough function of d and δ. Then there exists a distribution $\mathcal{M}$ which is a mixture of* `evens` *and* `odds` *so that*

$$\|f(\mathcal{U}^m) - \mathcal{M}\|_{\mathsf{TV}} \le O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) + \delta.$$

The following two results used in proving Theorem 4.7 will also be useful within this section. (Their proofs can likewise be found in Section 5.) The first allows us to reason about the distance between two distributions A and B over $\{0,1\}^n$ with B symmetric by reasoning about their weight distributions and A's symmetry. It is essentially a consequence of the observation that A is far from B if A is either far from its own symmetrization or if its weight distribution noticeably differs from B's. Recall $|A|$ denotes the distribution over the Hamming weight of strings $x \sim A$, and A_{sym} denotes the distribution resulting from randomly permuting strings $x \sim A$.

Lemma 4.8. *Let A and B be two distributions on $\{0,1\}^n$ with B symmetric. Then*

$$\|A - B\|_{\mathsf{TV}} = \Theta(\||A| - |B|\|_{\mathsf{TV}} + \|A - A_{\mathrm{sym}}\|_{\mathsf{TV}}).$$

The second result can be viewed as a continuity property of the weight distribution of $f(\mathcal{U}^m)$ in this section's regime.

Proposition 4.9. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of d). Let $\Psi \subseteq \{0,1,\dots,n\}$ be a set containing an element $n(1/2 \pm c(d))$ for some $c(d) > 0$ a small enough function of d. Then the distribution $f(\mathcal{U}^m)$ can be written as a mixture $aE + (1-a)X$ with $a = O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}})$ so that for any even Δ and $x \in \{0,1,\dots,n\}$,*

$$\big|\, \mathbf{Pr}\left[|X| = x\right] - \mathbf{Pr}\left[|X| = x + \Delta\right] \big| = O_d\left(\frac{|\Delta|}{n}\right).$$

Here, one should view E as the "error" part of $f(\mathcal{U}^m)$ which is far from every roughly centered uniform symmetric distribution, while the remaining part X has a smooth weight distribution (modulo parity constraints). These properties are ultimately inherited via the structure of the independent neighborhoods obtained after conditioning, as sketched in Section 2. E corresponds to neighborhoods that are far from uniform, while the smoothness of X is a consequence of the anticoncentration properties of many roughly unbiased independent neighborhoods (formalized in Appendix A).

We proceed with an important corollary of Theorem 4.7:

Corollary 4.10. *Let $\delta > 0$ and $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of δ and d). Let $\Psi \subseteq \{0,1,2,\dots,n\}$ contain some element within $n^{2/3}$ of $n/2$. Then there exists a distribution $\mathcal{D} \in \{$*`evens`*, *`odds`*, *`all`*$\}$ so that*

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \le O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) + \delta.$$

Proof. For sufficiently large n, we may apply Theorem 4.7 to deduce that for some mixture $\mathcal{M} = \eta \cdot \texttt{evens} + (1-\eta) \cdot \texttt{odds}$, we have

$$\|f(\mathcal{U}^m) - \mathcal{M}\|_{\mathsf{TV}} \leq O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) + \frac{\delta}{3}. \tag{1}$$

We will show a similar upper bound for $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}}$. This essentially follows from a number of applications of the triangle inequality. We have

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq \|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} + \|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}}\,, \tag{2}$$

so it remains to bound $\|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}}$. We first show $\|\mathcal{M} - \mathcal{D}\|_{\mathsf{TV}} \leq 2\,\|\mathcal{M} - \mathcal{D}_\Psi\|_{\mathsf{TV}}$. Let S be a partition of $\{0,1\}^n$ into pairs of elements (x_e, x_o) such that each pair contains one element x_e of even weight and one element x_o of odd weight. Then,

$$\begin{aligned}
\|\mathcal{M} - \mathcal{D}\|_{\mathsf{TV}} &= \frac{1}{2} \sum_{x \in \{0,1\}^n} |\mathcal{M}(x) - \mathcal{D}(x)| \\
&= \frac{1}{2} \sum_{(x_e,x_o) \in S} |\mathcal{M}(x_e) - \mathcal{D}(x_e)| + |\mathcal{M}(x_o) - \mathcal{D}(x_o)| \\
&= 2^{n-2} \left(\left| \frac{\eta}{2^{n-1}} - \mathcal{D}(x_e) \right| + \left| \frac{1-\eta}{2^{n-1}} - \mathcal{D}(x_o) \right| \right).
\end{aligned}$$

Breaking into cases, we find that

$$\|\mathcal{M} - \mathcal{D}\|_{\mathsf{TV}} = \begin{cases} 1-\eta & \text{if } \mathcal{D} = \texttt{evens}, \\ \eta & \text{if } \mathcal{D} = \texttt{odds}, \\ \left|\eta - \frac{1}{2}\right| & \text{if } \mathcal{D} = \texttt{all}. \end{cases}$$

By choosing $\mathcal{D}$ appropriately, we may assume $\|\mathcal{M} - \mathcal{D}\|_{\mathsf{TV}} = \zeta := \min\{\eta, |\eta - 1/2|, 1-\eta\}$. Similarly,

$$\begin{aligned}
\|\mathcal{M} - \mathcal{D}_\Psi\|_{\mathsf{TV}} &= \frac{1}{2} \sum_{x \in \{0,1\}^n} |\mathcal{M}(x) - \mathcal{D}_\Psi(x)| \\
&= \frac{1}{2} \sum_{(x_e,x_o) \in S} \left| \frac{\eta}{2^{n-1}} - \mathcal{D}_\Psi(x_e) \right| + \left| \frac{1-\eta}{2^{n-1}} - \mathcal{D}_\Psi(x_o) \right|.
\end{aligned}$$

Since $\mathcal{D}_\Psi$ is uniform, each pair $(x_e, x_o) \in S$ either satisfies $\mathcal{D}_\Psi(x_e) = \mathcal{D}_\Psi(x_o)$ or one of $\mathcal{D}_\Psi(x_e), \mathcal{D}_\Psi(x_o)$ is zero. Again breaking into cases,

$$2^{n-1} \left(\left| \frac{\eta}{2^{n-1}} - \mathcal{D}_\Psi(x_e) \right| + \left| \frac{1-\eta}{2^{n-1}} - \mathcal{D}_\Psi(x_o) \right| \right) \geq \begin{cases} \eta & \text{if } \mathcal{D}_\Psi(x_e) = 0, \\ 1-\eta & \text{if } \mathcal{D}_\Psi(x_o) = 0, \\ |2\eta - 1| & \text{if } \mathcal{D}_\Psi(x_e) = \mathcal{D}_\Psi(x_o). \end{cases}$$

In other words,

$$\left| \frac{\eta}{2^{n-1}} - \mathcal{D}_\Psi(x_e) \right| + \left| \frac{1-\eta}{2^{n-1}} - \mathcal{D}_\Psi(x_o) \right| \geq \frac{\zeta}{2^{n-1}},$$

so

$$\|\mathcal{M} - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq \frac{1}{2} \cdot 2^{n-1} \cdot \frac{\zeta}{2^{n-1}} = \frac{1}{2}\,\|\mathcal{M} - \mathcal{D}\|_{\mathsf{TV}}\,. \tag{3}$$

Therefore,

$$\begin{aligned}\|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}} &\le \|\mathcal{D}_\Psi - \mathcal{M}\|_{\mathsf{TV}} + \|\mathcal{M} - \mathcal{D}\|_{\mathsf{TV}} \\ &\le 3\,\|\mathcal{M} - \mathcal{D}_\Psi\|_{\mathsf{TV}} && \text{(by (3))} \\ &\le 3\,(\|\mathcal{M} - f(\mathcal{U}^m)\|_{\mathsf{TV}} + \|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}})\,.\end{aligned}$$

Combining with (1) and (2) yields the result. □

Ideally, we would want to prove Theorem 4.4 by simply applying Corollary 4.10 with $\delta = O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}})$. However, if δ is too small, n may not be large enough (in terms of δ and d) to satisfy the assumption of Corollary 4.10. Thus, we can only deduce that Theorem 4.4 holds unless $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}}$ is less than a small constant. Hence, we will turn our attention to this special case. The two main consequences of such an assumption can be encapsulated in the following claims.

Claim 4.11. Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function. If $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} < 2^{-d}$ for some $\mathcal{D} \in \{\mathtt{evens}, \mathtt{odds}\}$, then $\mathsf{supp}\,(f(\mathcal{U}^m)) \subseteq \mathsf{supp}\,(\mathcal{D})$.

Proof. Observe that we may write $|f(x)| \bmod 2$ as an $\mathbb{F}_2$-polynomial p (of the input bits) of degree at most d. If p is not constant, then it must take each possible value with probability at least 2^{-d}, contradicting our distance assumption. □

We call a distribution over $\{0,1\}^n$ *k-wise independent* if the projection onto any $k' \le k$ indices is uniformly distributed over $\{0,1\}^{k'}$.

Claim 4.12. Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function. If $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} < 2^{-kd}$ for some $\mathcal{D} \in \{\mathtt{evens}, \mathtt{odds}, \mathtt{all}\}$ and positive integer $k < n$, then $f(\mathcal{U}^m)$ is a k-wise independent distribution.

Proof. For any set $T \subseteq [n]$ of coordinates of size k, the data processing inequality implies

$$2^{-kd} > \|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \ge \|f(\mathcal{U}^m)[T] - \mathcal{D}[T]\|_{\mathsf{TV}}\,. \tag{4}$$

Note that $\mathcal{D}[T]$ is the uniform distribution on $\{0,1\}^T$. On the other hand, the outputs of f on T only depend on at most kd input bits, so the probability that $f(\mathcal{U}^m)[T]$ assigns to any given string must be a multiple of 2^{-kd}. Thus, $\|f(\mathcal{U}^m)[T] - \mathcal{D}[T]\|_{\mathsf{TV}}$ is also a multiple of 2^{-kd}. Combining this observation with (4) yields $f(\mathcal{U}^m)[T] = \mathcal{D}[T]$, concluding the proof. □

Using these claims, we will show that the distribution on weights of our sampled distribution $|f(\mathcal{U}^m)|$ is "missing" a noticeable amount of mass around $n/2$. For this, we record the observation that we can write $|f(\mathcal{U}^m)|(x) = |\mathcal{D}|(x) + \kappa A(x) - \kappa S(x)$ for some disjoint distributions A, S and $\kappa \coloneqq \||f(\mathcal{U}^m)| - |\mathcal{D}|\|_{\mathsf{TV}} \le \|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}}$.

Proposition 4.13. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of d). Suppose $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} = \lambda$ for some $\mathcal{D} \in \{\mathtt{evens}, \mathtt{odds}, \mathtt{all}\}$ and λ less than a sufficiently small constant (in terms of d). Moreover, let A, S be distributions with disjoint support satisfying $|f(\mathcal{U}^m)| = |\mathcal{D}| + \kappa A - \kappa S$ for some $\kappa \le \lambda$. If $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \le O(\lambda)$ for some sufficiently small implicit constant and $\Psi \subseteq \{0,1,\ldots,n\}$, then*

1. *$\kappa = \Omega(\lambda)$, and*
2. *S assigns at least a constant fraction of its mass to weights within $O(\sqrt{n})$ of $n/2$.*

Note that if the assumption $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \le O(\lambda)$ does not hold, there is nothing left to prove (i.e., the conclusion of Theorem 4.4 is already satisfied).

Remark 4.14. In the following proof, we will want to apply hypercontractivity (Lemma 3.6) for expectations over any $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}, \texttt{all}\}$ (as opposed to over $\mathcal{U}^n$). Observe, however, that each of these distributions is $(n-1)$-wise independent. Thus, $\mathbb{E}[p(\mathcal{U}^n)] = \mathbb{E}[p(\mathcal{D})]$ for any polynomial p of degree at most $n-1$. In particular, Lemma 3.6 still holds when q is even and $q \cdot \deg(p) < n$.

Similarly, we will need to apply Fact 3.4 for $|\mathcal{D}|$. In the case of $\mathcal{D} = \texttt{evens}$, for example, we observe that

$$\Pr_{x\sim|\mathcal{D}|}[x = z] = \Pr_{x\sim|\mathcal{U}^n|}\left[x = z \,\middle|\, |x| \text{ is even}\right] \leq \frac{\mathbf{Pr}_{x\sim|\mathcal{U}^n|}[x = z]}{\mathbf{Pr}_{x\sim|\mathcal{U}^n|}[|x| \text{ is even}]} = 2 \Pr_{x\sim|\mathcal{U}^n|}[x = z].$$

Thus, we may use Fact 3.4 up to some small constant loss.

Proof of Proposition 4.13. For clarity, let $W = |f(\mathcal{U}^m)|$. To prove the first part of the claim, we note that by Lemma 4.8,

$$\begin{aligned}\lambda = \|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} &= \Theta\left(\|f(\mathcal{U}^m) - f(\mathcal{U}^m)_{\mathrm{sym}}\|_{\mathsf{TV}} + \|W - |\mathcal{D}|\|_{\mathsf{TV}}\right) \\ &= \Theta(\|f(\mathcal{U}^m) - f(\mathcal{U}^m)_{\mathrm{sym}}\|_{\mathsf{TV}} + \kappa).\end{aligned}$$

This means that either $\kappa = \Omega(\lambda)$ or $\|f(\mathcal{U}^m) - f(\mathcal{U}^m)_{\mathrm{sym}}\|_{\mathsf{TV}} = \Omega(\lambda)$. However, if the latter is true, then Lemma 4.8 shows that $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq \Omega(\lambda)$. This contradicts our assumption that $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq O(\lambda)$ for a sufficiently small implicit constant. Henceforth, we will assume

$$\kappa = \Theta(\lambda). \tag{5}$$

We now turn to showing S assigns a constant fraction of its mass to weights close to $n/2$. First note that $\mathsf{supp}\,(W) \subseteq \mathsf{supp}\,(|\mathcal{D}|)$ by Claim 4.11. Therefore, we may assume that $\Psi \subseteq \mathsf{supp}\,(|\mathcal{D}|)$ as well, since replacing Ψ with $\Psi \cap \mathsf{supp}\,(|\mathcal{D}|)$ will not increase the distance between $\mathcal{D}_\Psi$ and $f(\mathcal{U}^m)$.

In particular, one should view $|\mathcal{D}_\Psi|$ as the result of redistributing mass from weights in $\mathsf{supp}\,(|\mathcal{D}|)\setminus \Psi$ evenly over the elements in Ψ. Since W is close to $|\mathcal{D}_\Psi|$, we can roughly view S as corresponding to the missing weights, and A as corresponding to the small "boost" from the redistributed mass. A priori, it is difficult to reason about S's distribution, as we have limited information about where the missing weights are. However,

$$\gamma \coloneqq \|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}} \leq \|\mathcal{D}_\Psi - f(\mathcal{U}^m)\|_{\mathsf{TV}} + \|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq O(\lambda) \tag{6}$$

is less than a small constant, so many weights around $n/2$ must be included in Ψ, which will receive further weight from A. Therefore, we will proceed by arguing A assigns substantial mass to weights in a small interval around $n/2$, and then convert that conclusion to one about S.

We begin by observing that all but a small constant fraction of the support of $|\mathcal{D}|$ in $I = [n/2 - \sqrt{n}/10, n/2 + \sqrt{n}/10]$ are in Ψ. Indeed by Fact 3.7 and Fact 3.8, each element of $\mathsf{supp}\,(|\mathcal{D}|) \cap (I \setminus \Psi)$ contributes at least

$$2^{-n}\binom{n}{\frac{n}{2} + \frac{\sqrt{n}}{10}} \geq \frac{2^{n\left(\mathcal{H}\left(\frac{1}{2} - \frac{1}{10\sqrt{n}}\right) - 1\right)}}{\sqrt{2n\left(1 - \frac{1}{25n}\right)}} \geq \frac{1}{2\sqrt{n}}$$

to the distance $\|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}}$, so by (6) such elements can only consist of an $O(\lambda)$-fraction of I.

Next, note that

$$1 = \sum_{s\in\Psi} |\mathcal{D}_\Psi|(s) = \gamma + \sum_{s\in\Psi} |\mathcal{D}|(s),$$

so $|\mathcal{D}_\Psi|(s) = |\mathcal{D}|(s)/(1-\gamma)$ for all $s \in \Psi$. This implies

$$\sum_{x\in I} \max\left\{|\mathcal{D}_\Psi|(x) - |\mathcal{D}|(x), 0\right\} = \sum_{x\in I\cap\Psi} \frac{|\mathcal{D}|(x)}{1-\gamma} - |\mathcal{D}|(x)$$

$$\begin{aligned}
&\geq \gamma \sum_{x \in I \cap \Psi} |\mathcal{D}|(x) \\
&\geq \|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}} \cdot \Omega(1-\lambda) \\
&\geq \Omega(\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} - \|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) \quad \text{(by triangle inequality)} \\
&\geq \Omega(\lambda). \qquad (7)
\end{aligned}$$

Now recall we originally assumed for some sufficiently small implicit constant that

$$\begin{aligned}
O(\lambda) &= \|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \\
&\geq \|W - |\mathcal{D}_\Psi|\|_{\mathsf{TV}} \qquad \text{(by data processing inequality)} \\
&\geq \sum_{x \in I} \max\{|\mathcal{D}_\Psi|(x) - W(x), 0\} \\
&\geq \sum_{x \in I} \max\{|\mathcal{D}_\Psi|(x) - |\mathcal{D}|(x) - \kappa A(x), 0\} \\
&\geq \sum_{x \in I} \max\{|\mathcal{D}_\Psi|(x) - |\mathcal{D}|(x), 0\} - \kappa A(x) \\
&\geq \Omega(\lambda) - \sum_{x \in I} \kappa A(x). \qquad \text{(by (7))}
\end{aligned}$$

Therefore, $\sum_{x \in I} \kappa A(x) = \Omega(\lambda)$. Moreover, A must assign constant mass to I by (5).

Our strategy will be to leverage this fact to show S must also assign a constant fraction of its mass to an interval of width $O(\sqrt{n})$ around $n/2$. Let r be the largest odd integer so that $2r < \log(1/\lambda)/d$. Combining with (5), we record the following equation for later convenience:

$$\frac{1}{\kappa} = \Theta\left(\frac{1}{\lambda}\right) = \exp\left(O(dr)\right). \tag{8}$$

By Claim 4.12, we have that W matches its first $2r$ moments with the binomial distribution $|\mathcal{U}^n|$, or equivalently, with $|\mathcal{D}|$. Thus, any polynomial p of degree at most $2r$ satisfies $\mathbb{E}[p(W)] = \mathbb{E}[p(|\mathcal{D}|)]$. In particular by the definition of W and the fact that for any distribution $\mathcal{P}$ and function q, $\mathbb{E}_{x \sim \mathcal{P}}[q(x)]$ is a linear functional on $\mathcal{P}$, we have

$$\mathbb{E}[p(A)] = \mathbb{E}[p(S)]. \tag{9}$$

Suppose momentarily we were able to set p to be the indicator for the interval I. Then the proposition would follow, since our earlier deduction about A's distribution gives $\mathbb{E}[p(A)] = \mathbf{Pr}[A \in I] = \Omega(1)$, so (9) implies $\mathbf{Pr}[S \in I] = \Omega(1)$. While this scenario is of course unrealistic, as the indicator for I will have an unaffordably high degree, the overall idea motivates our strategy of working with a carefully chosen polynomial p.

We morally want to choose p to approximate the indicator for I. However, we must be a bit discerning in our choice; polynomials blow-up in their tails, so any choice of p will inevitably be a poor approximation over the entire domain. We can distill the desired behavior for our low-degree polynomial p into the following three constraints:

1. p puts substantial mass on weights around $n/2$,
2. p puts minimal mass on weights further from the center, and
3. p's inevitable blow-up occurs as far into the tails as possible.

Ultimately, we set $p(x) = (T_r(y/r)/y)^2$ where $y := (x - n/2)/\sqrt{n}$ and T_r is the Chebyshev polynomial (of the first kind) $T_r(\cos(\theta)) = \cos(r\theta)$. Intuitively, Chebyshev polynomials remain relatively flat on the interval $[-1, 1]$, which allows us to (partially) achieve Item 2. The division by r "stretches" the flat region to further guarantee Item 2 and to address Item 3 by pushing the blow-up further into the tails. Finally, the additional $1/y$ factor and the choice of the largest possible (affordable) degree satisfies Item 1. We formalize these properties below.

Fact 4.15. *We have that:*

1. *$p(x)$ is a polynomial of degree at most $2r$.*
2. *$p(x) \geq 1/2$ when $|y| \leq 1/10$.*
3. *$p(x) \leq \min(1, 1/y^2)$ when $|y| \leq r$.*
4. *$p(x) \leq O(y/r)^{2r}$ when $|y| \geq r$.*
5. *$p(x) \geq 0$ for all x.*

Proof.

1. Recall T_r is a polynomial of degree r. Since r is odd, T_r is odd, and thus has a root at $y = 0$. In particular, $T_r(y/r)/y$ is a polynomial of degree at most $r - 1$.
2. Let $\sin(\theta) = y/r$. Then $p(x) = |\sin(r\theta)/y|^2$. Note that $|\theta| \leq 1/(5r)$ when $|y| < 1/10$, so both $\sin(r\theta)/(r\theta)$ and $\sin(\theta)/\theta$ are in $[0.99, 1]$. Thus,
$$p(x) = |\sin(r\theta)/(r\sin(\theta))|^2 = |\sin(r\theta)/(r\theta)|^2 |\sin(\theta)/\theta|^{-2} \geq 1/2.$$
3. Note that $|T_r(x)| \leq 1$ for $|x| \leq 1$ and so $p(x) \leq 1/y^2$. Finally, as above
$$p(x) = |\sin(r\theta)/(r\sin(\theta))|^2 \leq 1.$$
4. If $a = y/r$ we note that $a = (z + z^{-1})/2$ for some real z with $1 \leq |z| \leq 2|a|$. Setting $z = e^{i\theta}$, we can write $\cos(\theta) = (z + z^{-1})/2$, so $T_r(a) = T_r((z + z^{-1})/2) = (z^r + z^{-r})/2 \leq (2|a|)^r$.
5. Note that p is a square. □

Recall A must assign constant mass to $I = [n/2 - \sqrt{n}/10, n/2 + \sqrt{n}/10]$. Moreover, Item 2 and Item 5 imply $\mathbb{E}[p(A)] = \Omega(1)$. In particular, for some large enough constant C, we have that $\mathbb{E}[p(A)] > 1/C$. Therefore $\mathbb{E}[p(S)]$ is also at least $1/C$ by (9).

We consider the contribution to $\mathbb{E}[p(S)]$ coming from three ranges: $|y| \leq C$, $C \leq |y| \leq r$, and $|y| \geq r$. (If $C \geq r$, we ignore the second and third ranges.) By Item 3, the contributions from the first and second ranges are at most $\mathbf{Pr}[|S - n/2| \leq C\sqrt{n}]$ and $1/C^2$, respectively. For the third range, recall that $0 \leq W(x) = |\mathcal{D}|(x) - \kappa S$ for any $x \in \mathsf{supp}\,(S)$. Thus, κS must assign less mass to each point than $|\mathcal{D}|$ does. Therefore this contribution to $\mathbb{E}[p(S)]$ is at most

$$\begin{aligned}
&(1/\kappa) \mathop{\mathbb{E}}_{x\sim|\mathcal{D}|}[p(x)\mathbb{1}\{|x - n/2| > r\sqrt{n}\}] \\
&\leq (1/\kappa) \mathop{\mathbb{E}}_{x\sim|\mathcal{D}|}[p^2(x)\mathbb{1}\{|x - n/2| > r\sqrt{n}\}]^{1/2} \mathop{\mathbf{Pr}}_{x\sim|\mathcal{D}|}\left[|x - n/2| > r\sqrt{n}\right]^{1/2} && \text{(by Cauchy-Schwarz)} \\
&\leq (1/\kappa) \mathop{\mathbb{E}}_{x\sim|\mathcal{D}|}\left[O\left(\frac{|x - n/2|/\sqrt{n}}{r}\right)^{4r}\right]^{1/2} \exp(-\Omega(r^2)) && \text{(by Item 4 \& Fact 3.4)} \\
&\leq (1/\kappa) \cdot r^{-2r} \cdot O(r)^r \cdot \exp(-\Omega(r^2)) && \text{(by Lemma 3.6)}
\end{aligned}$$

$$= \exp(O(dr) - \Omega(r^2)) = \exp(-\Omega(r^2)), \qquad \text{(by (8))}$$

with the last line holding for r large enough relative to d (which is achievable by enforcing λ sufficiently small). Thus,

$$\frac{1}{C} < \mathbb{E}[p(S)] \leq \mathbf{Pr}[|S - n/2| \leq C\sqrt{n}] + \frac{1}{C^2} + \exp(-\Omega(r^2)),$$

so S must assign at least constant mass to the interval $[n/2 - C\sqrt{n}, n/2 + C\sqrt{n}]$. □

We can now prove the main result of this subsection: Theorem 4.4. Recall we aim to show that $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq O(\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}})$ for Ψ containing an element close to $n/2$. Our strategy will be to restrict to the small error case via our local limit theorem, and then use the preceding results to enact a pairing argument (as sketched in Section 2).

Proof. By Corollary 4.10, there exists a distribution $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}, \texttt{all}\}$ so that

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) + \delta$$

for some sufficiently small $\delta > 0$. The result immediately follows if $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} > \delta$, so we may assume $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq \delta$. In particular, we have $\lambda := \|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}}$ is less than some sufficiently small constant. From here, we assume by contradiction that

$$\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq O(\lambda) \tag{10}$$

for a sufficiently small implicit constant.

As above, let $W := |f(\mathcal{U}^m)| = |\mathcal{D}| + \kappa A - \kappa S$ for $\kappa := \|W - |\mathcal{D}|\|_{\mathsf{TV}} \leq \lambda$ and A, S some distributions with disjoint support. Once again by Claim 4.11, we have that W is supported only on the elements in the support of $|\mathcal{D}|$, and we assume without loss of generality that $\Psi \subseteq \mathsf{supp}\,(|\mathcal{D}|)$. This implies

1. $A(x) = 0$ for $x \in \mathsf{supp}\,(|\mathcal{D}|)$,
2. $S(x) = 0$ for $x \notin \mathsf{supp}\,(|\mathcal{D}|)$, and
3. $|\mathcal{D}_\Psi|(x) \geq |\mathcal{D}|(x)$ for $x \in \Psi$.

Next, we apply Proposition 4.13 to deduce

$$\lambda \geq \kappa \geq \Omega(\lambda) \tag{11}$$

and S assigns at least a constant fraction of its mass to weights within an interval I of width $O(\sqrt{n})$ around $n/2$. In particular, Item 2 gives

$$\sum_{x \in I \cap \mathsf{supp}(|\mathcal{D}|)} S(x) \geq \Omega(1). \tag{12}$$

At a high level, we aim to show there exist many nearby pairs $x, y \in I \cap \mathsf{supp}\,(|\mathcal{D}|)$ with $x \notin \Psi$ and $y \in \Psi$. Summing over all pairs will produce the desired contradiction $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} = \Omega(\lambda)$.

We first show there are many $x \in I \cap \mathsf{supp}\,(|\mathcal{D}|)$ with $x \notin \Psi$. Note that $|\mathcal{D}|(x) - W(x) = \kappa S(x)$ for $x \in \mathsf{supp}\,(|\mathcal{D}|)$ by Item 1. Thus,

$$\sum_{x \in I \cap \mathsf{supp}(|\mathcal{D}|) \cap \Psi} S(x) \leq \sum_{x \in I \cap \mathsf{supp}(|\mathcal{D}|) \cap \Psi} \frac{|\mathcal{D}_\Psi|(x) - W(x)}{\kappa} \qquad \text{(by Item 3)}$$

$$\begin{aligned}
&\leq \frac{2}{\kappa} \cdot \big\| |\mathcal{D}_\Psi| - W \big\|_{\mathsf{TV}} \\
&\leq \frac{2}{\kappa} \cdot \| \mathcal{D}_\Psi - f(\mathcal{U}^m) \|_{\mathsf{TV}} && \text{(by data processing inequality)} \\
&\leq \frac{2}{\kappa} \cdot O(\lambda) = O(1). && \text{(by (10) \& (11))}
\end{aligned}$$

for some sufficiently small implicit constant. Combining with (12) yields

$$\sum_{x \in I \cap \mathsf{supp}(|\mathcal{D}|) \cap \Psi^c} S(x) = \Omega(1).$$

However for any $x \in I \cap \mathsf{supp}\,(|\mathcal{D}|) \cap \Psi^c$, we have

$$S(x) = \frac{|\mathcal{D}|(x)}{\kappa} = \Theta\left(\frac{1}{\kappa\sqrt{n}}\right) = \Theta\left(\frac{1}{\lambda\sqrt{n}}\right).$$

In conclusion, it must be the case that there are $\Theta(\lambda\sqrt{n})$ many $x \in I \cap \mathsf{supp}\,(|\mathcal{D}|)$ with $x \notin \Psi$.

We consider a procedure where we take these elements one at a time and pair them with the closest unpaired element of $I \cap \Psi$ with the same parity. As each element only has to avoid at most $O(\lambda\sqrt{n})$ elements that are either not in Ψ or have already been paired, we end up with $\Theta(\lambda\sqrt{n})$ disjoint pairs of elements $(x_i, y_i) \in I \cap \mathsf{supp}\,(|\mathcal{D}|)$ so that $x_i \notin \Psi$, $y_i \in \Psi$, $|x_i - y_i| = O(\lambda\sqrt{n})$, and x_i and y_i have the same parity. Note for any such $y_i \in I \cap \Psi$, Fact 3.7 and Fact 3.8 imply

$$|\mathcal{D}_\Psi|(y_i) \geq 2^{-n}\binom{n}{\frac{n}{2} + O(\sqrt{n})} \geq \frac{2^{n\left(\mathcal{H}\left(\frac{1}{2} - O\left(n^{-1/2}\right)\right) - 1\right)}}{\sqrt{2n\left(1 - O\left(n^{-1}\right)\right)}} \geq \Omega\left(\frac{1}{\sqrt{n}}\right). \tag{13}$$

Now we apply Proposition 4.9 to write $W = aE + (1-a)X$ where $a = O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}})$ and

$$\big|X(x_i) - X(y_i)\big| = O_d(\lambda/\sqrt{n}) \tag{14}$$

for each i. In particular, we have that

$$\begin{aligned}
\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} &\geq \big\| |f(\mathcal{U}^m)| - |\mathcal{D}_\Psi| \big\|_{\mathsf{TV}} && \text{(by data processing inequality)} \\
&\geq \frac{1}{2}\sum_i \big|W(x_i) - |\mathcal{D}_\Psi|(x_i)\big| + \big|W(y_i) - |\mathcal{D}_\Psi|(y_i)\big| \\
&\geq \frac{1}{2}\sum_i \big||\mathcal{D}_\Psi|(x_i) - |\mathcal{D}_\Psi|(y_i)\big| - \big|W(x_i) - W(y_i)\big| && \text{(by triangle inequality)} \\
&\geq \frac{1}{2}\sum_i |\mathcal{D}_\Psi|(y_i) - (1-a)\big|X(x_i) - X(y_i)\big| - a\big|E(x_i) - E(y_i)\big| && \text{(since } x_i \notin \Psi\text{)} \\
&\geq \frac{1}{2}\sum_i \Omega\left(\frac{1}{\sqrt{n}}\right) - \big|X(x_i) - X(y_i)\big| + a\left(\big|X(x_i) - X(y_i)\big| - \big|E(x_i) - E(y_i)\big|\right) \\
& && \text{(by (13))} \\
&\geq \frac{1}{2}\sum_i \Omega\left(\frac{1}{\sqrt{n}}\right) - O_d\left(\frac{\lambda}{\sqrt{n}}\right) - a\left[|X(x_i) - E(x_i)| + |X(y_i) - E(y_i)|\right] \\
& && \text{(by (14) and triangle inequality)} \\
&= \Theta(\lambda\sqrt{n}) \cdot \Omega\left(\frac{1}{\sqrt{n}}\right) - a\sum_{i=0}^{n} W(i)
\end{aligned}$$

$$= \Omega(\lambda) - O\left(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}\right).$$

That is, $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} = \Omega(\lambda)$, which contradicts our assumption that $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} = O(\lambda)$ for a sufficiently small implicit constant. □

5 Local Limit Theorem

In this section, we will prove that any low-depth sampleable distribution that is reasonably close (in total variation distance) to a roughly centered uniform symmetric distribution is also close to a mixture of `evens` and `odds`. More precisely, we show:

Theorem (Theorem 4.7 Restated)**.** *Let $\delta > 0$ and $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of δ and d). Let $\Psi \subseteq \{0,1,2,\dots,n\}$ be a set containing some element $n(1/2 \pm c(d,\delta))$ for some $c(d,\delta) > 0$ a small enough function of d and δ. Then there exists a distribution $\mathcal{M}$ which is a mixture of* `evens` *and* `odds` *so that*

$$\|f(\mathcal{U}^m) - \mathcal{M}\|_{\mathsf{TV}} \le O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) + \delta.$$

The proof of this result is in two steps. First, we will show (Proposition 5.1) that the output weight distribution $|f(\mathcal{U}^m)|$ is close in Kolmogorov distance to the binomial distribution, even when restricting the weight's parity. Then we will show (Proposition 4.9) that the distribution on $|f(\mathcal{U}^m)|$ satisfies a continuity property. Together these will imply that the distribution on the weights of $f(\mathcal{U}^m)$ is comparable to those of $\mathcal{M}$. To obtain the final result, we apply the following lemma.

Lemma (Lemma 4.8 Restated)**.** *Let A and B be two distributions on $\{0,1\}^n$ with B symmetric. Then*

$$\|A - B\|_{\mathsf{TV}} = \Theta(\|A - A_{\mathrm{sym}}\|_{\mathsf{TV}} + \||A| - |B|\|_{\mathsf{TV}}).$$

Proof. First, we prove the upper bound. By the triangle inequality,

$$\|A - B\|_{\mathsf{TV}} \le \|A - A_{\mathrm{sym}}\|_{\mathsf{TV}} + \|A_{\mathrm{sym}} - B\|_{\mathsf{TV}} = \|A - A_{\mathrm{sym}}\|_{\mathsf{TV}} + \||A| - |B|\|_{\mathsf{TV}}.$$

For the lower bound, we note on the one hand by the data processing inequality that

$$\|A - B\|_{\mathsf{TV}} \ge \||A| - |B|\|_{\mathsf{TV}}.$$

On the other hand by the triangle inequality,

$$\|A - B\|_{\mathsf{TV}} \ge \|A - A_{\mathrm{sym}}\|_{\mathsf{TV}} - \|A_{\mathrm{sym}} - B\|_{\mathsf{TV}} \ge \|A - A_{\mathrm{sym}}\|_{\mathsf{TV}} - \||A| - |B|\|_{\mathsf{TV}}.$$

Combining, we find

$$\begin{aligned} 3\,\|A - B\|_{\mathsf{TV}} &\ge (\|A - A_{\mathrm{sym}}\|_{\mathsf{TV}} - \||A| - |B|\|_{\mathsf{TV}}) + 2\,\||A| - |B|\|_{\mathsf{TV}} \\ &= \|A - A_{\mathrm{sym}}\|_{\mathsf{TV}} + \||A| - |B|\|_{\mathsf{TV}}. \end{aligned}$$

□

5.1 The Kolmogorov Bound

In this section, we prove the following result.

Proposition 5.1. *Let $\delta > 0$ and $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of δ and d). Let $\Psi \subseteq \{0,1,\dots,n\}$ be a set containing some element $n(1/2 \pm c(d,\delta))$ for some $c(d,\delta) > 0$ a small enough function of d and δ. Then the distribution $f(\mathcal{U}^m)$ can be written as a mixture $aE + (1-a)X$ with $a = O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}})$ so that for some $\eta \in [0,1]$ and all $t \in \mathbb{R}$:*

$$\left| \mathbf{Pr}\left[|X| > t \text{ and } |X| \text{ is even}\right] - \eta\, \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] \right| = O(\delta),$$

and

$$\left| \mathbf{Pr}\left[|X| > t \text{ and } |X| \text{ is odd}\right] - (1-\eta)\, \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] \right| = O(\delta).$$

We briefly note the necessity of a in the above decomposition of $f(\mathcal{U}^m)$. It is possible for $f(\mathcal{U}^m)$ to have some small part that is far from all such $\mathcal{D}_\Psi$ (where "small" depends on the distance between $f(\mathcal{U}^m)$ and $\mathcal{D}_\Psi$), but the remainder of the distribution must have weight close in Kolmogorov distance to the binomial distribution $|\mathcal{U}^n|$, even accounting for parity.

We begin by proving a special case. Recall a distribution $\mathcal{D}$ over $\{0,1\}^n$ is k-wise independent if the projection of $\mathcal{D}$ onto any $k' \le k$ indices is uniformly distributed over $\{0,1\}^{k'}$. The following result handles the case where $f(\mathcal{U}^m)$ is nearly k-wise independent and none of the input degrees are too large.

Lemma 5.2. *Let $k \ge 2$ be an integer, and let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large. Suppose that no input bit of f affects more than n/A output bits for some $A > 0$. Suppose furthermore that there is a set $S \subseteq [n]$ of size $|S| \le \sqrt{n/A}$ such that $f(\mathcal{U}^m)$ is k-wise independent on $[n] \setminus S$. Then there exists $\eta \in [0,1]$ so that for any $\delta \in (0,1/2)$ and $t \in \mathbb{R}$, we have*

$$\left| \mathbf{Pr}\left[|f(\mathcal{U}^m)| > t \text{ and } |f(\mathcal{U}^m)| \text{ is even}\right] - \eta\, \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] \right|$$

and

$$\left| \mathbf{Pr}\left[|f(\mathcal{U}^m)| > t \text{ and } |f(\mathcal{U}^m)| \text{ is odd}\right] - (1-\eta)\, \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] \right|$$

are at most

$$O\left(\frac{\log(1/\delta)^{O(d)^d}}{\sqrt{A\delta}} + \frac{\log(k)}{\sqrt{k}} + \delta \right).$$

In particular, we note that as long as k and A are sufficiently large, we can make the error as small as we like.

Proof. The basic idea is to use the fact that k-wise independence fools threshold functions. In particular, the following result says that if the output coordinates of $f(\mathcal{U}^m)$ are k-wise independent, then the probability that $|f(\mathcal{U}^m)|$ (which is simply the sum of those outputs) is bigger than t will approximate the same expression for the binomial distribution.

Theorem 5.3 ([DGJ+10])**.** *Let $\mathcal{D}$ be a k-wise independent distribution on $\{0,1\}^n$, and let $\mathcal{U}^n$ be the uniform distribution over $\{0,1\}^n$. Then for any $w_1, \dots, w_n, \theta \in \mathbb{R}$, we have*

$$\left| \mathop{\mathbf{Pr}}_{x \sim \mathcal{D}}[w_1 x_1 + \cdots + w_n x_n \ge \theta] - \mathop{\mathbf{Pr}}_{x \sim \mathcal{U}^n}[w_1 x_1 + \cdots + w_n x_n \ge \theta] \right| \le O\left(\frac{\log(k)}{\sqrt{k}} \right).$$

To deal with the parities, we observe that the parity of $|f(\mathcal{U}^m)|$ is a degree at most d polynomial over $\mathbb{F}_2$ in the input bits. Using the following consequence of [CHH+20, Theorem 3.1], we have that randomizing a small number of input bits can be used to effectively re-randomize the parity of the output.

Theorem 5.4. *Let p be a degree-d polynomial over $\mathbb{F}_2^n$ and $\delta \in (0,1/2)$. There exists a subset $R \subseteq [n]$ with $|R| \leq \log(1/\delta)^{O(d)^d}$ so that if we write $p(x) = p(x_{R^c}, x_R)$ where x_R and x_{R^c} are the coordinates in R and not in R respectively, then with probability at least $1-\delta$ over the choice of a random value of x_{R^c} we have that*

$$\left| \mathop{\mathbf{Pr}}_{x_R} [p(x_{R^c}, x_R) = 1] - \mathop{\mathbf{Pr}}_{x} [p(x) = 1] \right| < \delta. \tag{15}$$

That is, if we randomly fix the coordinates in R^c, then with high probability re-randomizing the coordinates in R will essentially re-randomize the output of p.

Now, let the parity of the output of $f(x)$ be given by the $\mathbb{F}_2$-polynomial $p(x)$. Applying Theorem 5.4 gives a set R of at most $\log(1/\delta)^{O(d)^d}$ input bits. Let $B \subseteq [n]$ be the set of output bits affected by these input bits along with the output bits in S. Note that by our assumptions, $|B| \leq n\log(1/\delta)^{O(d)^d}/A$.

We set η to be the probability that $p(x) = 0$ over $x \sim \mathcal{U}^m$ and will show that

$$\mathbf{Pr}\left[|f(\mathcal{U}^m)| > t \text{ and } |f(\mathcal{U}^m)| \text{ is even}\right] \geq \eta\, \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] - O\left(\frac{\log(1/\delta)^{O(d)^d}}{\sqrt{A\delta}} + \frac{\log(k)}{\sqrt{k}} + \delta\right).$$

Observe that the case of odd $|f(\mathcal{U}^m)|$ is almost identical, while the upper bounds follow from considering the complement distribution $1^n - f(\mathcal{U}^m)$.

Before proceeding further, we define several variables and events for the sake of future clarity. Let $C := |B|/2 - \sqrt{|B|/\delta} - |S|$. Let EVEN be the event that $|f(\mathcal{U}^m)|$ is even (and similarly for ODD). Additionally, let BIG be the event that $|f(\mathcal{U}^m)[B^c]| > t - C$. Finally, let GOOD be the event that x_{R^c} satisfies (15) (and BAD be the complement event). We have

$$\begin{aligned}
\mathbf{Pr}\left[|f(\mathcal{U}^m)| > t \text{ and } \mathsf{EVEN}\right] &\geq \mathbf{Pr}\left[\mathsf{BIG} \text{ and } |f(\mathcal{U}^m)[B]| \geq C \text{ and } \mathsf{EVEN}\right] \\
&= \mathbf{Pr}\left[\mathsf{BIG} \text{ and } \mathsf{EVEN}\right] - \mathbf{Pr}\left[\mathsf{BIG} \text{ and } |f(\mathcal{U}^m)[B]| < C \text{ and } \mathsf{EVEN}\right] \\
&\geq \mathbf{Pr}\left[\mathsf{BIG}\right] \cdot \mathbf{Pr}\left[\mathsf{EVEN} \;\middle|\; \mathsf{BIG}\right] - \mathbf{Pr}\left[|f(\mathcal{U}^m)[B]| < C\right].
\end{aligned}$$

Since the coordinates of $f(\mathcal{U}^m)$ not in B are k-wise independent by assumption, Theorem 5.3 implies

$$\begin{aligned}
\mathbf{Pr}\left[\mathsf{BIG}\right] &\geq \mathbf{Pr}\left[|\mathcal{U}^{n-|B|}| > t - C\right] - O\left(\frac{\log(k)}{\sqrt{k}}\right) \\
&= \mathbf{Pr}\left[|\mathcal{U}^n| > t - C + |\mathcal{U}^{|B|}|\right] - O\left(\frac{\log(k)}{\sqrt{k}}\right) \\
&\geq \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] - \mathbf{Pr}\left[t \leq |\mathcal{U}^n| \leq t - C + |\mathcal{U}^{|B|}|\right] - O\left(\frac{\log(k)}{\sqrt{k}}\right) \\
&\geq \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] - \max_x |\mathcal{U}^n|(x) \cdot \mathbb{E}\left[\left||\mathcal{U}^{|B|}| - C\right|\right] - O\left(\frac{\log(k)}{\sqrt{k}}\right) \\
&\geq \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] - O\left(\frac{|C - \frac{|B|}{2}| + \sqrt{|B|}}{\sqrt{n}} + \frac{\log(k)}{\sqrt{k}}\right).
\end{aligned} \tag{16}$$

To lower bound the conditional probability $\mathbf{Pr}\left[\mathsf{EVEN} \;\middle|\; \mathsf{BIG}\right]$, it will be slightly more convenient to upper bound $\mathbf{Pr}\left[\mathsf{ODD} \;\middle|\; \mathsf{BIG}\right]$. For this, observe that the event BIG does not depend on any of the input bits in R. Hence by Theorem 5.4, for all but a δ-fraction of the settings of the bits in R^c, the probability over the bits in R that $|f(\mathcal{U}^m)|$ is odd is at most $(1-\eta) + \delta$. Therefore,

$$\mathbf{Pr}\left[\mathsf{ODD} \;\middle|\; \mathsf{BIG}\right] = \mathbf{Pr}\left[\mathsf{GOOD} \;\middle|\; \mathsf{BIG}\right] \cdot \mathbf{Pr}\left[\mathsf{ODD} \;\middle|\; \mathsf{GOOD} \text{ and } \mathsf{BIG}\right]$$

$$
\begin{aligned}
&+ \mathbf{Pr}\left[\mathsf{BAD} \mid \mathsf{BIG}\right] \cdot \mathbf{Pr}\left[\mathsf{ODD} \mid \mathsf{BAD} \text{ and } \mathsf{BIG}\right] \\
&\leq \mathbf{Pr}\left[\mathsf{ODD} \mid \mathsf{GOOD}\right] + \mathbf{Pr}\left[\mathsf{BAD} \mid \mathsf{BIG}\right] \leq (1-\eta) + \delta + \frac{\delta}{\mathbf{Pr}\left[\mathsf{BIG}\right]}.
\end{aligned}
$$

Combining with (16), we have that

$$
\begin{aligned}
\mathbf{Pr}\left[\mathsf{BIG}\right] \mathbf{Pr}\left[\mathsf{EVEN} \mid \mathsf{BIG}\right] &\geq \eta\, \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] - O\left(\frac{|C - \frac{|B|}{2}| + \sqrt{|B|}}{\sqrt{n}} + \frac{\log(k)}{\sqrt{k}} + \delta\right) \\
&\geq \eta\, \mathbf{Pr}\left[|\mathcal{U}^n| > t\right] - O\left(\frac{\log(1/\delta)^{O(d)^d}}{\sqrt{A\delta}} + \frac{\log(k)}{\sqrt{k}} + \delta\right).
\end{aligned}
$$

To complete our proof, it remains to bound $\mathbf{Pr}\left[|f(\mathcal{U}^m)[B]| < C\right]$. We have

$$
\begin{aligned}
\mathbf{Pr}\left[|f(\mathcal{U}^m)[B]| < C\right] &\leq \mathbf{Pr}\left[|f(\mathcal{U}^m)[B \setminus S]| < |B|/2 - \sqrt{|B|/\delta} - |S|\right] \\
&\leq \mathbf{Pr}\left[|f(\mathcal{U}^m)[B \setminus S]| < |B \setminus S|/2 - \sqrt{|B \setminus S|/\delta}\right] < \delta,
\end{aligned}
$$

where the final inequality follows from Chebyshev's inequality and the observation that the outputs in $B \setminus S$ are 2-wise independent. □

We note that the degree bound in Lemma 5.2 can be achieved by restricting high degree inputs, which d-locality guarantees relatively few of. To reduce to the case where most output coordinates are k-wise independent, we use the following lemma.

Lemma 5.5. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function, k be a positive integer, and $\delta, \kappa \in (0,1]$. Additionally, let $\Psi \subseteq \{0,1,\ldots,n\}$ contain some element within n/c of $n/2$ for some c sufficiently large given the values of $d, k, \delta,$ and κ. Then there exists a partition of $\{0,1\}^m$ into subcubes of three types so that:*

1. *For each subcube C of Type-I there is a set of $O_{d,k,\delta,\kappa}(1)$ coordinates in $[n]$ so that the other coordinates of $f(\mathcal{U}(C))$ are k-wise independent.*
2. *Letting W be the union of subcubes of Type-II, we have that $\|f(\mathcal{U}(W)) - \mathcal{D}_\Psi\|_{\mathsf{TV}} > 1 - \kappa$.*
3. *The total probability mass of all subcubes of Type-III is at most δ.*

We will later apply Lemma 5.5 to each subcube generated by conditioning on high degree input coordinates. This partitions the input space $\{0,1\}^m$ into subcubes, where almost all the mass is on subcubes C such that either $f(\mathcal{U}(C))$ is nearly k-wise independent with bounded degree inputs (in which case Lemma 5.2 is applicable) or whose union W has $f(\mathcal{U}(W))$ far from $\mathcal{D}_\Psi$.

Our proof will require the following two consequences of hypercontractivity to analyze bounded-degree polynomials. The first follows from combining Lemma 3.6 with Markov's inequality. Recall for $f\colon \{-1,1\}^n \to \mathbb{R}$, we define the norm $\|f\|_2 = \left(\mathbb{E}_{X\sim\{-1,1\}^n}\left[|f(X)|^2\right]\right)^{1/2}$.[4]

Lemma 5.6 (See, e.g., [Kan17, Corollary 26])**.** *Let $K > 0$. If $p\colon \{\pm1\}^n \to \{\pm1\}$ is a degree-d polynomial, then*

$$
\mathop{\mathbf{Pr}}_{X\sim\{-1,1\}^n}\left[|p(X)| > K \cdot \|p\|_2\right] = O\left(2^{-(K/2)^{2/d}}\right).
$$

[4] We originally defined the norm for functions on the domain $\{0,1\}^n$, but it will be more convenient in the subsequent proof to define it for the domain $\{\pm1\}^n$.

The second is a consequence of Lemma 3.6 and the Paley-Zygmund inequality.

Lemma 5.7 (See, e.g., [KKL17, Theorem 2.4])**.** *If $p\colon \{\pm 1\}^n \to \{\pm 1\}$ is a degree-d multilinear polynomial, then*

$$\mathop{\mathbf{Pr}}_{X\sim\{-1,1\}^n}\left[|p(X)| \geq (1/2)\cdot\|p\|_2\right] \geq (1/2)\cdot 9^{-d}.$$

Proof of Lemma 5.5. It will be convenient to view f as a function $\{\pm 1\}^m \to \{\pm 1\}^n$ rather than $\{0,1\}^m \to \{0,1\}^n$ (as well as $\mathcal{U}^n$ as the uniform distribution over $\{\pm 1\}^n$ and similarly for $\mathcal{D}_\Psi$). As this change does not affect any of our claims, we will assume it throughout the following. We also assume throughout that n is at least a sufficiently large function of d, k, δ, κ as otherwise, we can simply assign $\{\pm 1\}^m$ to be a single cube of Type-I, noting that there are at most n coordinates which are not k-wise independent.

We will show the lemma statement holds via induction on $\lceil \log(1/\delta)/\log(1/(1-2^{-Bkd}))\rceil$ for some sufficiently large constant B and all d, k, δ, κ. For the base case $\lceil \log(1/\delta)/\log(1/(1-2^{-Bkd}))\rceil = 0$, it must be that $\delta = 1$, so we declare all of $\{\pm 1\}^m$ to be a cube of Type-III. For the inductive step, we assume that our result holds for all d', k', δ', κ satisfying

$$\left\lceil \frac{\log\left(\frac{1}{\delta'}\right)}{\log\left(\frac{1}{1-2^{-Bk'd'}}\right)} \right\rceil < \left\lceil \frac{\log\left(\frac{1}{\delta}\right)}{\log\left(\frac{1}{1-2^{-Bkd}}\right)} \right\rceil.$$

Our basic strategy will be an iterative decomposition of $\{\pm 1\}^m$.

Consider all of the non-constant monomials of degree at most k in the output bits whose expectations when evaluated on $f(\mathcal{U}^m)$ are non-zero. Take $N = 2^{6kd}\log^k(L/\kappa)$ for L some sufficiently large constant. If there are fewer than N such monomials, we can declare all of $\{\pm 1\}^m$ a single subcube of Type-I, noting that other than the at most Nk coordinates involved in these monomials, the outputs of $f(\mathcal{U}^m)$ are k-wise independent. Otherwise let p be the sum of N of these monomials times the sign of their individual expectations.

Observe that each of these monomials is a function of at most kd input bits, so its expectation is a multiple of 2^{-kd}. In particular, this implies that $\mathbb{E}[p(f(\mathcal{U}^m))] \geq N2^{-kd}$. On the other hand, since it has no constant term, $\mathbb{E}[p(\mathcal{U}^n)] = 0$ and $\mathrm{Var}(p(\mathcal{U}^n)) = N$. Hence Lemma 5.6 implies a concentration bound:

$$\mathbf{Pr}\left[|p(\mathcal{U}^n)| > N2^{-kd-1}\right] = \mathbf{Pr}\left[|p(\mathcal{U}^n)| > \sqrt{N}2^{-kd-1}\cdot\|p\|_2\right] \leq O\left(2^{-(N2^{-2kd-4})^{1/k}}\right) < \kappa/10,$$

where the final inequality follows from choosing L (and thus N) sufficiently large. Additionally, we claim the restriction of $\mathcal{D}_\Psi$ to the at most Nk bits that p depends on is within a factor of two of uniform. Indeed, let $Z \subseteq [n]$ be the set of output bits affecting p.

Claim 5.8. For all $x \in \{\pm 1\}^Z$ and n sufficiently large, we have $\mathcal{D}_\Psi[Z](x) \leq 2\cdot\mathcal{U}^Z(x)$.

For clarity, we will finish the remainder of the proof before proving Claim 5.8. The claim implies

$$\mathbf{Pr}\left[|p(\mathcal{D}_\Psi)| > N2^{-kd-1}\right] < \kappa/5. \tag{17}$$

Moreover, take the at most Nkd input coordinates that affect the value of p. We split $\{\pm 1\}^m$ into subcubes based on all possible conditionings of these coordinates. We claim that at least a decent fraction of them can be assigned as Type-II. In particular, note that $p(f(x))$ is some polynomial $q(x)$ of degree at most kd satisfying $\mathbb{E}[q(\mathcal{U}^m)] \geq N2^{-kd}$. Therefore by Lemma 5.7, with probability at least 2^{-Bkd} over this conditioning (which completely determines the value of q), the resulting subcube C satisfies

$$|q(C)| \geq \|q(\mathcal{U}^m)\|_2/2 \geq \mathbb{E}[q(\mathcal{U}^m)]/2 \geq N2^{-kd-1},$$

where the second inequality follows from Jensen's inequality.

Hence, for at least a 2^{-Bkd}-fraction of these subcubes we have that $|q| \geq N2^{-kd-1}$. In particular, (17) implies the uniform distribution over the union of these subcubes is $(1-\kappa/5)$-far from $\mathcal{D}_\Psi$. We declare all of these subcubes to be Type-II, and use our inductive hypothesis on each of the remaining $M \leq 2^{Nkd}$ subcubes with parameters $\kappa' = \kappa/(5M)$ and $\delta' = \delta/(1-2^{-Bkd})$, noting that

$$\left\lceil \frac{\log\left(\frac{1}{\delta'}\right)}{\log\left(\frac{1}{1-2^{-Bkd}}\right)} \right\rceil = \left\lceil \frac{\log\left(\frac{1}{\delta}\right) + \log\left(1-2^{-Bkd}\right)}{\log\left(\frac{1}{1-2^{-Bkd}}\right)} \right\rceil = \left\lceil \frac{\log\left(\frac{1}{\delta}\right)}{\log\left(\frac{1}{1-2^{-Bkd}}\right)} \right\rceil - 1.$$

(If $M = 0$, we have already obtained our desired partition.) For each subcube i for $i = 1, 2, \ldots, M$, this gives a partition into three types such that

1. For each subcube C of Type-I there is a set of $O_{d,k,\delta,\kappa}(1)$ coordinates in $[n]$ so that the other coordinates of $f(\mathcal{U}(C))$ are k-wise independent.
2. Letting W_i be the union of subcubes of Type-II, we have that $\|f(\mathcal{U}(W_i)) - \mathcal{D}_\Psi\|_{\mathsf{TV}} > 1 - \frac{\kappa}{5M}$.
3. The total probability mass of all subcubes of Type-III is at most δ'.

It remains to combine these partitions together, along with the union W' of Type-II subcubes C with $|q(C)| \geq N2^{-kd-1}$. The Type-I subcubes all remain Type-I during this combination. For Type-II subcubes, we have already shown that $\|f(\mathcal{U}(W')) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - \frac{\kappa}{5}$. Appealing to Lemma 3.3, we find that the union over all Type-II subcubes satisfies

$$\left\|f(\mathcal{U}(W_1 \cup \cdots \cup W_M \cup W')) - \mathcal{D}_\Psi\right\|_{\mathsf{TV}} \geq 1 - (M+2) \cdot \frac{\kappa}{5M} \geq 1 - \kappa.$$

We conclude by calculating that the total probability mass of all subcubes of Type-III is at most $(1-2^{-Bkd})\delta' = \delta$, as desired. □

We now prove Claim 5.8, showing that $\mathcal{D}_\Psi[Z](x) \leq 2 \cdot \mathcal{U}^Z(x)$ for all $x \in \{\pm 1\}^Z$. This allowed us to reason about the concentration of $|p(\mathcal{D}_\Psi)|$ using information about the concentration of $|p(\mathcal{U})|$.

Proof of Claim 5.8. For this proof, it will be more convenient to revert back to working with $\{0, 1\}$ rather than $\{\pm 1\}$. We may restrict our attention to the weights $\overline{\Psi} = \{s \in \Psi : |s - n/2| \leq 2n/c\}$ at a small cost. Indeed, Fact 3.9 implies

$$A := \sum_{s \in \Psi : |s-n/2| > \frac{2n}{c}} \binom{n}{s} \leq 2 \cdot \sum_{0 \leq s < \frac{n}{2} - \frac{2n}{c}} \binom{n}{s} \leq 2 \cdot 2^{n \cdot \mathcal{H}\left(\frac{1}{2} - \frac{2}{c}\right)}. \tag{18}$$

Then by Fact 3.7, we have

$$B := \binom{n}{\iota(\Psi)} \geq \binom{n}{\frac{n}{2} - \frac{n}{c}} \geq \frac{2^{n \cdot \mathcal{H}\left(\frac{1}{2} - \frac{1}{c}\right)}}{\sqrt{8n\left(\frac{1}{2} - \frac{1}{c}\right)^2}}. \tag{19}$$

Thus,

$$\mathop{\mathbf{Pr}}_{s \sim \mathcal{D}_\Psi}\left[\left|s - \frac{n}{2}\right| > \frac{2n}{c}\right] = \frac{A}{|\mathsf{supp}\left(\mathcal{D}_\Psi\right)|} \leq \frac{A}{B} \leq \sqrt{8n\left(\frac{1}{2} - \frac{1}{c}\right)^2} \cdot 2^{n\left(\mathcal{H}\left(\frac{1}{2} - \frac{2}{c}\right) - \mathcal{H}\left(\frac{1}{2} - \frac{1}{c}\right)\right)}.$$

By Fact 3.8, we have

$$\begin{aligned}\mathcal{H}\left(\frac{1}{2}-\frac{2}{c}\right)-\mathcal{H}\left(\frac{1}{2}-\frac{1}{c}\right) &= \frac{1}{2\ln(2)}\sum_{m\geq 1}\frac{(2/c)^{2m}-(4/c)^{2m}}{m\cdot(2m-1)}\\ &\leq \frac{(2/c)^{2}-(4/c)^{2}}{2\ln(2)} \leq -\frac{1}{c^{2}}.\end{aligned}$$

Hence

$$\mathop{\mathbf{Pr}}_{s\sim\mathcal{D}_{\Psi}}\left[\left|s-\frac{n}{2}\right|>\frac{2n}{c}\right] \leq \sqrt{8n\left(\frac{1}{2}-\frac{1}{c}\right)^{2}}\cdot 2^{-n/c^{2}}.$$

For sufficiently large c and n, this implies

$$\mathcal{D}_{\Psi}[Z](x) \leq \mathop{\mathbf{Pr}}_{s\sim\mathcal{D}_{\Psi}}\left[|s-n/2|>2n/c\right]+\mathcal{D}_{\overline{\Psi}}[Z](x) \leq \frac{1}{2}\cdot\mathcal{U}^{Z}(x)+\mathcal{D}_{\overline{\Psi}}[Z](x),$$

so it remains to show $\mathcal{D}_{\overline{\Psi}}[Z](x) \leq (3/2)\cdot\mathcal{U}^{Z}(x)$. Assume without loss of generality $Z=[z]$. Then for all $x\in\{\pm1\}^{z}$ and $s\in\overline{\Psi}$, if we let $y\sim\mathcal{D}_{\{s\}}[Z]$, we have

$$\mathcal{D}_{\{s\}}[Z](x) = \mathbf{Pr}[y_1=x_1]\cdot\mathbf{Pr}\left[y_2=x_2\,|\,y_1=x_1\right]\cdots\mathbf{Pr}\left[y_z=x_z\,|\,y_i=x_i \text{ for all } i\in[z-1]\right].$$

We will show each factor is at most $\frac{1}{2}+\frac{\lambda}{z}$ for some sufficiently small constant $\lambda>0$, so that

$$\mathcal{D}_{\overline{\Psi}}[Z](x) \leq \max_{s\in\overline{\Psi}}\mathcal{D}_{\{s\}}[Z](x) \leq \left(\frac{1}{2}+\frac{\lambda}{z}\right)^{z} = 2^{-z}\cdot\left(1+\frac{2\lambda}{z}\right)^{z} \leq \mathcal{U}^{Z}(x)\cdot e^{2\lambda} \leq \frac{3}{2}\cdot\mathcal{U}^{Z}(x).$$

Consider the i-th factor in the case of $x_i=1$. (The case of $x_i=0$ is similar.) Then

$$\mathbf{Pr}\left[y_i=1\,|\,y_j=x_j \text{ for all } j\in[i-1]\right] = \frac{s-\sum_{j<i}\mathbb{1}(x_j=1)}{n-(i-1)} \leq \frac{\frac{n}{2}+\frac{2n}{c}}{n-z} \leq \frac{1}{2}+\frac{\lambda}{z}$$

for c,n sufficiently large. □

We are now ready to prove Proposition 5.1. Recall our goal is to show $f(\mathcal{U}^{m})$ can be written as a mixture $aE+(1-a)X$ with $a=O(\|f(\mathcal{U}^{m})-\mathcal{D}_{\Psi}\|_{\mathsf{TV}})$, where $|X|$ is $O(\delta)$-close in Kolmogorov distance to the binomial distribution $|\mathcal{U}^{n}|$, even accounting for parity. Our strategy will be to use the previously proven inductive decomposition lemma to partition the input space into structured subcubes. Notably, one type of resulting subcube is amiable to the application of our Kolmogorov distance / parity result (Lemma 5.2).

Proof of Proposition 5.1. We will show that for any $\delta>0$, we can achieve error $O(\delta)$ provided n is big enough. We assume $\delta\leq 1/2$ as otherwise we can take $a=0$, $X=f(\mathcal{U}^{m})$, and E arbitrary.

Let k be an integer at least $\log^{2}(1/\delta)/\delta^{2}$ and let $A=(1/\delta)^{3}\log(1/\delta)^{\Omega(d)^{d}}$ sufficiently large. We begin by conditioning on every input coordinate which affects more than n/A many output coordinates. Note that since the sum of the degrees of our coordinates is at most nd, the number M of these coordinates is at most Ad. This leaves us with a partition of $\{0,1\}^{m}$ into 2^{M} subcubes. On each of these subcubes we apply Lemma 5.5 with parameters d, k, δ, and $\kappa=\delta 2^{-M}/10$. This yields a partition of all of $\{0,1\}^{m}$ into subcubes of types I, II, and III so that:

1. For each subcube C of Type-I there is a set of $O_{d,k,\delta}(1)$ coordinates in $[n]$ so that the other coordinates of $f(\mathcal{U}(C))$ are k-wise independent.

2. Let W be the union of subcubes of Type-II. By Lemma 3.3 and the choice of κ, we have

$$\|f(\mathcal{U}(W)) - \mathcal{D}_\Psi\|_{\mathsf{TV}} > 1 - (2^M + 1)\kappa \geq 1 - \frac{\delta}{2}.$$

3. The total probability mass of all subcubes of Type-III is at most δ.

Recall that for each subcube C of Type-I, Lemma 5.2 guarantees some $\eta_C \in [0,1]$ such that as long as n is sufficiently large, we have

$$\left|\,\mathbf{Pr}\left[|f(\mathcal{U}(C))| > t \text{ and } |f(\mathcal{U}(C))| \text{ is even}\right] - \eta_C\,\mathbf{Pr}\left[|\mathcal{U}^n| > t\right]\,\right|$$

and

$$\left|\,\mathbf{Pr}\left[|f(\mathcal{U}(C))| > t \text{ and } |f(\mathcal{U}(C))| \text{ is odd}\right] - (1-\eta_C)\,\mathbf{Pr}\left[|\mathcal{U}^n| > t\right]\,\right|$$

are at most

$$O\left(\frac{\log(1/\delta)^{O(d)^d}}{\sqrt{A\delta}} + \frac{\log(k)}{\sqrt{k}} + \delta\right) = O(\delta).$$

Letting U be the union of all Type-I cubes and η be the appropriate weighted average of the η_C's, we see that

$$\left|\,\mathbf{Pr}\left[|f(\mathcal{U}(U))| > t \text{ and } |f(\mathcal{U}(U))| \text{ is even}\right] - \eta\,\mathbf{Pr}\left[|\mathcal{U}^n| > t\right]\,\right| = O(\delta),$$

and

$$\left|\,\mathbf{Pr}\left[|f(\mathcal{U}(U))| > t \text{ and } |f(\mathcal{U}(U))| \text{ is odd}\right] - (1-\eta)\,\mathbf{Pr}\left[|\mathcal{U}^n| > t\right]\,\right| = O(\delta).$$

Furthermore, let V be the union of U and all the Type-III cubes, as well as the Type-II cubes if their total mass is at most 4δ. We note that the same inequality will hold for $f(\mathcal{U}(V))$, which we set equal to X. If the total mass of the subcubes of Type-II is less than 4δ, we set $a = 0$ and are done. Otherwise, we set a to be the total mass of these Type-II subcubes and let $E = f(\mathcal{U}(W))$, where recall W is the union over the subcubes of Type-II. However, we see by Item 2 that $\|E - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - \delta/2$, so there exists an event $\mathcal{E}$ with mass at least $1-\delta$ in E but at most δ in $\mathcal{D}_\Psi$. Moreover $f(\mathcal{U}^m) = aE + (1-a)X$, so this event $\mathcal{E}$ has mass at least $a(1-\delta)$ in $f(\mathcal{U}^m)$. Hence,

$$\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq a(1-\delta) - \delta \geq a/4,$$

where we have used the fact that $a \geq 4\delta$ and $\delta \leq 1/2$. This establishes the upper bound on a and completes our proof. □

5.2 Continuity Bound

Here we prove the following proposition. It will later be combined with the Kolmogorov bound to prove our local limit theorem.

Proposition (Proposition 4.9 Restated)**.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function with n sufficiently large (in terms of d). Let $\Psi \subseteq \{0,1,\ldots,n\}$ be a set containing an element $n(1/2 \pm c(d))$ for some $c(d) > 0$ a small enough function of d. Then the distribution $f(\mathcal{U}^m)$ can be written as a mixture $aE + (1-a)X$ with $a = O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}})$ so that for any even Δ and $x \in \{0,1,\ldots,n\}$,*

$$\left|\,\mathbf{Pr}\left[|X| = x\right] - \mathbf{Pr}\left[|X| = x+\Delta\right]\,\right| = O_d\left(\frac{|\Delta|}{n}\right).$$

The proof requires machinery from [KOW24]. In particular, we recall the alternative viewpoint of a function as a hypergraph G on the output bits $[n]$ with an edge for each input bit containing all of the output bits that depend on it. Observe the locality assumption implies G has maximum degree at most d. As a consequence of [KOW24, Corollary 4.11] we have:

Corollary 5.9. *Let G be a hypergraph on n vertices with maximum degree at most d. For any increasing function $F\colon \mathbb{N} \to \mathbb{N}$, there exists a set S of edges in G whose removal yields at least $r = n/O_{d,F}(1)$[5] vertices in G whose neighborhoods have size at most $t = O_{d,F}(1)$ and are pairwise non-adjacent, and satisfies $|S| \le r/F(t)$.*

We will also need the following technical density comparison result, whose full generality and proof are deferred to Appendix A.

Theorem 5.10 (Special case of Theorem A.1)**.** *Let $t \ge 1$ be an integer, and let $X_1, \dots, X_n$ be independent random variables in $\{0, 1, \dots, t\}$. For each $i \in [n]$ and integer $r \ge 1$, define $p_{r,i} = \max_{x\in\mathbb{Z}} \mathbf{Pr}\left[X_i \equiv x \pmod r\right]$ and assume*

$$\sum_{i\in[n]} (1 - p_{r,i}) \ge \Omega_t(n) \quad \textit{holds for all } r \in \{3, 4, \dots, t\}.$$

Then for any $x \in \mathbb{Z}$ and even $\Delta \in \mathbb{Z}$, we have

$$\mathbf{Pr}\left[\sum_{i\in[n]} X_i = x\right] - \mathbf{Pr}\left[\sum_{i\in[n]} X_i = x + \Delta\right] \le O_t\left(\frac{|\Delta|}{n}\right).$$

Proof of Proposition 4.9. Let S be the set of input coordinates promised by Corollary 5.9, taking $F(t)$ to be a sufficiently large multiple of 2^{2dt}. Each setting of the variables in S produces a subcube of the inputs. We call a subcube C *weird* if for at least half of the neighborhoods of outputs promised by Corollary 5.9, the distribution of $f(\mathcal{U}(C))$ on those outputs is not uniform.

Claim 5.11. Any weird subcube C satisfies

$$\|f(\mathcal{U}(C)) - \mathcal{D}_\Psi\|_{\mathsf{TV}} > 1 - 2^{-\Omega(r\cdot 2^{-2dt})}.$$

Proof. By Corollary 5.9, the outputs of $f(\mathcal{U}(C))$ on any two of the resulting neighborhoods are independent. Moreover, each neighborhood only depends on at most dt input bits, so if its corresponding marginal distribution is not uniform, it must be at least 2^{-dt}-far from uniform. Furthermore, Fact 3.7 and Fact 3.8 imply

$$\eta := \min_x \frac{\mathcal{U}^n(x)}{\mathcal{D}_\Psi(x)} \ge \frac{\binom{n}{n\left(\frac{1}{2}\pm c(d)\right)}}{2^n} \ge \frac{2^{n\left(\mathcal{H}\left(\frac{1}{2}\pm c(d)\right)-1\right)}}{\sqrt{8n\left(\frac{1}{4} - c(d)^2\right)}} \ge 2^{-\Omega\left(n\cdot c(d)^2\right)}.$$

Choose $c(d)$ to be at most a small multiple of $\left(\frac{r}{n}\right)^{1/2} \cdot 2^{-dt}$. The claim then follows from applying Lemma 3.2 to deduce

$$\|f(\mathcal{U}(C)) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \ge 1 - 2\cdot e^{-2^{-2dt-2}\cdot r}/\eta > 1 - 2^{-\Omega(r\cdot 2^{-2dt})}. \qquad \square$$

[5] Recall $O_{d,F}(1)$ denotes a quantity whose value is constant once d and F are fixed.

Let w be the fraction of cubes (resulting from conditioning on the variables in S) that are weird. Combining our bound $|S| \leq r/F(t)$ and our choice of $F = \Omega(2^{2dt})$ with Lemma 3.3 yields

$$\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq w\left(1 - 2^{|S| - \Omega(r \cdot 2^{-2dt})}\right) \geq \frac{w}{2}.$$

Thus in our conclusion, we can take $a = w$, E to be the mixture of f applied to the weird subcubes, and X to be f applied to the mixture of the non-weird subcubes. It remains to show that

$$\left|\, \mathbf{Pr}\left[|f(\mathcal{U}(C))| = x\right] - \mathbf{Pr}\left[|f(\mathcal{U}(C))| = x + \Delta\right] \right| = O_d\left(\frac{|\Delta|}{n}\right)$$

for any non-weird subcube C. Taking the mixture will yield our result.

For a non-weird subcube C, consider the r non-adjacent neighborhoods promised to us by Corollary 5.9, and remove the (at most $r/2$) ones for which the output distribution is not uniform. By renaming which neighborhoods we are considering and decreasing r by a factor of 2, we may assume that no such neighborhoods exist.

Each neighborhood is $N(v_i)$ for some central element v_i. Take a random assignment of all of the input bits that do not affect some central element, which we call *extraneous inputs*. This fixes the value of every output bit not in one of these neighborhoods, and the weight of the output bits of v_i's neighborhood becomes a random variable X_i. Note that the total weight of the output of f is some constant plus the sum of the X_i's, which are independent. We would like to claim that Theorem 5.10 can be applied to this situation with high probability.

From here, we proceed similarly to [KOW24, Claims 5.16 & 5.23]. In particular, consider v_i's neighborhood N_i for some i and some integer modulus $3 \leq s \leq t$. Observe that because the distribution over $f(\mathcal{U}(C))[N_i]$ is uniform, the distribution of the weight modulo s conditioned on the v_i^{th} coordinate being 1 and conditioned on the v_i^{th} coordinate being 0 are not the same. On the other hand, conditioning on the extraneous inputs, this coordinate is equally likely to be 0 as 1. Hence it cannot be the case that X_i is always constant mod s, as this would imply that both the distribution of $|f(\mathcal{U}(C))[N_i]| \pmod{s}$ conditioned on $f(\mathcal{U}(C))[v_i] = 0$ and the distribution conditioned on $f(\mathcal{U}(C))[v_i] = 1$ would be equal to the distribution of $X_i \pmod{s}$.

However, after fixing the extraneous inputs, X_i only depends on at most d input bits. Thus, if it is not constant mod s, it must be at least 2^{-d}-far from constant. Furthermore, the bits in the neighborhood only depend on at most dt input bits, so with probability at least 2^{-dt} over the choice of values for the extraneous bits, X_i is at least 2^{-d}-far from constant mod s. Finally, note that since the neighborhoods are non-adjacent (after removing the edges in S), the extraneous bits used to determine X_i are disjoint from the extraneous bits used to determine X_j for $i \neq j$. Thus, whether or not X_i is constant mod s is independent of whether X_j is.

By Chernoff's inequality (Fact 3.5) and a union bound, except with probability $2^{-\Omega_d(n)}$ over the values of the extraneous bits, we have that there are $\Omega_d(n)$ neighborhoods where X_i is at least 2^{-d}-far from constant mod s for each $3 \leq s \leq t$. We can now apply Theorem 5.10 to show that if this event occurs, then the corresponding subcube C' defined by fixing the bits in S and the extraneous bits satisfies

$$\left|\mathbf{Pr}\left[|f(\mathcal{U}(C'))| = x\right] - \mathbf{Pr}\left[|f(\mathcal{U}(C'))| = x + \Delta\right]\right| = O_d\left(\frac{|\Delta|}{n}\right).$$

Taking the mixture over all such subcubes gives our result. □

5.3 Putting it Together

We are now prepared to prove Theorem 4.7 by combining our previous Kolmogorov and continuity bounds. Recall we wish to show the existence of a distribution $\mathcal{M}$ which is a mixture of $\texttt{evens}$ and $\texttt{odds}$ such that $\|f(\mathcal{U}^m) - \mathcal{M}\|_{\mathsf{TV}} \leq O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}) + \delta$ for any $\delta > 0$, provided n is sufficiently large in terms of d and δ.

Proof of Theorem 4.7. Let $c > 0$ be a small constant, and set $\delta' = \max\{\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}}/c, \delta\}$. Then we have

$$\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq c\delta'. \tag{20}$$

Note for any distribution $\mathcal{M}$ which is a mixture of $\texttt{evens}$ and $\texttt{odds}$, two applications of Lemma 4.8 yield

$$\begin{aligned}\|f(\mathcal{U}^m) - \mathcal{M}\|_{\mathsf{TV}} &= \Theta(\||f(\mathcal{U}^m)| - |\mathcal{M}|\|_{\mathsf{TV}}) + \Theta(\|f(\mathcal{U}^m)_{\mathrm{sym}} - f(\mathcal{U}^m)\|_{\mathsf{TV}}) \\ &= \Theta(\||f(\mathcal{U}^m)| - |\mathcal{M}|\|_{\mathsf{TV}}) + \Theta(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} - \||f(\mathcal{U}^m)| - |\mathcal{D}_\Psi|\|_{\mathsf{TV}}) \\ &= \Theta(\||f(\mathcal{U}^m)| - |\mathcal{M}|\|_{\mathsf{TV}}) + O(\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}})\end{aligned}$$

Thus, it suffices to construct a suitable $\mathcal{M}$ and show $\||f(\mathcal{U}^m)| - |\mathcal{M}|\|_{\mathsf{TV}} = O(\delta'')$ for some $\delta'' = \Theta(\delta')$ with a sufficiently small implicit constant, since

$$\Theta\left(\||f(\mathcal{U}^m)| - |\mathcal{M}|\|_{\mathsf{TV}}\right) = O(\delta'') \leq \delta' \leq \frac{1}{c} \cdot \|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} + \delta.$$

We begin by applying Proposition 5.1 with the δ from that proposition taken to be κ^2 for some $\kappa > 0$ sufficiently small in terms of d and δ''. Let $\mathcal{M} = \eta \cdot \texttt{evens} + (1-\eta) \cdot \texttt{odds}$ for the η coming from that result. Next, we partition $\{0, 1, \ldots, n\}$ into *parity intervals*, each of which is an interval of length $\Theta(\kappa\sqrt{n})$ intersected with either the set of even integers or the set of odd integers. In particular, we have by Proposition 5.1 and (20) that $f(\mathcal{U}^m)$ is $O(\delta'')$-close to a distribution X so that for any parity interval $\mathcal{I}$, we have that

$$\left|\, \mathbf{Pr}\left[|X| \in \mathcal{I}\right] - \mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right] \right| = O(\kappa^2). \tag{21}$$

We next apply Proposition 4.9 to find that $f(\mathcal{U}^m)$ is $O(\delta'')$-close to a distribution X' satisfying

$$\left|\, \mathbf{Pr}\left[|X'| = x\right] - \mathbf{Pr}\left[|X'| = x + \Delta\right] \right| = O_d(\Delta/n) \tag{22}$$

for any x and even Δ. Note also that $\|X - X'\|_{\mathsf{TV}} = O(\delta'')$. We claim that this is enough to show that $\||X'| - |\mathcal{M}|\|_{\mathsf{TV}}$ is small. In particular, we have that

$$\begin{aligned}\||X'| - |\mathcal{M}|\|_{\mathsf{TV}} &= \frac{1}{2}\sum_{x=0}^{n} \left| \mathbf{Pr}\left[|X'| = x\right] - \mathbf{Pr}\left[|\mathcal{M}| = x\right] \right| \\ &= \frac{1}{2}\sum_{\mathcal{I}}\sum_{x \in \mathcal{I}} \left| \mathbf{Pr}\left[|X'| = x\right] - \mathbf{Pr}\left[|\mathcal{M}| = x\right] \right| \\ &\leq \sum_{\mathcal{I}}\sum_{x \in \mathcal{I}} \left| \mathbf{Pr}\left[|X'| = x\right] - \frac{\mathbf{Pr}\left[|X'| \in \mathcal{I}\right]}{|\mathcal{I}|} \right| + \left| \frac{\mathbf{Pr}\left[|X'| \in \mathcal{I}\right] - \mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right]}{|\mathcal{I}|} \right| \\ &\qquad + \left| \mathbf{Pr}\left[|\mathcal{M}| = x\right] - \frac{\mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right]}{|\mathcal{I}|} \right|.\end{aligned} \tag{23}$$

To analyze this, we note that Fact 3.4 implies all but $O(\delta'')$ of the mass of $|\mathcal{M}|$ is supported on $O(\log(1/\delta'')/\kappa)$ many parity intervals, which we call *big*. The total mass that $|X|$ assigns to these big intervals is

$$\sum_{\text{big } \mathcal{I}} \mathbf{Pr}\left[|X| \in \mathcal{I}\right] = \sum_{\text{big } \mathcal{I}} \left(\mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right] + O(\kappa^2)\right) \qquad \text{(by (21))}$$

$$= 1 - O(\delta'') + O(\kappa^2 \log(1/\delta'')/\kappa) = 1 - O(\delta''),$$

with the final equality holding for κ sufficiently small. Therefore, since $\|X - X'\|_{\mathsf{TV}} = O(\delta'')$, X' also assigns all but an $O(\delta'')$-fraction of its mass to big intervals. Hence, up to this $O(\delta'')$ error, we can restrict the sum in (23) to big intervals. Thus,

$$\big\||X'| - |\mathcal{M}|\big\|_{\mathsf{TV}} \le O(\delta'') + \sum_{\text{big } \mathcal{I}} \sum_{x \in \mathcal{I}} \left|\mathbf{Pr}\left[|X'| = x\right] - \frac{\mathbf{Pr}\left[|X'| \in \mathcal{I}\right]}{|\mathcal{I}|}\right| + \left|\frac{\mathbf{Pr}\left[|X'| \in \mathcal{I}\right] - \mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right]}{|\mathcal{I}|}\right|$$

$$+ \left|\mathbf{Pr}\left[|\mathcal{M}| = x\right] - \frac{\mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right]}{|\mathcal{I}|}\right|.$$

Clearly,

$$\sum_{x \in \mathcal{I}} \left|\frac{\mathbf{Pr}\left[|X'| \in \mathcal{I}\right] - \mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right]}{|\mathcal{I}|}\right| = \left|\, \mathbf{Pr}\left[|X'| \in \mathcal{I}\right] - \mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right] \right|$$

$$\le \left|\, \mathbf{Pr}\left[|X'| \in \mathcal{I}\right] - \mathbf{Pr}\left[|X| \in \mathcal{I}\right] \right| + \left|\, \mathbf{Pr}\left[|X| \in \mathcal{I}\right] - \mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right] \right|.$$

The first term summed over all $\mathcal{I}$ is at most $\|X - X'\|_{\mathsf{TV}} = O(\delta'')$. The second term is at most $O(\kappa^2)$ by (21), so summed over all big intervals is $O(\delta'')$ (for small enough κ).

Additionally, note that

$$\left|\mathbf{Pr}\left[|X'| = x\right] - \frac{\mathbf{Pr}\left[|X'| \in \mathcal{I}\right]}{|\mathcal{I}|}\right| \le \max_{y \in \mathcal{I}} \left|\, \mathbf{Pr}\left[|X'| = x\right] - \mathbf{Pr}\left[|X'| = y\right] \right|.$$

Since x and y have the same parity in any parity interval, (22) implies this is at most

$$\max_{y \in \mathcal{I}} O_d\left(\frac{|x - y|}{n}\right) = O_d\left(\frac{\kappa\sqrt{n}}{n}\right) = O_d\left(\frac{\kappa}{\sqrt{n}}\right).$$

Summing this over all $x \in \mathcal{I}$ gives $O_d(\kappa^2)$, and summing over all big intervals gives $O_d(\kappa \log(1/\delta'')) = O(\delta'')$. The sum of the

$$\left|\mathbf{Pr}\left[|\mathcal{M}| = x\right] - \frac{\mathbf{Pr}\left[|\mathcal{M}| \in \mathcal{I}\right]}{|\mathcal{I}|}\right|$$

terms can be bounded similarly. We infer that $\||X'| - |\mathcal{M}|\|_{\mathsf{TV}} = O(\delta'')$, and thus by the triangle inequality

$$\||f(\mathcal{U}^m)| - |\mathcal{M}|\|_{\mathsf{TV}} \le \||f(\mathcal{U}^m)| - |X|\|_{\mathsf{TV}} + \big\||X| - |X'|\big\|_{\mathsf{TV}} + \big\||X'| - |\mathcal{M}|\big\|_{\mathsf{TV}} = O(\delta''),$$

completing our proof. □

Acknowledgments

We thank anonymous reviewers for helpful comments on an earlier version of this manuscript, and we thank Jason Gaitonde for pointing us to [CHH+20] on MathOverflow [Gai23]. KW wants to thank Tiancheng He for answering questions in probability theory. AO wants to thank Farzan Byramji, Shachar Lovett, and Jackson Morris for suggesting references about low-depth circuits.

References

[Bab87] Lászió Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Information Processing Letters*, 26(1):51–53, 1987. 2

[BDK+13] Olaf Beyersdorff, Samir Datta, Andreas Krebs, Meena Mahajan, Gido Scharfenberger-Fabian, Karteek Sreenivasaiah, Michael Thomas, and Heribert Vollmer. Verifying proofs in constant depth. *ACM Transactions on Computation Theory (TOCT)*, 5(1):1–23, 2013. 1

[BGK18] Sergey Bravyi, David Gosset, and Robert König. Quantum advantage with shallow circuits. *Science*, 362(6412):308–311, 2018. 1

[BGKT20] Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. Quantum advantage with noisy shallow circuits. *Nature Physics*, 16(10):1040–1045, 2020. 1

[BIL12] Chris Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for AC0-circuits. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 101–110. IEEE, 2012. 1

[BL87] Ravi B Boppana and Jeffrey C Lagarias. One-way functions and circuit complexity. *Information and Computation*, 74(3):226–240, 1987. 2

[Bon70] Aline Bonami. Étude des coefficients de fourier des fonctions de $L^p(G)$. In *Annales de l'institut Fourier*, volume 20, pages 335–402, 1970. 8

[Bru12] Franqis Brunault. Estimates for Bezout coefficients. MathOverflow, 2012. URL:https://mathoverflow.net/q/108723 (version: 2012-10-05). 39

[CGZ22] Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The space complexity of sampling. In *13th Innovations in Theoretical Computer Science Conference,(ITCS 2022)*, 2022. 1

[CHH+20] 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. 21, 31

[CS16] 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. 1

[CT06] Thomas M Cover and Joy A Thomas. Elements of information theory, 2006. 8

[CZ16] 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. 1

[DGJ+10] 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. 21

[DW12] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory (TOCT)*, 4(1):1–21, 2012. 1

[FLRS23] Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. Sampling and certifying symmetric functions. In *Approximation, Randomization, and Combinatorial Optimization. (APPROX/RANDOM)*, 2023. 1, 2, 3, 4, 11

[Gai23] Jason Gaitonde. Are there a few input bits that randomize the output of an $\mathbb{F}_2$ polynomial? MathOverflow, 2023. URL:https://mathoverflow.net/q/460879 (version: 2023-12-22). 31

[GGH$^+$07] Shafi Goldwasser, Dan Gutfreund, Alexander Healy, Tali Kaufman, and Guy N Rothblum. Verifying and decoding in constant depth. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 440–449, 2007. 1, 2

[GGNS23] Karthik Gajulapalli, Alexander Golovnev, Satyajeet Nagargoje, and Sidhant Saraogi. Range avoidance for constant-depth circuits: Hardness and algorithms. *arXiv preprint arXiv:2303.05044*, 2023. 1

[GLW22] Venkatesan Guruswami, Xin Lyu, and Xiuhan Wang. Range avoidance for low-depth circuits and connections to pseudorandomness. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2022. 1

[GW20] Mika Göös and Thomas Watson. A lower bound for sampling disjoint sets. *ACM Transactions on Computation Theory (TOCT)*, 12(3):1–13, 2020. 1

[Hag91] 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. 3

[Hås86] Johan Håstad. *Computational limitations for small depth circuits.* PhD thesis, Massachusetts Institute of Technology, 1986. 2

[JVV86] 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. 1

[Kan17] Daniel M Kane. A structure theorem for poorly anticoncentrated polynomials of Gaussians and applications to the study of polynomial threshold functions. 2017. 23

[KKL17] Valentine Kabanets, Daniel M Kane, and Zhenjian Lu. A polynomial restriction lemma with applications. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 615–628, 2017. Available at `https://cseweb.ucsd.edu/~dakane/PTFblockRestriction.pdf`. 24

[KLMS16] Andreas Krebs, Nutan Limaye, Meena Mahajan, and Karteek Sreenivasaiah. Small depth proof systems. *ACM Transactions on Computation Theory (TOCT)*, 9(1):1–26, 2016. 1

[KOW24] 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. Available at `https://arxiv.org/abs/2402.14278`. 1, 2, 3, 4, 7, 11, 28, 29

[Lug17] Michael Lugo. Sum of the first k binomial coefficients for fixed n. MathOverflow, 2017. URL:https://mathoverflow.net/q/17236 (version: 2017-10-01). 8

[LV11] 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. 1

[MV91] Yossi Matias and Uzi Vishkin. Converting high probability into nearly-constant time—with applications to parallel hashing. In *Proceedings of the twenty-third annual ACM symposium on Theory of Computing*, pages 307–316, 1991. 3

[RSW22] Hanlin Ren, Rahul Santhanam, and Zhikun Wang. On the range avoidance problem for circuits. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 640–650. IEEE, 2022. 1

[SS24] 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. 1

[Ush86] Nikolai G Ushakov. Upper estimates of maximum probability for sums of independent random vectors. *Theory of Probability & Its Applications*, 30(1):38–49, 1986. 4

[Vio12a] Emanuele Viola. Bit-probe lower bounds for succinct data structures. *SIAM Journal on Computing*, 41(6):1593, 2012. 1

[Vio12b] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012. 1, 2, 3

[Vio12c] Emanuele Viola. Extractors for Turing-machine sources. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 663–671. Springer, 2012. 1

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014. 1

[Vio16] Emanuele Viola. Quadratic maps are hard to sample. *ACM Transactions on Computation Theory (TOCT)*, 8(4):1–4, 2016. 1

[Vio20] Emanuele Viola. Sampling lower bounds: boolean average-case and permutations. *SIAM Journal on Computing*, 49(1):119–137, 2020. 1, 3, 7

[Vio23] Emanuele Viola. New sampling lower bounds via the separator. In *38th Computational Complexity Conference (CCC 2023)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. Available at https://eccc.weizmann.ac.il/report/2021/073/, 2023. 1, 2, 3

[Wik23a] Wikipedia. Binary entropy function — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Binary%20entropy%20function&oldid=1071507954, 2023. [Online; accessed 04-December-2023]. 8

[Wik23b] Wikipedia. Binomial coefficient — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Binomial%20coefficient&oldid=1187835533, 2023. [Online; accessed 15-December-2023]. 8

[Wik23c] Wikipedia. Bézout's identity — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=B%C3%A9zout's%20identity&oldid=1179305736, 2023. [Online; accessed 05-December-2023]. 39

[WKST19] Adam Bene Watts, Robin Kothari, Luke Schaeffer, and Avishay Tal. Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 515–526, 2019. 1

[WP23] Adam Bene Watts and Natalie Parham. Unconditional quantum advantage for sampling with shallow circuits. *arXiv preprint arXiv:2301.00995*, 2023. 1, 2

A Density Comparison of Sum of Integral Random Variables

The goal of this section is to prove the following density comparison result for sums of integral random variables.

Theorem A.1. *Let $t \ge 1$ be an integer, and let $X_1, \ldots, X_n$ be independent random variables in $\{0, 1, \ldots, t\}$. Let $\Phi \subseteq \{2, 3, \ldots, t\}$. Define ϕ as the least common multiple of values in $[t] \setminus \Phi$.*

For each $i \in [n]$ and integer $r \ge 1$, define $p_{r,i} = \max_{x \in \mathbb{Z}} \mathbf{Pr}\left[X_i \equiv x \pmod r\right]$ and assume[6]

$$\sum_{i \in [n]} (1 - p_{r,i}) \ge L > 0 \quad \textit{holds for all } r \in \Phi. \tag{24}$$

Let $\alpha = \left(\frac{L}{4n(t+1)}\right)^{t^2\phi}$ and assume $m := \lfloor L/(16t^4\phi) \rfloor \ge 1$. Then for any $x \in \mathbb{Z}$, we have

$$\mathbf{Pr}\left[\sum_{i \in [n]} X_i = x\right] - \mathbf{Pr}\left[\sum_{i \in [n]} X_i = x + \Delta\right] \le \frac{22|\Delta|}{\phi \cdot \alpha m}$$

holds for any $\Delta \in \mathbb{Z}$ that is a multiple of ϕ.

The typical setting for Theorem A.1 is when we have small t and $L = \Theta_t(n)$; then α is also a constant depending only on t.

Remark A.2. The assumption of Δ being a multiple of ϕ is necessary. If the X_i's have some joint congruence relation not shared with Δ, the bound can fail. Consider the case where n is even and each X_i is uniform in $\{1, 3\}$, which violates (24) only for $r = 2$. Then we set $x = 2n$ and $\Delta = 1$. Since the sum is n plus twice an n-bit binomial distribution, we have $\mathbf{Pr}\left[\sum_i X_i = x\right] \approx 1/\sqrt{n}$ but $\mathbf{Pr}\left[\sum_i X_i = x + \Delta\right] = 0$. However, $m \approx n$ and α is a constant. Hence the final bound does not hold.

We also note that the quantitative bound of α and m can be slightly improved by tightening our analysis. Since it does not change our final bounds by much, we choose the cleaner presentation here.

We will need the following simple bound on the difference of nearby binomial coefficients.

[6] Note that if (24) holds for some r, then it also holds for r' that is a multiple of r as $\mathbf{Pr}\left[X_i \equiv x \pmod{r'}\right] \le \mathbf{Pr}\left[X_i \equiv x \pmod r\right]$. Hence we may assume that Φ contains all the multiples of r (up to t) if $r \in \Phi$.

Fact A.3. *For any integers $n, b \geq 1$, we have $2^{-n} \cdot \left(\binom{n}{b} - \binom{n}{b+1}\right) \leq \frac{7}{n}$. Moreover for any integer $\Delta \geq 0$, we have*

$$2^{-n} \cdot \left(\binom{n}{b} - \binom{n}{b+\Delta}\right) \leq \frac{7\Delta}{n}.$$

Proof. The moreover part follows from a telescoping sum. Hence we focus on the first bound and divide into the following cases:

- If $b < n/2$, then $\binom{n}{b} \leq \binom{n}{b+1}$ and the bound trivially holds.
- If $b \geq n$, then $\binom{n}{b} \leq 1$, $\binom{n}{b+1} = 0$, and the bound trivially holds.
- If $n/2 \leq b \leq n-1$, then

$$\begin{aligned} \binom{n}{b} - \binom{n}{b+1} &= \binom{n}{b} \cdot \left(1 - \frac{n-b}{b+1}\right) = \binom{n}{b} \cdot \frac{2b+1-n}{b+1} \\ &\leq \frac{2^{n\cdot\mathcal{H}(b/n)}}{\sqrt{\pi b(1-b/n)}} \cdot \frac{2b+1-n}{b+1} && \text{(by Fact 3.7)} \\ &\leq \frac{2^{n\cdot\mathcal{H}(b/n)}}{\sqrt{n/2}} \cdot \frac{2b+1-n}{n/2}. && \text{(since } n/2 \leq b \leq n-1\text{)} \end{aligned}$$

Define $x = \frac{2b}{n} - 1$. Then $x \in [0, 1)$. By Fact 3.8, we have $\mathcal{H}(b/n) - 1 = \mathcal{H}\left(\frac{1+x}{2}\right) - 1 \leq -\frac{x^2}{2\ln(2)}$ and hence

$$\begin{aligned} 2^{-n} \cdot \left(\binom{n}{b} - \binom{n}{b+1}\right) &\leq 2^{-n\cdot\frac{x^2}{2\ln(2)}} \cdot \frac{1}{\sqrt{n/2}} \cdot \frac{n\cdot x + 1}{n/2} \\ &\leq \underbrace{x \cdot 2^{-n\cdot\frac{x^2}{2\ln(2)}}}_{A} \cdot \frac{2\sqrt{2}}{\sqrt{n}} + \left(\frac{2}{n}\right)^{1.5}. \end{aligned}$$

Writing $x = \sqrt{\frac{2\ln(2)\cdot\log(y)}{n}}$ for some $y \geq 1$, we transform A above into

$$A = \sqrt{\frac{2\ln(2)\cdot\log(y)}{n}} \cdot \frac{1}{y} \leq \sqrt{\frac{2\ln(2)}{n}}.$$

Hence $2^{-n} \cdot \left(\binom{n}{b} - \binom{n}{b+1}\right) \leq \frac{4\sqrt{\ln(2)}}{n} + \left(\frac{2}{n}\right)^{1.5} \leq \frac{7}{n}$ as claimed. □

To prove Theorem A.1, we observe that intuitively $\sum_{i\in[n]} X_i$ should converge to a (discrete) Gaussian distribution with large variance. Then in this (discrete) Gaussian distribution,

- if x lies much outside the standard deviation regime around the mean, then it has small density already;
- otherwise its density, compared with the density of $x + \Delta$, is only off by a small multiplicative factor, which means the quantity of interest is in fact small.

We first prove a simpler case where each random variable always has a "neighboring" pair of values in its support. Note that in this case we do not need to assume that the random variables are bounded. Later we will reduce the case of Theorem A.1 to this setting.

Lemma A.4. *Let $Y_1, \ldots, Y_m$ be independent integer random variables, and let $\phi \geq 1$ be an integer. Assume that $\alpha > 0$ is a parameter such that for each $i \in [m]$, there exists $u_i \in \mathbb{Z}$ satisfying*

$$\mathbf{Pr}[Y_i = u_i] \geq \alpha \quad \textit{and} \quad \mathbf{Pr}[Y_i = u_i + \phi] \geq \alpha.$$

Then for any $y \in \mathbb{Z}$, we have

$$\mathbf{Pr}\left[\sum_{i\in[m]} Y_i = y\right] - \mathbf{Pr}\left[\sum_{i\in[m]} Y_i = y + \Delta\right] \leq \frac{22|\Delta|}{\phi \cdot \alpha m} \tag{25}$$

holds for any $\Delta \in \mathbb{Z}$ that is a multiple of ϕ.

Proof. The bound trivially holds if $\Delta = 0$. If $\Delta < 0$, then we work with negated Y_i's. Hence we assume $\Delta \geq \phi$. By subtracting u_i from Y_i and y, we assume that each u_i equals zero. Then we decompose each $Y_i = W_i \cdot B_i + (1 - W_i) \cdot Z_i$, where B_i is uniform over $\{0, \phi\}$, W_i be an α-biased coin (i.e., $\mathbf{Pr}[W_i = 1] = \alpha$ and $\mathbf{Pr}[W_i = 0] = 1 - \alpha$), and Z_i is some integer random variable. In addition, W_i, B_i, Z_i are independent.

Now define $\mathcal{E}$ to be the event that $\sum_{i\in[m]} W_i \leq \alpha \cdot m/2$. Then by Fact 3.5 with $\delta = 1/2$ and $\mu = \alpha \cdot m$, we have

$$\mathbf{Pr}[\mathcal{E}] \leq e^{-\alpha \cdot m/8}. \tag{26}$$

For fixed $W = (W_1, \ldots, W_m)$ under which $\mathcal{E}$ does not happen, let $S = \{i \in [m] : W_i = 1\}$ of size $k = |S| \geq \alpha \cdot m/2$. Then for any fixed $Z = (Z_1, \ldots, Z_m)$, the LHS of (25) equals

$$\mathbf{Pr}\left[\sum_{i\in S} B_i = b \,\middle|\, W, Z, \neg\mathcal{E}\right] - \mathbf{Pr}\left[\sum_{i\in S} B_i = b + \Delta \,\middle|\, W, Z, \neg\mathcal{E}\right],$$

where $b = y - \sum_{i\notin S} Z_i$. Recall that each B_i is uniform over $\{0, \phi\}$. If b is not a multiple of ϕ, then the above quantity equals zero since Δ is a multiple of ϕ. Otherwise, let $b' = b/\phi$ and $\Delta' = \Delta/\phi \geq 1$. Then the above quantity equals

$$2^{-k} \cdot \left(\binom{k}{b'} - \binom{k}{b' + \Delta'}\right) \leq \frac{7\Delta'}{k} \tag{27}$$

by Fact A.3. This, combined with (26), establishes (25):

$$\begin{aligned}
\text{LHS of (25)} &\leq \mathbf{Pr}[\mathcal{E}] + \mathbb{E}\left[\mathbb{1}_{\phi \text{ divides } b} \cdot 2^{-k} \cdot \left(\binom{k}{b'} - \binom{k}{b' + \Delta'}\right) \,\middle|\, \neg\mathcal{E}\right] \\
&\leq e^{-\alpha \cdot m/8} + \mathbb{E}\left[\frac{7\Delta'}{k} \,\middle|\, \neg\mathcal{E}\right] && \text{(by (26) and (27))} \\
&\leq \frac{8}{\alpha \cdot m} + \frac{14\Delta'}{\alpha \cdot m} && \text{(since } e^{-x} \leq \tfrac{1}{x} \text{ and } k \geq \alpha m/2\text{)} \\
&\leq \frac{22\Delta}{\phi \cdot \alpha m} && \text{(since } \Delta' = \Delta/\phi \geq 1\text{)} \\
&= \text{RHS of (25)}.
\end{aligned}$$

□

Now we prove Theorem A.1 by reducing it to Lemma A.4. To this end, we will divide $X_1, \ldots, X_n$ into many parts, and the sum within each part will have two neighboring values with noticeable probability weights.

Proof of Theorem A.1. Denote $\Phi = \{r_1, r_2, \ldots, r_k\}$ where $k = |\Phi| \le t$. For each r_j, let $S_{r_j} \subseteq [n]$ be the set of X_i's with $1 - p_{r_j,i} \ge L/(2n)$, i.e.,

$$S_{r_j} = \left\{ i \in [n] : 1 - p_{r_j,i} \ge L/(2n) \right\}.$$

Since $0 \le 1 - p_{r_j,i} \le 1$ and by (24), we have

$$L \le \sum_{i \in [n]} (1 - p_{r_j,i}) \le \left|S_{r_j}\right| \cdot 1 + \left(n - \left|S_{r_j}\right|\right) \cdot \frac{L}{2n},$$

which implies $\left|S_{r_j}\right| \ge L/2$. Now we remove multiple appearances of indices across S_{r_j}'s to make them pairwise disjoint. Formally, for each $j = 1, 2, \ldots, k$, we keep $n' := \lfloor L/(4t) \rfloor$ elements in S_{r_j} and update $S_{r_{j'}} \leftarrow S_{r_{j'}} \setminus S_{r_j}$ for all $j' > j$. Since $k \le t$ and originally $\left|S_{r_j}\right| \ge L/2$, each S_{r_j} contains at least n' elements after this pruning.

For each $j \in [k]$ and $i \in S_{r_j}$, by an averaging argument, there exists $c_i \in \mathbb{Z}/r_j\mathbb{Z}$ such that $\mathbf{Pr}\left[X_i \equiv c_i \pmod{r_j}\right] \ge \frac{1}{r_j}$. Hence, by another averaging argument, there exists $z_i \in \{0, 1, \ldots, t\}$ such that $z_i \equiv c_i \pmod{r_j}$ and

$$\mathbf{Pr}\left[X_i = z_i\right] \ge \frac{1}{r_j} \cdot \frac{1}{\lceil (t+1)/r_j \rceil} \ge \frac{1}{2(t+1)} \ge \frac{L}{4n(t+1)},$$

where we used the fact that $0 < L \le n$. Since $i \in S_{r_j}$, we also have $\mathbf{Pr}\left[X_i \equiv c_i \pmod{r_j}\right] \le 1 - \frac{L}{2n}$ and hence, by an averaging argument, there exists $c'_i \in \mathbb{Z}/r_j\mathbb{Z}$ such that $c'_i \ne c_i$ and $\mathbf{Pr}\left[X_i \equiv c'_i \pmod{r_j}\right] \ge \frac{L}{2n \cdot (r_j - 1)}$. Similarly by another averaging argument, there exists $z'_i \in \{0, 1, \ldots, t\}$ such that $z'_i \equiv c'_i \pmod{r_j}$ and

$$\mathbf{Pr}\left[X_i = z'_i\right] \ge \frac{L}{2n \cdot (r_j - 1)} \cdot \frac{1}{\lceil (t+1)/r_j \rceil} \ge \frac{L}{4n(t+1)}.$$

Since both z_i and z'_i are in $\{0, 1, \ldots, t\}$, by a final averaging argument, there exists z_{r_j}, z'_{r_j} such that

1. $z_{r_j}, z'_{r_j} \in \{0, 1, \ldots, t\}$ and $z_{r_j} \not\equiv z'_{r_j} \pmod{r_j}$, and
2. at least a $1/\binom{t+1}{2} \ge 1/t^2$ fraction of $i \in S_{r_j}$ satisfy $\mathbf{Pr}\left[X_i = z_{r_j}\right], \mathbf{Pr}\left[X_i = z'_{r_j}\right] \ge \frac{L}{4n(t+1)}$.

Let $n'' = \lceil n'/t^2 \rceil = \lceil |S_{r_j}|/t^2 \rceil$. Based on Item 1 and Item 2, for each $j \in [k]$ we define $T_{r_j} \subseteq S_{r_j}$ to be of size n'' and contain indices satisfying Item 2.

Recall that ϕ is the least common multiple of values in $[t] \setminus \Phi$. Now we show that the sum of $t\phi \cdot k$ random variables ($t\phi$ from each one of $T_{r_1}, \ldots, T_{r_k}$) is a random variable that satisfies the conditions in Lemma A.4. Formally, let $m = \lfloor n''/(t\phi) \rfloor$ and arbitrarily select m disjoint subsets $T^1_{r_j}, \ldots, T^m_{r_j}$ of size $t\phi$ from each T_{r_j}. Define random variables

$$Y_\ell = \sum_{j \in [k]} \sum_{i \in T^\ell_{r_j}} X_i \quad \text{for each } \ell \in [m]$$

and define

$$Y_0 = \sum_{i \notin \bigcup_{j \in [k], \ell \in [m]} T^\ell_{r_j}} X_i$$

to be the sum of the remaining X_i's. We will show that for each $\ell \in [m]$, there exists $u_\ell \in \mathbb{Z}$ such that both $\mathbf{Pr}[Y_\ell = u_\ell]$ and $\mathbf{Pr}[Y_\ell = u_\ell + \phi]$ are at least $\alpha = \left(\frac{L}{4n(t+1)}\right)^{t^2\phi}$. Then Theorem A.1 follows from Lemma A.4 by conditioning on Y_0 and observing $m = \left\lfloor \left\lceil \lfloor L/(4t) \rfloor / t^2 \right\rceil / (t\phi) \right\rfloor \geq \lfloor L/(16t^4\phi) \rfloor$.

Fix an arbitrary $\ell \in [m]$ and define $w_j = z_{r_j} - z'_{r_j}$ for each $j \in [k]$. By Item 1, $|w_j| \leq t$ and r_j does not divide it. Hence the greatest common divisor g of $|w_1|, \ldots, |w_k|$ lies in $[t] \setminus \Phi$, which must divide ϕ. Thus by Bézout's identity (see e.g., [Wik23c]), there exist $s_1, \ldots, s_k \in \mathbb{Z}$ such that

$$\sum_{j \in [k]} s_j \cdot w_j = \phi. \tag{28}$$

In addition, we can assume that $|s_j| \leq \phi/g \cdot \max_{j\in[k]} |w_j|/g \leq t\phi$ [Bru12]. Now we define u_ℓ as

$$u_\ell = \sum_{j\in[k]: s_j < 0} z_{r_j} \cdot t\phi + \sum_{j \in [k]: s_j \geq 0} z'_{r_j} \cdot t\phi.$$

Then the probability of $Y_\ell = u_\ell$ is at least the probability that every $X_i \in T^\ell_{r_j}$ equals z_{r_j} if $s_j < 0$, and every $X_i \in T^\ell_{r_j}$ equals z'_{r_j} if $s_j \geq 0$. Hence by Item 2 and the independence of the X_i's, we have $\mathbf{Pr}\left[Y_\ell = u_\ell\right] \geq \left(\frac{L}{4n(t+1)}\right)^{t\phi\cdot k} \geq \alpha$ as desired. To analyze $u_\ell + \phi$, we rewrite it as

$$\begin{aligned} u_\ell + \phi &= \sum_{j\in[k]:s_j<0} \left(z_{r_j} \cdot t\phi + s_j \cdot w_j\right) + \sum_{j\in[k]:s_j\geq 0} \left(z'_{r_j} \cdot t\phi + s_j \cdot w_j\right) && \text{(by (28))} \\ &= \sum_{j\in[k]:s_j<0} \left(z'_{r_j} \cdot |s_j| + z_{r_j} \cdot (t\phi - |s_j|)\right) + \sum_{j\in[k]:s_j\geq 0} \left(z_{r_j} \cdot |s_j| + z'_{r_j} \cdot (t\phi - |s_j|)\right). \\ & && \text{(since } w_j = z_{r_j} - z'_{r_j}\text{)} \end{aligned}$$

Hence the probability of $Y_\ell = u_\ell + \phi$ is at least the probability that $|s_j|$ (resp., $t\phi - |s_j|$) many $X_i \in T^\ell_{r_j}$ equal z'_{r_j} (resp., z_{r_j}) if $s_j < 0$, and $|s_j|$ (resp., $t\phi - |s_j|$) many $X_i \in T^\ell_{r_j}$ equal z_{r_j} (resp., z'_{r_j}) if $s_j \geq 0$. Therefore $\mathbf{Pr}\left[Y_\ell = u_\ell + \phi\right] \geq \alpha$ follows again from Item 2 and the independence of X_i's. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il