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 also 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 example, NC^0 circuits can sample the uniform distribution over the preimage $\mathrm{PARITY}^{-1}(0)$

(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 $x_1, \ldots, x_n$ to $x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_n \oplus x_1$ [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 $3 \leq q \ll \sqrt{n}$ [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. *Let $f \colon \{0,1\}^m \to \{0,1\}^n$ be computable by an NC^0 circuit. If $f(\mathcal{U}^m)$ is ε-close (in total variation distance) to a uniform symmetric distribution and n is sufficiently large, 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] as follows.

Theorem 1.2 (Consequence of Theorem 4.1)**.** *Let d be a fixed constant, and suppose $f \colon \{0,1\}^m \to \{0,1\}^n$ is computable by an NC^0 circuit of depth at most d. If $f(\mathcal{U}^m)$ is ε-close (in total variation distance) to a uniform symmetric distribution and n is sufficiently large, then $f(\mathcal{U}^m)$ is $O(\varepsilon)$-close to one of the six distributions in Conjecture 1.1.*

Note that this result is optimal up to the implicit constant in $O(\varepsilon)$. 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 the appendices.

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]. View the inputs and outputs of f as the right and left vertices, respectively, of a bipartite graph. Following their terminology, we define the *neighborhood* of a left vertex v as the set of all left vertices adjacent to any right vertex that v is adjacent to. Furthermore, we call two neighborhoods N_1, N_2 *connected* if there exist left vertices $v_1 \in N_1, v_2 \in N_2$ and right vertex u such that both v_1 and v_2 are adjacent to u. By [KOW24, Corollary 4.11], there exists a relatively small set of right vertices 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 closely match the marginals of $\mathcal{D}_k$. Hence by conditioning on all the input bits that do not affect output bits in these neighborhoods, 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 $s = o(n)$ or $s = (1 - o(1))n$, we will show that it must be the case of $s \in \{0, n\}$. Otherwise $o(n) \ll s \ll (1 - o(1))n$, and we will show that it must be the case of $s \approx n/2$ and Ψ is effectively all-even, all-odd, or everything.

The $s = o(n)$ or $s = (1 - o(1))n$ case is essentially handled by [FLRS23]. For completeness, we give a simple (albeit quantitatively weaker) argument that suffices for our purposes. Our treatment (Theorem 4.3) is similar to [Vio20] and is presented in Appendix B.

Now we turn to the more interesting case where $o(n) \ll s \ll (1 - o(1))n$ and aim to show that $\mathcal{D}_\Psi$ is essentially uniform over even parities, odd parities, or everything. In this scenario, since s is relatively far from the boundaries, the framework explained in Subsection 2.1 is now applicable. However, we need a number of new ideas since Ψ is unstructured in general. For illustration and to introduce these new ingredients, we start with some concrete examples.

The Mixture of Opposite Hamming Slices. The first illustrating example is when Ψ contains opposite Hamming slices, both of which are significant. For concreteness, consider $\Psi = \{n/3, 2n/3\}$. To rule out this possibility, we follow the previous outline to reduce the local function f to more structured local functions where one has many small neighborhoods that are not connected. Assume after this graph elimination process, we obtained a local function g with many non-connected neighborhoods of small sizes. Since the marginal distribution of coordinates of $\mathcal{D}_\Psi$ is an unbiased coin, one may attempt to classify these neighborhoods by their distance to unbiased distributions as in Subsection 2.1; then use their independence to derive a $1 - e^{-\Omega_d(n)}$ bound when they are far from being unbiased, and use anticoncentration inequalities to show that the output of g rarely hits the desired Hamming weights when they are almost unbiased.

The second case indeed works. The first case, however, unfortunately fails, the reason for which is that g being far from unbiased does not necessarily mean that f cannot sample from $\mathcal{D}_\Psi$. For example, f could use one uniform bit to choose to sample from $\mathcal{D}_{n/3}$ or $\mathcal{D}_{2n/3}$, where g only needs to handle the former without the need to approximate $\mathcal{D}_\Psi$ itself. Though this example is artificial

4

since [Vio23, KOW24] have ruled out local sampling schemes for $\mathcal{D}_{n/3}$, the argument is oblivious to this and cannot work for more contrived examples that are similar in spirit to the above one.

To get around this, we observe that though the individual marginals of $\mathcal{D}_\Psi$ are unbiased, their joint distribution is far from *any* biased distribution (in fact, any product distribution, but focusing only on biased ones is sufficient for us). Hence we alter the criteria of our two cases for handling the behavior on marginals. We call a neighborhood Type-1 if it is not close to any biased distribution, and a neighborhood Type-2 if it is close to some biased distribution. (Note that these definitions are slightly different from those used in [KOW24].) At this point, if g has many small non-connected Type-1 neighborhoods, then it is $(1 - e^{-\Omega_d(n)})$-far from both $\mathcal{D}_{n/3}$ and $\mathcal{D}_{2n/3}$, which, by a union bound, means it is also $(1 - e^{-\Omega_d(n)})$-far from any mixture of $\mathcal{D}_{n/3}, \mathcal{D}_{2n/3}$ and, particularly, $\mathcal{D}_\Psi = \mathcal{D}_{\{n/3, 2n/3\}}$. On the other hand, we can similarly argue that the Hamming weight of a Type-2 neighborhood is a random variable with noticeable variance, since the Hamming weight of the biased distribution that it implicitly approximates has such a property. Therefore, if g has many small non-connected Type-2 neighborhoods, the previous argument using anticoncentration inequalities still works.

We remark that technically, the Hamming weight of a biased distribution has noticeable variance only if the bias itself is not extremely close to 0 or 1. Fortunately we can ignore those (Claim 4.10) by truncating the tail of Ψ. Since $o(n) \ll s \ll (1 - o(1))n$, this does not incur much error. In fact, this is the *only* place where we actually use the assumption that $o(n) \ll s \ll (1 - o(1))n$.

The Majority Distribution. The second interesting example is the majority case where $\mathcal{D}_\Psi$ is uniform over strings that have more ones than zeros, i.e., $\Psi = \{n/2 + 1, n/2 + 2, \ldots, n\}$. Note that this distribution is not close to any of the six special distributions, and thus, if we believe that the classification result is indeed true, it cannot be approximated by local distributions. Results are known on distinguishing this distribution with local functions [GGH+07, Vio12b, FLRS23]; however, their arguments do not seem to generalize.

To prove that any local function's output is far from $\mathcal{D}_\Psi$, we still follow the previous outline to obtain a d-local function g that has n/C_d non-connected neighborhoods of size C_d, where $C_d \gg 1$ is a large constant depending only on d from the graph elimination results. As before, if many of these neighborhoods are Type-1, then we readily obtain (Lemma 4.12) a $1 - e^{-\Omega_d(n)}$ distance bound by their independence. To handle the case where most of these neighborhoods are Type-2, we recall that past work [Vio12b, KOW24] used anticoncentration inequalities to show that the local distribution cannot hit the support of the desired distribution with as much mass as it is supposed to.

However this simply does not work, since Ψ covers a consecutively wide range of Hamming weights that significantly exceeds the amount of independence we could use in g. More precisely, g only has n/C_d independent small neighborhoods, hence their Hamming weight sum is essentially a discrete Gaussian of variance $\Omega_d(n)$. Due to the random shift from other neighborhoods that we cannot control, the support of this distribution can, in the worst case, completely lie in Ψ which has range $n/2$. Therefore, we cannot hope to establish a similar result saying with high probability the output of g does not hit the support of $\mathcal{D}_\Psi$.

To circumvent this, we explore the sharp threshold phenomenon in Ψ. Observe that the range of Ψ, though wide, has a sharp cutoff at Hamming weight $n/2$. This should not happen in a distribution that behaves like a Gaussian, unless the cutoff appears at the tail. In other words, we expect to see that in the probability density function of the Hamming weight of the output of g, the mass on the right boundary of $n/2$ does not differ from the mass on the left boundary of $n/2$,

in contrast to $\mathcal{D}_\Psi$. To be more precise, we consider the potential function

$$\phi = \mathbf{Pr}\left[|X| \in [n/2 + 1, n/2 + \sqrt{n}]\right] - C_d' \cdot \mathbf{Pr}\left[|X| \in [n/2 - \sqrt{n}, n/2]\right],$$

where C_d' is another large constant depending on d. Let ϕ_1 be ϕ when $X \sim \mathcal{D}_\Psi$ and ϕ_2 be ϕ when X is the output of g. Then obviously ϕ_1 is roughly 1. However ϕ_2 should be $o_d(1)$: since g has many small non-connected Type-2 neighborhoods, the Hamming weight distribution of g can be shown to be a Gaussian-like distribution with variance n/C_d. If the mean of the distribution is $\ll n/2 + \Theta_d(\sqrt{n})$, then ϕ_2 is negative since C_d' is sufficiently large and the two intervals in the definition are consecutive. On the other hand if the mean is $\gg n/2 + \Theta_d(\sqrt{n})$, then ϕ_2 is $o_d(1)$ by standard concentration for the left-hand side of ϕ_2. In summary, this shows that $\phi_1 - \phi_2$ is always at least $1 - o_d(1)$, which implies a $1/C_d'$ distance bound since ϕ has range bounded by C_d'. Later it is put together with the Type-1 analysis by a union bound.

A Unified Construction of the Potential Function. Let $\mathcal{P}_\Psi$ be the distribution over Hamming weights of strings drawn from $\mathcal{D}_\Psi$. To extend the analysis of the above majority distribution to a general Ψ, the key point is to discover the threshold phenomenon in Ψ and to construct the corresponding potential function to separate $\mathcal{P}_\Psi$ from a discrete Gaussian $\mathcal{P}$ with variance $\Theta(n)$ and an unknown mean. To this end, we follow the same approach to select some Hamming weights $S \subseteq \Psi$ and $T \subseteq \{0, 1, \ldots, n\} \setminus \Psi$, then define

$$\phi = \mathbf{Pr}[|X| \in S] - \Theta(1) \cdot \mathbf{Pr}[|X| \in T].$$

To ensure ϕ is large under $\mathcal{P}_\Psi$, S needs to take a significant amount of mass in $\mathcal{P}_\Psi$. To ensure ϕ is small under $\mathcal{P}$, S and T should not be too far apart (Lemma 3.9), i.e., elements in S should be paired with distinct elements in T such that the distance of each pair is not too large. Guided by this intuition, we construct S and T in a greedy way: we iteratively pick out $s \in \Psi, t \notin \Psi$ to update S, T, where the selected pair needs to minimize the distance $|s - t|$. If we select K pairs in total, then the absolute distance between each pair is $O(K)$ by a simple induction argument and the minimality of the distance of each pair upon selection. Therefore, we only need to make sure K is small while ϕ is large, say at least δ, under $\mathcal{P}_\Psi$.

For this purpose, we truncate the δ-tail and only look for (s, t) pairs thereinto. Formally, let $0 \leq \ell \leq r \leq n$ be such that Hamming weights in $[\ell, r]$ cover $1 - \delta$ fraction of 2^n. Then we iteratively pick out $s \in \Psi \cap [\ell, r]$ and $t \in [\ell, r] \setminus \Psi$ while minimizing their distance $|s - t|$ until $\phi \gg \delta$ under $\mathcal{P}_\Psi$. Similarly, if we selected K pairs, the distance of each pair is at most $O(K)$. In addition, since we have removed the δ-tail, we can show (Claim 4.17) that every Hamming slice in $\Psi \cap [\ell, r]$ has probability mass $\gg \frac{\delta\sqrt{\log(1/\delta)}}{\sqrt{n}}$ under the uniform distribution (and thus under $\mathcal{P}_\Psi$ by its definition). Therefore $K \ll \sqrt{n/\log(1/\delta)}$. Crucially, this additional log factor helps us in upper bounding ϕ under the discrete Gaussian $\mathcal{P}$. Let $\Delta = s - n/2$. Then

$$\mathcal{P}(s) \approx 2^{-\Delta^2/n}/\sqrt{n} \quad \text{and} \quad \mathcal{P}(t) \approx \mathcal{P}(s + K) \approx 2^{-(\Delta+K)^2/n}/\sqrt{n}.$$

Thus

- if $\Delta \gg \sqrt{n\log(1/\delta)}$, then $\mathcal{P}(s) - C_d' \cdot \mathcal{P}(t) \leq \mathcal{P}(s) \ll \delta/\sqrt{n}$,
- otherwise $\Delta \ll \sqrt{n\log(1/\delta)}$ and $\mathcal{P}(s) - \Theta(1) \cdot \mathcal{P}(t) \approx \left(2^{-\Delta^2/n} - \Theta(1) \cdot 2^{-(\Delta+K)^2/n}\right)/\sqrt{n} < 0$,

which implies that $\phi \ll K \cdot \delta/\sqrt{n} \ll \delta$ under $\mathcal{P}$.

6

One caveat here is that the actual Hamming weight distribution of the sum of small non-connected Type-2 neighborhoods may embed a periodic pattern. For example, $\mathcal{P}$ may be a discrete Gaussian on even numbers, which is exactly the case in the locally sampleable distribution $\mathcal{D}_{\{0,2,4,\dots\}}$. Luckily, by the argument for ruling out periodic Hamming slices in [KOW24], periods other than two are also ruled out for $\mathcal{P}$ (Lemma 4.13). Therefore, we only need to additionally make sure the selected pairs (s,t) have an even distance (Lemma 4.16), since otherwise we may have $\mathcal{P}(s) \approx 1/\sqrt{n}$ but $\mathcal{P}(t) = 0$ when s and t have different parities.

Extremely Small Error Regime. Combining the above arguments, we can prove a weaker classification result: if the local distribution $\mathcal{P}$ is ε-close to $\mathcal{D}_\Psi$ where there exists some $s \in \Psi$ such that $o(n) \ll s \ll (1-o(1))n$, then it is $\left(O_d(\varepsilon) + e^{-\Theta_d(n)}\right)$-close to the uniform distribution or the uniform distribution over even / odd Hamming weights. The additional exponentially small factor comes from analyzing Type-2 neighborhoods (Lemma 4.13), where an additive $e^{-\Theta_d(n)}$ is used to rule out the atypical case that the Hamming weight distribution is not Gaussian-like due to correlations outside those non-connected Type-2 neighborhoods. While this seems inevitable in our current analysis framework, this factor is minor when ε is not extremely small. To further shave it, we explore the structure of d-local functions from different angles.

Let $\varepsilon \ll e^{-\Theta_d(n)}$. Assume our distribution $\mathcal{P}$ is ε-close to $\mathcal{D}_\Psi$ and is $\left(O_d(\varepsilon) + e^{-\Theta_d(n)}\right)$-close to the uniform distribution. Let $\delta \ll O_d(\varepsilon) + e^{-\Theta_d(n)}$ be the distance between $\mathcal{D}_\Psi$ and the uniform distribution. Then we will use the variance V of the Hamming weight of $\mathcal{D}_\Psi$ to show (Theorem 4.9) that δ is in fact $O_d(\varepsilon)$. On the one hand, since $\mathcal{P}$ is extremely close to the uniform, its (pairwise) marginals should be perfectly unbiased by granularity (Lemma 4.21). Hence $V \geq n/4 - n^2 \cdot \varepsilon$ where $n/4$ is the variance of the Hamming weight of $\mathcal{P}$. On the other hand, since $\mathcal{D}_\Psi$ is already $e^{-\Theta_d(n)}$-close to the uniform distribution, it shall not miss central Hamming weights. In fact, every $m \notin \Psi$ must be $\gg \Theta_d(n)$. Therefore we can upper bound $V \leq \left(n/4 - \Theta_d(n^2) \cdot \delta\right)/(1-\delta)$ by comparing the variance calculation of $\mathcal{D}_\Psi$ and the uniform distribution. Then $\delta \leq O_d(\varepsilon)$ follows from rearranging.

Now we turn to the case that our distribution is $\left(O_d(\varepsilon) + e^{-\Theta_d(n)}\right)$-close to the uniform distribution over strings of even (or odd) Hamming weights. Then the above argument still works if Ψ contains only even numbers. Otherwise Ψ may include some odd Hamming weights and thus make the upper bound on V larger than $n/4$. To amend this, we show (Lemma 4.22) that $\mathcal{P}$ supports on strings with even Hamming weights and thus we can discard odd weights in Ψ without increasing the distance ε. Then the previous argument carries over. To this end, we notice that the Hamming weight of $\mathcal{P}$ has degree d over $\mathbb{F}_2$ because f is d-local. Hence by granularity, the expectation of the Hamming weight parity is an integer multiple of 2^{-d}, which has to be zero as the distance $\varepsilon \ll 2^{-d}$.

Putting Everything Together. Combining the above new ideas, we now give a streamlined proof outline for proving the classification result in the regime where there exists some $s \in \Psi$ such that $o(n) \ll s \ll (1-o(1))n$. Assume the output of a d-local function f is close to $\mathcal{D}_\Psi$.

- First we apply the graph elimination results (Proposition 4.6) to reduce the d-local function f to a number of more structured functions g, where each g is d-local and has $\Omega_d(n)$ non-connected neighborhoods of size $O_d(1)$.

- Then we classify each such neighborhood as Type-1 or Type-2, depending on whether its output distribution is far from any biased distribution.

  - If g has $\Omega_d(n)$ Type-1 non-connected small neighborhoods, then we show (Lemma 4.12) that the output of g is extremely far from $\mathcal{D}_\Psi$ by their independence.

7

– Otherwise we have $\Omega_d(n)$ Type-2 non-connected small neighborhoods. Unless $\mathcal{D}_\Psi$ is close to the uniform distribution, uniform over even strings, or uniform over odd strings, we can construct (Lemma 4.16) a potential function ϕ such that its expectation is large under the Hamming weight distribution of $\mathcal{D}_\Psi$ but is small under any Gaussian-like distribution on even or on odd numbers. Then we show (Lemma 4.13) that the Hamming weight distribution of g's output is typically a sum of independent integral random variables that mimics a discrete Gaussian over odd or even numbers, which means ϕ is indeed small under this distribution.

• Combining the large distance bound from the Type-1 case and the noticeable deviation of the expected value of ϕ from the Type-2 case, we show (Theorem 4.8) that the output of f is relatively far from $\mathcal{D}_\Psi$. Since this contradicts the starting assumption, we must have that $\mathcal{D}_\Psi$ is close to one of the three special distributions, implying that the output of f is also close to the very distribution.

• Finally we use additional treatment (Theorem 4.9) sketched above to further sharpen the bound.

3 Preliminaries

For a positive integer n, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. We use $\mathbb{R}$ to denote the set of real numbers, use $\mathbb{N} = \{0, 1, 2, \ldots\}$ 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 $e \approx 2.71828\ldots$ respectively. For $a \in \mathbb{N}$, we use $\mathrm{tow}(a)$ to denote the power tower of base 2 and order a, where

$$\mathrm{tow}(a) = \begin{cases} 1 & a = 0, \\ 2^{\mathrm{tow}(a-1)} & a \geq 1. \end{cases}$$

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.

Probability. We reserve $\mathcal{U}$ to denote the uniform distribution over $\{0, 1\}$, and more generally for $\gamma \in [0, 1]$, reserve $\mathcal{U}_\gamma$ to denote the γ-biased distribution, i.e., $\mathcal{U}_\gamma(1) = \gamma = 1 - \mathcal{U}_\gamma(0)$. Note that $\mathcal{U} = \mathcal{U}_{1/2}$.

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, for any non-empty set $S \subseteq [n]$, we use $\mathcal{P}|_S$ to denote the marginal distribution of $\mathcal{P}$ on coordinates in 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 $\mathcal{P}$.

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.[1] We say $\mathcal{P}$ is ε-close to $\mathcal{Q}$ if $\|\mathcal{P}(x) - \mathcal{Q}(x)\|_{\mathsf{TV}} \leq \varepsilon$, and ε-far otherwise.

---

[1] 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.

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

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

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 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 $\mathcal{P} = \sum_{i \in [t]} \alpha_i \cdot \mathcal{P}_i$.

We will use several inequalities from [KOW24] about total variation distance. The first shows that distance is largely preserved under conditioning.

Fact 3.2 (See e.g., [KOW24, Fact 4.1])**.** *Assume $\mathcal{P}$ is ε-close to $\mathcal{Q}$, and let $\mathcal{P}', \mathcal{Q}'$ be the distributions of $\mathcal{P}, \mathcal{Q}$ conditioned on some event $\mathcal{E}$, respectively. Then for any function f,*

$$\left\|f(\mathcal{P}') - f(\mathcal{Q}')\right\|_{\mathsf{TV}} \leq \frac{2\varepsilon}{\mathcal{Q}(\mathcal{E})}.$$

The second allows us to argue 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 ([KOW24, Lemma 4.3]; see also [Vio20, Section 4.1])**.** *Let $\mathcal{P}_1, \ldots, \mathcal{P}_t$ and $\mathcal{Q}$ be distributions. Assume there exists an event $\mathcal{E}$ and values $\varepsilon_1, \varepsilon_2$ such that for each $i \in [t]$,*

- *either $\|\mathcal{P}_i - \mathcal{Q}\|_{\mathsf{TV}} \geq 1 - \varepsilon_1$ holds,*
- *or $\mathcal{P}_i(\mathcal{E}) \leq \varepsilon_2$ and $\mathcal{Q}(\mathcal{E}) \geq 1 - \varepsilon_3$ hold.*

Then for any distribution $\mathcal{P}$ as a convex combination of $\mathcal{P}_1, \ldots, \mathcal{P}_t$, we have

$$\|\mathcal{P} - \mathcal{Q}\|_{\mathsf{TV}} \geq 1 - (t+1) \cdot \varepsilon_1 - \varepsilon_2 - \varepsilon_3.$$

Finally, we will require the following lemma. It shows that two coupled random vectors with identical marginal distributions will still have Hamming weight mismatch (even modulo an integer) as long as parts of their entries are independent.

Lemma 3.4 ([KOW24, Lemma 4.4])**.** *Let (X, Y, Z, W) be a random variable where $X, Z \in \{0, 1\}$ and $Y, W \in \{0, 1\}^{t-1}$. Let $q \geq \min\{3, t+1\}$ be an integer.[2] Assume*

- *X is independent from (Z, W) and Z is independent from (X, Y),*
- *(X, Y) and (Z, W) have the same marginal distribution and is ε-close to $\mathcal{U}_\gamma^t$ for some $\gamma \in (0, 1/2]$[3] and*

$$\varepsilon \leq \frac{\gamma}{4q} \cdot 2^{-50\gamma(t-1)/q^2}.$$

Then we have

$$\mathbf{Pr}\left[X + |Y| \equiv Z + |W| \pmod{q}\right] \leq 1 - \frac{\gamma}{2q} \cdot 2^{-50\gamma(t-1)/q^2}.$$

We also follow much of the terminology and notation from [KOW24] below.

---

[2]If $q \geq t + 1$, then one may instead apply Lemma 3.4 with modulus $t + 1$, since $X + |Y| \equiv Z + |W| \pmod{q}$ is equivalent to $X + |Y| = Z + |W|$ for $q \geq t + 1$.

[3]Lemma 3.4 holds for $\gamma \in [1/2, 1)$ as well, with γ replaced by $1 - \gamma$ in the bounds. This can be achieved by simply flipping zeros and ones of (X, Y, Z, W). This trick carries over the ε-closeness to $\mathcal{U}_{1-\gamma}^t$ and preserves the congruence.

Locality. Let $f\colon \{0,1\}^m \to \{0,1\}^n$. For each output bit $i \in [n]$, we use $I_f(i) \subseteq [m]$ to denote the set of input bits that the i-th output bit depends on. We say f is a d-local function if $|I_f(i)| \leq d$ holds for all $i \in [n]$. Define $N_f(i) = \{i' \in [n]\colon I_f(i) \cap I_f(i') \neq \emptyset\}$ to be the neighborhood of i, which contains all the output bits that have potential correlation with the i-th output bit. For each input bit $j \in [m]$, we use $\deg_f(j) = |\{i \in [n]\colon j \in I_f(i)\}|$ to denote the number of output bits that it influences.

We say output bit i_1 is connected to i_2 if $I_f(i_1) \cap I_f(i_2) \neq \emptyset$. We say neighborhood $N_f(i_1)$ is connected to $N_f(i_2)$ if there exist $i'_1 \in N_f(i_1)$ and $i'_2 \in N_f(i_2)$ such that $I_f(i'_1) \cap I_f(i'_2) \neq \emptyset$. As such, every output bit is independent of any non-connected output bit, and the output of a neighborhood has no correlation with any non-connected neighborhood of it. When f is clear from the context, we will drop subscripts in $I_f(i), N_f(i), \deg_f(j)$ and simply use $I(i), N(i), \deg(j)$.

Bipartite Graphs. We sometimes take an alternative view, using bipartite graphs to model the dependency relations in f. Let $G = (V_1, V_2, E)$ be an undirected bipartite graph. For each $i \in V_1$, we use $I_G(i) \subseteq V_2$ to denote the set of adjacent vertices in V_2. We say G is d-left-bounded if $|I_G(i)| \leq d$ holds for all $i \in V_1$. Define $N_G(i) = \{i' \in V_1\colon I_G(i) \cap I_G(i') \neq \emptyset\}$ to be the left neighborhood of i.

We say left vertex i_1 is connected to i_2 if $I_G(i_1) \cap I_G(i_2) \neq \emptyset$. We say left neighborhood $N_G(i_1)$ is connected to $N_G(i_2)$ if there exist $i'_1 \in N_G(i_1)$ and $i'_2 \in N_G(i_2)$ such that $I_G(i'_1) \cap I_G(i'_2) \neq \emptyset$. For each $j \in V_2$, we use $\deg_G(j) = |\{i \in V_1\colon j \in I_G(i)\}|$ to denote its degree. When G is clear from the context, we will drop subscripts in $I_G(i), N_G(i), \deg_G(j)$ and simply use $I(i), N(i), \deg(j)$.

It is easy to see that the dependency relation in $f\colon \{0,1\}^m \to \{0,1\}^n$ can be visualized as a bipartite graph $G = G_f$ where $[n]$ is the left vertices (representing output bits of f) and $[m]$ is the right vertices (representing input bits of f), and an edge $(i,j) \in [n] \times [m]$ exists if and only if $j \in I_f(i)$. The notation and definitions of $I_f(i), N_f(i), \deg_f(j)$ are then equivalent to those of $I_G(i), N_G(i), \deg_G(j)$.

As mentioned in Section 2, it will be useful to reduce a general d-local function to one having many non-connected neighborhoods of small size by deleting a few input bits.

Theorem 3.5 ([KOW24, Corollary 4.11]). *Let $\lambda, \kappa \geq 1$ be parameters (not necessarily constant) and let $F(\cdot)$ be an increasing function. Let $G = ([n], [m], E)$ be a d-left-bounded bipartite graph.*
Define

$$\widetilde{F}(x) = \frac{1}{d} \cdot \exp\left\{32d^4 x^2 \cdot F(2d \cdot x)\right\}. \tag{1}$$

Assume $H(\cdot)$ is an increasing function and $H(x) \geq \widetilde{F}(x)$ for all $x \geq L$ where $L \geq 1$ is some parameter not necessarily constant. If $H(x) \geq 2x$ for $x \geq L$ and

$$F(x) \geq 1 \text{ holds for all } x \geq 1 \quad \text{and} \quad \kappa \geq \lambda \geq d \cdot H^{(2d+2)}(L), \tag{2}$$

where $H^{(k)}$ is the iterated H of order k[4], then there exists $S \subseteq [m]$ such that deleting those right vertices (and their incident edges) produces a bipartite graph with r non-connected left neighborhoods of size at most t satisfying

$$|S| \leq \frac{r}{F(t)} \quad \text{and} \quad r \geq \frac{n}{\lambda} \quad \text{and} \quad t \leq \kappa.$$

Observe that in Theorem 3.5, even if F is a constant function, $\widetilde{F}$ (and hence H) will grow faster than an exponential function. This implies that the lower bound on κ and λ will (at least) be a tower-type blowup in d. Surprisingly, this is necessary [KOW24, Appendix A.2].

---

[4] $H^{(1)}(x) = H(x)$ and $H^{(k)}(x) = H(H^{(k-1)}(x))$ for $k \geq 2$.

Note that in the language of functions, Theorem 3.5 reduces d-local function into (d, r, t)-local functions, as defined below.

Definition 3.6 $((d, r, t)$-Local Function). We say $g\colon \{0, 1\}^m \to \{0, 1\}^n$ is a (d, r, t)-local function if g is a d-local function with r non-connected neighborhoods of size at most t.

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*

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

Density Comparison. We will need the following comparison result for the probability density function of the sum of independent random variables. It is a special case of the more general Theorem A.1; the proof and discussion on tightness and typical parameter choices are deferred to Appendix A.

Lemma 3.9. *Let $t \ge 1$ be an integer, and let $X_1, \ldots, X_n$ be independent random variables in $\{0, 1, \ldots, t\}$. 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*

$$\sum_{i \in [n]} (1 - p_{r,i}) \ge L > 0 \quad \text{holds for all } r \ge 3.$$

Let $m = \lfloor L/(32t^4) \rfloor$ and $\alpha = \left(\frac{L}{4n(t+1)}\right)^{2t^2}$. Then for any $x \in \mathbb{Z}$ and $0 \le \kappa_1, \kappa_2 \le \alpha \cdot m/128$,

$$\mathbf{Pr}\left[\sum_{i \in [n]} X_i = x\right] - C \cdot \mathbf{Pr}\left[\sum_{i \in [n]} X_i = x + \Delta\right] \le \sqrt{\frac{32}{\alpha \cdot m}} \cdot e^{-2\kappa_2}$$

holds for any $\Delta \in \mathbb{Z}$ and $C \in \mathbb{R}$ satisfying

$$|\Delta| \le 2\sqrt{\kappa_1 \cdot \alpha m} \text{ is an even number} \quad \text{and} \quad C \ge 2 \cdot e^{12 \cdot (\sqrt{\kappa_1 \kappa_2} + \kappa_1)}.$$

4 Classification of Locally Sampleable Hamming Slices

In this section, we will prove a general classification result for uniform distributions with symmetric support that can be sampled by local functions. Let $\Psi \subseteq \{0, 1, \ldots, n\}$ 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

- zeros $= \mathcal{D}_{\{0\}}$, ones $= \mathcal{D}_{\{n\}}$, and zerones $= \mathcal{D}_{\{0,n\}}$.

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

- all $= \mathcal{D}_{\{0,1,\ldots,n\}}$.

Theorem 4.1. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function. Assume $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \le \varepsilon$ for some $\Psi \subseteq \{0,1,\ldots,n\}$ and $n \ge \mathrm{tow}(900(d+1))$. Then*

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \le \mathrm{tow}(850(d+1)) \cdot \varepsilon$$

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

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

- The six special distributions admit local sampling schemes: zeros and ones can be sampled by a 0-local function; all and zerones can be sampled by a 1-local function; evens and 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 evens, we randomly flip the first $c \in [n]$ output bits with probability $1/4$. This distribution is 4-local since both 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 all and 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 evens and with probability $1/4$ is odds. Observe $\mathcal{D}$ can be sampled by a 3-local function via a similar strategy to that for evens. The uniform distribution over n-bit strings of Hamming weight $0,1,2,$ or 4 mod 6 is approximately $(1/6)$-close to $\mathcal{D}$; however, all six special distributions are $(1/4)$-far from $\mathcal{D}$. Thus, the implicit constant in Theorem 4.1 must be strictly greater than 1.

To prove Theorem 4.1, we will classify Ψ into several 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:

$$\iota(\Psi) = \arg\min_{s \in \Psi} |s - n/2|$$

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 $\frac{1}{2}\left(\mathcal{D}_{\iota(\Psi)} + \mathcal{D}_{n-\iota(\Psi)}\right)$, where we recall that $\mathcal{D}_k$ is the uniform distribution over the Hamming slice of weight k. Based on this intuition, we divide Ψ into the following cases:

- TAIL REGIME: $\iota(\Psi) \leq n/2^{d+2}$ or $\iota(\Psi) \geq n - n/2^{d+2}$.
- CENTRAL REGIME: $n/2^{d+2} < \iota(\Psi) < n - n/2^{d+2}$.

In the tail regime, we wish to show that $\mathcal{D}_\Psi$ can only be $\mathtt{zeros}$, $\mathtt{ones}$, or $\mathtt{zerones}$. This is formalized by the following consequence of [FLRS23, Theorem 1.2]. For completeness, we include a simple self-contained proof in Appendix B.

Theorem 4.3. *Let* $f\colon \{0,1\}^m \to \{0,1\}^n$ *be a d-local function where* $n \geq 2^{2^{8 \cdot (d+1)^2}}$. *Assume* $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq 1/2$ *for some* Ψ *in the tail regime. Then* $\mathcal{D}_\Psi \in \{\mathtt{zeros}, \mathtt{ones}, \mathtt{zerones}\}$.

In the central regime, we aim to show that $\mathcal{D}_\Psi$ is essentially $\mathtt{evens}$, $\mathtt{odds}$, or $\mathtt{all}$ by Theorem 4.4, which will be proved in Subsection 4.1.

Theorem 4.4. *Let* $f\colon \{0,1\}^m \to \{0,1\}^n$ *be a d-local function where* $n \geq \mathrm{tow}(900(d+1))$. *Assume* $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq \varepsilon$ *for some* Ψ *in the central regime. Then*

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq \mathrm{tow}(850(d+1)) \cdot \varepsilon$$

for some $\mathcal{D} \in \{\mathtt{evens}, \mathtt{odds}, \mathtt{all}\}$.

Combining Theorem 4.3 and Theorem 4.4, we can now establish Theorem 4.1.

Proof of Theorem 4.1. First note that the bound in Theorem 4.1 is trivial if $\varepsilon \geq \mathrm{tow}(850(d+1))^{-1}$. Hence we assume without loss of generality that $\varepsilon < \mathrm{tow}(850(d+1))^{-1} \leq 1/2$. If Ψ is in the tail regime, then by Theorem 4.3, we have $\mathcal{D}_\Psi \in \{\mathtt{zeros}, \mathtt{ones}, \mathtt{zerones}\}$. Hence by setting $\mathcal{D} = \mathcal{D}_\Psi$, we have $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq \varepsilon \leq \mathrm{tow}(850(d+1)) \cdot \varepsilon$. If Ψ is in the central regime, then the bound follows directly from Theorem 4.4. □

4.1 Central Regime

In this section, we deal with the central regime where strings from $\mathcal{D}_\Psi$ are spread out in the middle layers, i.e., $n/2^{d+2} < \iota(\Psi) < n - n/2^{d+2}$ where $\iota(\Psi)$ is the Hamming weight in Ψ closest to $n/2$.

Theorem (Theorem 4.4 Restated)**.** *Let* $f\colon \{0,1\}^m \to \{0,1\}^n$ *be a d-local function where* $n \geq \mathrm{tow}(900(d+1))$. *Assume* $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq \varepsilon$ *for some* Ψ *in the central regime. Then*

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq \mathrm{tow}(850(d+1)) \cdot \varepsilon$$

for some $\mathcal{D} \in \{\mathtt{evens}, \mathtt{odds}, \mathtt{all}\}$.

Recall that g is a (d, r, t)*-local function* if (1) g is a d-local function, and (2) it has r non-connected neighborhoods of size at most t. The high-level idea underlying the proof of Theorem 4.4 is to use graph elimination results from [KOW24] to reduce an arbitrary d-local function to a mixture of (d, r, t)-local functions with r as a small multiple of n and t as a large constant. This additional structure will make the analysis substantially easier. After proving bounds for each of the subfunctions, we amalgamate them into one for the original function via a union bound argument. We will first prove the following distance lower bounds for (d, r, t)-local functions in Subsection 4.1.1.

Proposition 4.5. *Let Ψ be in the central regime and assume $\|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}} \geq \delta$ holds for all $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}, \texttt{all}\}$, where*

$$\delta \geq \exp\left\{-n/\left(2^{30(dt+1)} \cdot n/r\right)^{3(t+1)^2}\right\} \quad and \quad n \geq \left(2^{30(dt+1)} \cdot n/r\right)^{6(t+1)^2} \quad and \quad r \geq 2^{100(dt+1)}.$$

Then there exists a function $\phi\colon \{0,1\}^n \to \mathbb{R}$ such that $-\exp\left\{2^{100(dt+1)^3} \cdot (n/r)^{4(t+1)^2}\right\} \leq \phi(x) \leq 1$ holds for all $x \in \{0,1\}^n$.

Moreover, let $g\colon \{0,1\}^m \to \{0,1\}^n$ be an arbitrary (d,r,t)-local function and define $\mathcal{P}_g = g(\mathcal{U}^m)$. Then either

$$\|\mathcal{P}_g - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - n^2 \cdot \exp\left\{-r/2^{20(dt+1)}\right\}$$

or

$$\mathop{\mathbb{E}}_{X \sim \mathcal{D}_\Psi}[\phi(X)] - \mathop{\mathbb{E}}_{X \sim \mathcal{P}_g}[\phi(X)] \geq \delta \cdot 2^{-120} - (t+1) \cdot \exp\left\{-r/2^{20(dt+1)}\right\}.$$

We remark that in [KOW24], the potential function ϕ is merely the indicator function of the support of the desired distribution, which has tiny probability mass in the actual d-local distribution. Since our Ψ is arbitrary in the central regime, the actual construction of ϕ in Proposition 4.5 is much more delicate and is no longer a simple indicator function.

To reduce the actual d-local function to a (d,r,t)-local function, we use graph elimination results from [KOW24] as before.

Proposition 4.6. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function. There exists a set $S \subseteq [m]$ such that any fixing of input bits in S reduces f to a (d,r,t)-local function g where*

$$|S| \leq \frac{r}{2^{30(dt+1)}} \quad and \quad r \geq \frac{n}{\text{tow}(600d)} \quad and \quad t \leq \text{tow}(600d).$$

Proof. The statement is trivial when $d = 0$ since then we can set $S = \emptyset, r = n, t = 0$. For $d \geq 1$, we apply Theorem 3.5. Set $F(x) = 2^{30 \cdot (dx+1)}$. Then

$$\widetilde{F}(x) = \frac{1}{d} \cdot \exp\left\{32d^4 x^2 \cdot 2^{30d \cdot (2d^2 x + 1)}\right\}.$$

Define $H(x) = 2^{2^{2^x}}$ and let $L = 500d$. By setting

$$\kappa = \lambda = \text{tow}(600d) \geq d \cdot H^{(2d+2)}(L),$$

the conditions in Theorem 3.5 are satisfied, yielding the result. □

In [KOW24], Lemma 3.3 is used to combine Proposition 4.6 and Proposition 4.5. Here we need Lemma 4.7, which is a slight strengthening of Lemma 3.3, since we have a potential function that is not necessarily an indicator of an event. Nevertheless, the proof carries over and is deferred to Appendix C.

Lemma 4.7. *Let $\mathcal{P}_1, \ldots, \mathcal{P}_\ell$ and $\mathcal{Q}$ be distributions. Let ϕ be a function with output range $[a,b]$ where $a < b$. Assume for each $i \in [\ell]$,*

$$either \quad \|\mathcal{P}_i - \mathcal{Q}\|_{\mathsf{TV}} \geq 1 - \eta_1 \quad or \quad \mathop{\mathbb{E}}_{X \sim \mathcal{Q}}[\phi(X)] - \mathop{\mathbb{E}}_{X \sim \mathcal{P}_i}[\phi(X)] \geq \eta_2.$$

Then for any distribution $\mathcal{P}$ as a convex combination of $\mathcal{P}_1, \ldots, \mathcal{P}_\ell$, we have

$$\|\mathcal{P} - \mathcal{Q}\|_{\mathsf{TV}} \geq \frac{\eta_2}{b-a} - (\ell+1) \cdot \eta_1.$$

14

At this point, we can prove Theorem 4.8, a weaker version of Theorem 4.4 equipped with an additional assumption on ε being not too small.

Theorem 4.8. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function where $n \geq \mathrm{tow}(900(d+1))$. Assume $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq \varepsilon$ for some Ψ in the central regime and*

$$\varepsilon \geq \exp\left\{-\frac{n}{\mathrm{tow}(750(d+1))}\right\}.$$

Then

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq \mathrm{tow}(750(d+1)) \cdot \varepsilon$$

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

Proof. Let $\delta = \mathrm{tow}(750(d+1)) \cdot \varepsilon$, which satisfies

$$\delta \geq \exp\left\{-\frac{n}{\mathrm{tow}(800(d+1))}\right\}. \tag{3}$$

Assume towards a contradiction that $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} > \delta$ for all $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}, \texttt{all}\}$.

First we reduce f to (d,r,t)-local functions. By Proposition 4.6, we find a set $S \subseteq [m]$ such that any fixing ρ of input bits in S reduces f to a (d,r,t)-local function f_ρ where

$$|S| \leq \frac{r}{2^{30d(t+1)}} \quad \text{and} \quad r \geq \frac{n}{\mathrm{tow}(600d)} \quad \text{and} \quad t \leq \mathrm{tow}(600d). \tag{4}$$

Then we use Proposition 4.5 to analyze each f_ρ. Since $n \geq \mathrm{tow}(900(d+1))$, by (3) and (4), the conditions in Proposition 4.5 holds. Hence, by Proposition 4.5, there exists a function $\phi\colon \{0,1\}^n \to \mathbb{R}$, which depends only on Ψ, such that

(1) $a := 1 - \mathrm{tow}(700(d+1)) \leq \phi(x) \leq 1 =: b$ holds for all $x \in \{0,1\}^n$,

(2) for each f_ρ, define $\mathcal{P}_g = g(\mathcal{U}^m)$ then either

$$\left\|\mathcal{P}_{f_\rho} - \mathcal{D}_\Psi\right\|_{\mathsf{TV}} \geq 1 - n^2 \exp\left\{-r/2^{20d(t+1)}\right\} =: 1 - \eta_1$$

or

$$\mathbb{E}_{X\sim\mathcal{D}_\Psi}[\phi(X)] - \mathbb{E}_{X\sim\mathcal{P}_{f_\rho}}[\phi(X)] \geq \delta/2^{120} - \exp\left\{-\frac{n}{\mathrm{tow}(700(d+1))}\right\} \geq \delta/2^{130} =: \eta_2.$$

Note that the above bounds are simplified using (4) and $n \geq \mathrm{tow}(900(d+1))$.

Finally we use Lemma 4.7 with $\mathcal{P} = f(\mathcal{U}^m)$, $\mathcal{Q} = \mathcal{D}_\Psi$, and $\mathcal{P}_1, \ldots, \mathcal{P}_\ell$ being the output distributions of f_ρ's, as well as the parameters a, b, η_1, η_2 defined above. Then we have

$$
\begin{aligned}
\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} &\geq \frac{\delta/2^{130}}{\mathrm{tow}(700(d+1))} - \left(2^{|S|} + 1\right) \cdot n^2 \exp\left\{-r/2^{20d(t+1)}\right\} \\
&\geq \frac{\delta/2^{130}}{\mathrm{tow}(700(d+1))} - n^2 \cdot \exp\left\{-r/2^{25d(t+1)}\right\} && \text{(by (4))} \\
&\geq \frac{\delta/2^{130} - \delta/2^{200}}{\mathrm{tow}(700(d+1))} && \text{(by (4), } n \geq \mathrm{tow}(900(d+1)), \text{ and (3))} \\
&> \delta/\mathrm{tow}(750(d+1)) = \varepsilon,
\end{aligned}
$$

which is a contradiction. $\qquad\square$

To further lift the assumption on ε, we need to explore extra properties of local functions. This is handled in the following Theorem 4.9, the proof of which is presented in Subsection 4.1.2.

Theorem 4.9. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function where $n \geq \mathrm{tow}(900(d+1))$. Assume $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq \varepsilon$ for some Ψ in the central regime and*

$$\varepsilon \leq \exp\left\{-\frac{n}{\mathrm{tow}(800(d+1))}\right\}.$$

Then

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq \mathrm{tow}(850(d+1)) \cdot \varepsilon$$

for some $\mathcal{D} \in \{\mathtt{evens}, \mathtt{odds}, \mathtt{all}\}$.

The full version of Theorem 4.4 follows from the combination of Theorem 4.8 and Theorem 4.9.

Proof of Theorem 4.4. If $\varepsilon \geq \exp\{-n/\mathrm{tow}(750(d+1))\}$, then we apply Theorem 4.8. Otherwise we have $\varepsilon < \exp\{-n/\mathrm{tow}(750(d+1))\} \leq \exp\{-n/\mathrm{tow}(800(d+1))\}$, from which we apply Theorem 4.9. □

4.1.1 Analysis of More Structured Local Functions

In this part, we prove Proposition 4.5, which concerns the more structured (d,r,t)-local functions. For convenience, we additionally assume $d, t \geq 1$ in the restated version of Proposition 4.5 below. This immediately implies the original statement by shifting by one, since any (d,r,t)-local function is also (d',r,t')-local for any $t' \geq t$ and $d' \geq d$.

Proposition (Proposition 4.5 Restated). *Let $d \geq 1$, $t \geq 1$, and $r \geq 2^{100dt}$. Assume Ψ is in the central regime and assume $\|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}} \geq \delta$ for all $\mathcal{D} \in \{\mathtt{evens}, \mathtt{odds}, \mathtt{all}\}$, where*

$$\delta \geq \exp\left\{-n \cdot \left(2^{30dt} \cdot n/r\right)^{-3t^2}\right\} \quad and \quad n \geq \left(2^{30dt} \cdot n/r\right)^{6t^2}.$$

Then there exists a function $\phi\colon \{0,1\}^n \to \mathbb{R}$ such that $-\exp\left\{2^{100dt^3} \cdot (n/r)^{4t^2}\right\} \leq \phi(x) \leq 1$ holds for all $x \in \{0,1\}^n$.

Moreover, let $g\colon \{0,1\}^m \to \{0,1\}^n$ be an arbitrary (d,r,t)-local function and define $\mathcal{P}_g = g(\mathcal{U}^m)$. Then either

$$\|\mathcal{P}_g - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - n^2 \cdot \exp\left\{-r/2^{20dt}\right\}$$

or

$$\mathbb{E}_{X \sim \mathcal{D}_\Psi}[\phi(X)] - \mathbb{E}_{X \sim \mathcal{P}_g}[\phi(X)] \geq \delta \cdot 2^{-120} - t \cdot \exp\left\{-r/2^{20dt}\right\}.$$

To prove Proposition 4.5, we fix a (d,r,t)-local function $g\colon \{0,1\}^m \to \{0,1\}^n$, where r and t are parameters to be plugged in later. Recall that for each $i \in [n]$, its neighborhood $N_g(i) \subseteq [n]$ is the set of output bits that share common input bits with the i-th output. Let $\varepsilon \in [0,1]$ be a parameter to be optimized later. We classify each $N(i) = N_g(i)$ of size $s_i = |N(i)|$ into the following cases:

- Type-1. $\mathcal{P}_g|_{N(i)}$ is not ε-close to $\mathcal{U}^{s_i}_{s/n}$ for each $n/2^{d+3} \leq s \leq n - n/2^{d+3}$.

- Type-2. $\mathcal{P}_g|_{N(i)}$ is ε-close to $\mathcal{U}^{s_i}_{s/n}$ for some $n/2^{d+3} \leq s \leq n - n/2^{d+3}$.

The classification of Type-1 and Type-2 neighborhoods is different from the ones in [KOW24], where they only have a single unique biased distribution to compare with. The definition here takes into consideration all noticeable Hamming weights since $\mathcal{D}_\Psi$ can be a mixture of very different biased distributions, e.g., if $\Psi = \{n/3, 2n/3\}$, then $\mathcal{D}_\Psi$ is roughly a uniform mixture of the 1/3-biased and 2/3-biased distributions, which is not close to any product distribution.

Recall that Ψ being in the central regime means there exists some $s \in \Psi$ such that $n/2^{d+2} < s < n - n/2^{d+2}$. Hence focusing only on the Hamming weights within range $[n/2^{d+3}, n - n/2^{d+3}]$, as in the definition of Type-1 and Type-2 neighborhoods, does not differ too much. This is formalized by Claim 4.10 and is proved in Appendix C.

Claim 4.10. If Ψ is in the central regime, then $\overline{\Psi} \neq \emptyset$ and

$$\left\| \mathcal{D}_\Psi - \mathcal{D}_{\overline{\Psi}} \right\|_{\mathsf{TV}} \leq 8 \cdot \exp\left\{ -n \cdot 2^{-(d+4)} \right\},$$

where $\overline{\Psi} = \left\{ s \in \Psi \colon n/2^{d+3} \leq s \leq n - n/2^{d+3} \right\}$.

If g has many small non-connected Type-1 neighborhoods, then we show that $\mathcal{P}_g$ is extremely far from $\mathcal{D}_{\overline{\Psi}}$, and thus $\mathcal{D}_\Psi$ by Claim 4.10, in Lemma 4.12. This is proved via a simple reduction to the analysis of Type-1 neighborhoods in the single Hamming slice case in [KOW24], quoted as Lemma 4.11 below.

Lemma 4.11 ([KOW24, Lemma 5.14]). *Assume there are $r' \geq 1$ non-connected Type-1 neighborhoods. Then*

$$\left\| \mathcal{P}_g - \mathcal{D}_k \right\|_{\mathsf{TV}} \geq 1 - 2\sqrt{2n} \cdot \exp\left\{ -\varepsilon^2 r'/2 \right\}.$$

Lemma 4.12. *Assume there are $r' \geq 1$ non-connected Type-1 neighborhoods. Then*

$$\left\| \mathcal{P}_g - \mathcal{D}_\Psi \right\|_{\mathsf{TV}} \geq 1 - 4n^{1.5} \cdot \exp\left\{ -\varepsilon^2 r'/8 \right\} - 8 \cdot \exp\left\{ -n \cdot 2^{-(d+4)} \right\}.$$

Proof. Let $\overline{\Psi} = \left\{ s \in \Psi \colon n/2^{d+3} \leq s \leq n - n/2^{d+3} \right\}$, which is non-empty by Claim 4.10. For each $s \in \overline{\Psi}$, we apply Lemma 4.11 to obtain $\left\| \mathcal{P}_g - \mathcal{D}_s \right\|_{\mathsf{TV}} \geq 1 - 4\sqrt{n} \cdot \exp\left\{ -\varepsilon^2 r'/8 \right\}$. Since $\mathcal{D}_{\overline{\Psi}}$ is the convex combination of $\mathcal{D}_s$ for $s \in \overline{\Psi}$, we have

$$\left\| \mathcal{P}_g - \mathcal{D}_{\overline{\Psi}} \right\|_{\mathsf{TV}} \geq 1 - 4n^{1.5} \cdot \exp\left\{ -\varepsilon^2 r'/8 \right\},$$

where we used Lemma 3.3 and the fact that $|\overline{\Psi}| < n$. Then the final bound follows from Claim 4.10 and a triangle inequality. $\square$

If g has many small non-connected Type-2 neighborhoods, we show that the Hamming weight of $\mathcal{P}_g$ is the convex combination of sums of many independent bounded random integers with additional congruence properties.

This "open-boxes" the proofs of [KOW24, Lemmas 5.15 & 5.22]. The key point here is that ε only needs to be mildly small depending on d and t to obtain the congruence properties by Lemma 3.4, since every s in the definition of a Type-2 neighborhood is bounded away from the boundary (i.e., $1/2^{d+3} \leq s/n \leq 1 - 1/2^{d+3}$). The proof of Lemma 4.13 is deferred to Appendix C.

Lemma 4.13. *Assume there are $r' \geq 1$ non-connected Type-2 neighborhoods of size at most t. If $\varepsilon \leq 2^{-3t-d-3}$, then the distribution of the Hamming weight of $X \sim \mathcal{P}_g$ can be decomposed as $\sum_\rho \lambda_\rho \cdot \mathcal{P}_\rho$ where*

1. $\sum_\rho \lambda_\rho = 1$ and each $\lambda_\rho \geq 0$;

17

2. *each* $\mathcal{P}_\rho = X_\rho + \sum_{j\in[r']} X_{\rho,j}$ *where* X_ρ *is a fixed integer and* $X_{\rho,1}, \ldots, X_{\rho,r'}$ *are independent random variables in* $\{0, 1, \ldots, t\}$,

3. *we say* ρ *is bad if there exists some integer* $q \geq 3$ *such that* $\sum_{j\in[r']}(1 - p_{\rho,q,j}) \leq 2^{-7t-d} \cdot r'$, *where* $p_{\rho,q,j} = \max_{x\in\mathbb{Z}} \mathbf{Pr}\left[X_{\rho,j} \equiv x \pmod{q}\right]$. *Then* $\sum_{bad\ \rho} \lambda_\rho \leq t \cdot \exp\left\{-2^{-7t-d-2} \cdot r'\right\}$.

Note that Item 3 from Lemma 4.13 does not guarantee any congruence property for modulus 2. This suggests that we should handle even and odd numbers in Ψ separately. Indeed, in the following Lemma 4.14 we show that if $\mathcal{D}_\Psi$ is not close to `evens`, `odds`, and `all`, then either its even part is far from `evens` or its odd part is far from `odds`. We remark that Lemma 4.14 also works for Ψ not in the central regime.

Lemma 4.14. *Let* Γ *and* Ξ *be the set of even and odd numbers in* Ψ *respectively. Define* $\gamma = |\mathsf{supp}\,(\mathcal{D}_\Gamma)|/2^{n-1}$ *and* $\xi = |\mathsf{supp}\,(\mathcal{D}_\Xi)|/2^{n-1}$. *Then* $\mathcal{D}_\Psi$ *can be decomposed as* $\frac{\gamma}{\gamma+\xi} \cdot \mathcal{D}_\Gamma + \frac{\xi}{\gamma+\xi} \cdot \mathcal{D}_\Xi$.[5] *In addition, if* $\|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}} \geq \delta \geq 0$ *for every* $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}, \texttt{all}\}$, *then*

$$either \quad \frac{\gamma \cdot \|\mathcal{D}_\Gamma - \texttt{evens}\|_{\mathsf{TV}}}{\gamma + \xi} = \frac{\gamma(1-\gamma)}{\gamma+\xi} \geq \frac{\delta}{12} \quad or \quad \frac{\xi \cdot \|\mathcal{D}_\Xi - \texttt{odds}\|_{\mathsf{TV}}}{\gamma + \xi} = \frac{\xi(1-\xi)}{\gamma+\xi} \geq \frac{\delta}{12}. \quad (5)$$

Proof. The decomposition follows directly from the fact that $\mathcal{D}_\Psi, \mathcal{D}_\Gamma, \mathcal{D}_\Xi$ are uniform distributions over their supports respectively. Since $\mathsf{supp}\,(\mathcal{D}_\Gamma) \subseteq \mathsf{supp}\,(\texttt{evens})$, $\mathsf{supp}\,(\mathcal{D}_\Xi) \subseteq \mathsf{supp}\,(\texttt{odds})$, and $\Psi \neq \emptyset$, we have

$$0 \leq \gamma \leq 1, \quad 0 \leq \xi \leq 1, \quad \text{and} \quad \gamma + \xi > 0. \quad (6)$$

Now we can express $\|\mathcal{D}_\Psi - \mathcal{D}\|_{\mathsf{TV}}$. Starting with $\mathcal{D} = \texttt{all}$ and by Fact 3.2, we have

$$\|\mathcal{D}_\Psi - \texttt{all}\|_{\mathsf{TV}} = 1 - \frac{|\mathsf{supp}\,(\mathcal{D}_\Psi)|}{2^n} = 1 - \frac{|\mathsf{supp}\,(\mathcal{D}_\Gamma)| + |\mathsf{supp}\,(\mathcal{D}_\Xi)|}{2^n} = 1 - \frac{\gamma + \xi}{2}.$$

Let $\delta' = \delta/3 \in [0, 1/3]$. Since $\|\mathcal{D}_\Psi - \texttt{all}\|_{\mathsf{TV}} \geq \delta \geq \delta'$, this gives

$$\gamma + \xi \leq 2 \cdot (1 - \delta'). \quad (7)$$

For $\mathcal{D} = \texttt{evens}$, we work directly with the definition of total variation distance:

$$\|\mathcal{D}_\Psi - \texttt{evens}\|_{\mathsf{TV}} = \frac{1}{2}\left(\left|\frac{|\mathsf{supp}\,(\mathcal{D}_\Gamma)|}{|\mathsf{supp}\,(\mathcal{D}_\Psi)|} - \frac{|\mathsf{supp}\,(\mathcal{D}_\Gamma)|}{2^{n-1}}\right| + \frac{2^{n-1} - |\mathsf{supp}\,(\mathcal{D}_\Gamma)|}{2^{n-1}} + \frac{|\mathsf{supp}\,(\mathcal{D}_\Xi)|}{|\mathsf{supp}\,(\mathcal{D}_\Psi)|}\right)$$

$$= \frac{1}{2} \cdot \left(\left|\frac{\gamma}{\gamma+\xi} - \gamma\right| + 1 - \gamma + \frac{\xi}{\gamma+\xi}\right),$$

which implies

$$\delta' \leq \begin{cases} 1 - \gamma & \gamma + \xi \leq 1, \\ \frac{\xi}{\gamma+\xi} & \gamma + \xi > 1. \end{cases} \quad (8)$$

Similarly by inspecting $\mathcal{D} = \texttt{odds}$, we have

$$\delta' \leq \begin{cases} 1 - \xi & \gamma + \xi \leq 1, \\ \frac{\gamma}{\gamma+\xi} & \gamma + \xi > 1. \end{cases} \quad (9)$$

In addition, by Fact 3.2, the equalities in (5) hold and the inequalities follows from solving the following optimization problem, which is proved in Appendix C.

---

[5]Even if, say, $\mathcal{D}_\Gamma$ is undefined due to $\Gamma = \emptyset$, this decomposition is still valid, since then $\gamma = 0$ and $\mathcal{D}_\Psi = \mathcal{D}_\Xi$.

Claim 4.15. Given constraints (6), (7), (8), and (9), we have $\frac{\gamma^2 + \xi^2}{\gamma + \xi} \leq 1 - \delta'/2$.

Assuming (5) is false, then we should have

$$\frac{\gamma(1-\gamma)}{\gamma + \xi} + \frac{\xi(1-\xi)}{\gamma + \xi} = 1 - \frac{\gamma^2 + \xi^2}{\gamma + \xi} < \frac{\delta}{6} = \frac{\delta'}{2},$$

which, however, contradicts Claim 4.15. $\qquad\qquad\square$

Given Lemma 4.14, we can focus on the even part or the odd part of Ψ. Say, the even part Γ of Ψ witnesses a large deviation from evens. Then we construct a potential function ϕ such that (1) its expectation is small under each typical $\mathcal{P}_\rho$, and (2) it is large under $\mathcal{D}_\Gamma$. In retrospect, the potential function was defined to be the indicator of the support of the desired distribution in [KOW24, Lemmas 5.15 & 5.22], where (2) held trivially and (1) followed from anticoncentration (i.e. [Ush86, Theorem 3] or [KOW24, Lemma 3.7]). Since here we have an unstructured Γ, it is much more delicate to construct such a potential function and verify conditions (1) and (2).

Informally, ϕ is constructed to indicate part of the Hamming weights in Γ (i_ℓ's in Lemma 4.16) and subtracts some nearby Hamming weights that are not covered by Γ (j_ℓ's in Lemma 4.16). To guarantee (1), we will use density comparison results (Lemma 3.9); for (2), we will make sure that the set of indicated weights in Γ consists of a noticeable probability mass in $\mathcal{D}_\Gamma$.

Lemma 4.16. *Assume $n \geq 2^{10}$ and $\delta_1, \delta_2 \in (0,1)$ satisfying $\delta_1(1 - \delta_2) \leq 2^{-100}$ and $\delta_1 \leq \delta_2$. Let $\Gamma \subset \{0, 1, \ldots, n\}$ be non-empty and contain only even numbers. If $\|\mathcal{D}_\Gamma - \text{evens}\|_{\text{TV}} \geq \delta_2$, then there exist $i_1, \ldots, i_K, j_1, \ldots, j_K$ such that*

1. *$i_1, \ldots, i_K, j_1, \ldots, j_K$ are distinct even numbers in $\{0, 1, \ldots, n\}$,*

2. *$\mathbf{Pr}_{X \sim \mathcal{D}_\Gamma} [|X| \in \{i_1, \ldots, i_K\}] \geq \delta_1/60$ and $\mathbf{Pr}_{X \sim \mathcal{D}_\Gamma} [|X| \in \{j_1, \ldots, j_K\}] = 0$,*

3. *$K \leq 1 + 8\sqrt{n / \log\left(\frac{1}{\delta_1(1-\delta_2)}\right)}$ and $|i_\ell - j_\ell| \leq 4K$ holds for each $\ell \in [K]$.*

Proof. For each $m \in \{0, 1, \ldots, n\}$, define

$$\text{wt}_m = 2^{-n+1}\binom{n}{m} \quad \text{and} \quad \text{wt}_{\leq m} = 2^{-n+1}\sum_{i \geq 0}\binom{n}{m - 2i} \quad \text{and} \quad \text{wt}_{\geq m} = 2^{-n+1}\sum_{i \geq 0}\binom{n}{m + 2i}.$$

For even m's, wt_m captures the Hamming weight distribution of evens.

Let $\gamma = \sum_{m \in \Gamma} \text{wt}_m$, which equals $|\text{supp}(\mathcal{D}_\Gamma)|/2^{n-1}$. Then

$$\Pr_{X \sim \mathcal{D}_\Gamma} [|X| = m] = \text{wt}_m/\gamma \quad \text{for each } m \in \Gamma. \tag{10}$$

In addition, by Fact 3.2, we have

$$\|\mathcal{D}_\Gamma - \text{evens}\|_{\text{TV}} = \sum_{\text{even } m \notin \Gamma} \text{wt}_m = 1 - \gamma \geq \delta_2 > 0. \tag{11}$$

If $\delta_1 \gamma \leq 2^{-n+1}$, then we set $K = 1$ and select arbitrary $i_1 \in \Gamma, j_1 \notin \Gamma$ such that $|i_1 - j_1| = 2$. Since $\|\mathcal{D}_\Gamma - \text{evens}\|_{\text{TV}} > 0$, such a pair always exists. Then Lemma 4.16 follows immediately from (10) and the fact that $\text{wt}_m \geq 2^{-n+1}$ for every even m.

From now on we assume $\delta_1 \gamma > 2^{-n+1}$. Define

$$m_{\text{L}} = \min\{\text{even } m \colon \text{wt}_{\leq m} \geq \delta_1\gamma/4\} \quad \text{and} \quad m_{\text{R}} = \max\{\text{even } m \colon \text{wt}_{\geq m} \geq \delta_1\gamma/4\}.$$

Note that m_{L} and m_{R} are well-defined. We will need the following estimate for the binomial weights sandwiched between them. The proof of Claim 4.17 is a direct but tedious calculation on binomial coefficients and is deferred to Appendix C.

Claim 4.17. $m_\mathsf{L} \leq n/2 - \sqrt{n}$ and $m_\mathsf{R} \geq n/2 + \sqrt{n}$. Furthermore, for each even $m \in [m_\mathsf{L}, m_\mathsf{R}]$, we have $\mathsf{wt}_m \geq \frac{\delta_1 \gamma \sqrt{\log(1/(\delta_1\gamma))}}{16\sqrt{n}}$.

Define

$$I = \{m \in \Gamma \colon m_\mathsf{L} \leq m \leq m_\mathsf{R}\} \quad \text{and} \quad J = \{\text{even } m \colon m_\mathsf{L} \leq m \leq m_\mathsf{R}\} \setminus I.$$

We will select $i_1, \ldots, i_K$ from I and $j_1, \ldots, j_K$ from J. First observe that

$$\sum_{m \in I} \mathsf{wt}_m \geq \sum_{m \in \Gamma} \mathsf{wt}_m - \mathsf{wt}_{\leq m_\mathsf{L}-2} - \mathsf{wt}_{\geq m_\mathsf{R}+2} \geq \gamma - \delta_1\gamma/2 \geq \delta_1\gamma/2, \tag{12}$$

which, combined with (10), implies that the first half of Item 2 will be eventually satisfied if we keep picking out i_k's from I. We describe the actual process separately depending on whether J is abundant.

The $|J| \geq \sqrt{n}/9$ Case. We will simply select (i_k, j_k) iteratively until Item 2 is satisfied, and each pair is selected to minimize their absolute distance to guarantee Item 3.

- Initialize $I_0 = I$ and $J_0 = J$.
- For each $k = 0, 1, 2, \ldots$, if $\sum_{\ell \in [k]} \mathsf{wt}_{i_\ell} \geq \delta_1\gamma/16$, then set $K = k$ and terminate; otherwise define $I_{k+1} \leftarrow I_k \setminus \{i_k\}$, $J_{k+1} \leftarrow J_k \setminus \{j_k\}$ then continue, where

$$(i_k, j_k) = \operatorname*{arg\,min}_{(i,j) \in I_k \times J_k} |i - j|.$$

By Claim 4.17 and the definition of I, we know that if the process does not terminate at time k, then

$$\frac{\delta_1\gamma}{16} > \sum_{\ell \in [k]} \mathsf{wt}_{i_\ell} \geq k \cdot \frac{\delta_1\gamma\sqrt{\log(1/(\delta_1\gamma))}}{16\sqrt{n}} \geq k \cdot \frac{\delta_1\gamma\sqrt{\log\left(\frac{1}{\delta_1(1-\delta_2)}\right)}}{16\sqrt{n}},$$

where we used $\gamma \leq 1 - \delta_2$ from (11). This means that the process must terminate before $k > 1 + \sqrt{n/\log\left(\frac{1}{\delta_1(1-\delta_2)}\right)}$. Since $|J| \geq \sqrt{n}/9$ and $\delta_1(1-\delta_2) \leq 2^{-100}$, we have $|J| \geq 1 + \sqrt{n/\log\left(\frac{1}{\delta_1(1-\delta_2)}\right)}$, which implies that we will never run out of j_k before termination and the above process is indeed valid. This argument, combined with (10) and the definition of each I_k and J_k, already verifies Item 1, Item 2, and the first half of Item 3. For the second half of Item 3, a simple induction argument shows that the k-th selected pair has distance at most $4k - 2$, which is at most $4K$ as desired.

The $|J| < \sqrt{n}/9$ Case. Here, the naive greedy selection approach may not work since we may not have a sufficient amount of j_k's to pair with i_k's. However, by (11), the total weight in J is also noticeable:

$$\sum_{m \in J} \mathsf{wt}_m \geq \sum_{\text{even } m \notin \Gamma} \mathsf{wt}_m - \mathsf{wt}_{\leq m_\mathsf{L}-2} - \mathsf{wt}_{\geq m_\mathsf{R}+2} \geq \delta_2 - \delta_1\gamma/2 \geq \delta_1\gamma/2, \tag{13}$$

where for the last inequality we used $\delta_1\gamma \leq \delta_1 \leq \delta_2$. If we cleverly select pairs until $\sum_\ell \mathsf{wt}_{j_\ell}$ is large enough, $\sum_\ell \mathsf{wt}_{i_\ell}$ should also be large since $|J|$ itself is small and hence $\mathsf{wt}_{i_\ell}, \mathsf{wt}_{j_\ell}$ shouldn't be off by much. Now we describe the actual process.

20

- Initialize $I_0 = I$ and $J_0 = J$.
- For each $k = 0, 1, 2, \ldots$, if $\sum_{\ell \in [k]} \mathsf{wt}_{j_\ell} \geq \delta_1 \gamma / 2$ or $\sum_{\ell \in [k]} \mathsf{wt}_{i_\ell} \geq \delta_1 \gamma / 60$, then set $K = k$ and terminate; otherwise define $I_{k+1} \leftarrow I_k \setminus \{i_k\}$, $J_{k+1} \leftarrow J_k \setminus \{j_k\}$ then continue, where i_k, j_k are selected as follows:

    (A) if there exists $i \in I_k, j \in J_k$ that $n/2 \leq i < j$, then $(i_k, j_k) = \arg\min_{\substack{(i,j) \in I_k \times J_k \\ n/2 \leq i < j}} |i - j|$,

    (B) else if there exists $i \in I_k, j \in J_k$ that $n/2 \geq i > j$, then $(i_k, j_k) = \arg\min_{\substack{(i,j) \in I_k \times J_k \\ n/2 \geq i > j}} |i - j|$,

    (C) else if $I_k \neq \emptyset$ and $J_k \neq \emptyset$, then $(i_k, j_k) = \arg\min_{(i,j) \in I_k \times J_k} |i - j|$,

    (D) else, by (13), we must have $I_k = \emptyset$. Then the process errs.

To prove the validity of the above process, we first show that the process does not err. Since each $\mathsf{wt}_m, m \in J$ is also lower bounded by Claim 4.17, $\sum_{\ell \in [k]} \mathsf{wt}_{j_\ell} \geq \delta_1 \gamma / 2$ will be satisfied before $k > 1 + 8\sqrt{n / \log\left(\frac{1}{\delta_1(1-\delta_2)}\right)}$ by the same argument as in the previous case. On the other hand, Claim 4.17 also implies that $|I| + |J| \geq \sqrt{n} - 1$. Since $|J| < \sqrt{n}/9$, we have $|I| > 8\sqrt{n}/9 - 1$, which is larger than $1 + 8\sqrt{n / \log\left(\frac{1}{\delta_1(1-\delta_2)}\right)}$ as $\delta_1(1 - \delta_2) \leq 2^{-100}$ and $n \geq 2^{10}$. Hence we never run out of i_k in Item (D) before termination and the above process does not err. This argument, combined with the definition of each I_k and J_k, already verifies Item 1, the second half of Item 2, and the first half of Item 3. The second half of Item 3 can be analogously established by an induction argument that the k-th selected pair has distance at most $4k - 2 \leq 4K$.

All that remains is the first half of Item 2, i.e., $\sum_\ell \mathsf{wt}_{i_\ell}$ is large upon termination. To this end, we relate each wt_{i_k} and wt_{j_k} by Claim 4.18.

Claim 4.18. For any time k before termination, we have $\mathsf{wt}_{i_k} \geq \mathsf{wt}_{j_k}/30$.

Proof. If (i_k, j_k) is obtained in Item (A) or Item (B), then by the definition of wt, we have $\mathsf{wt}_{i_k} > \mathsf{wt}_{j_k} \geq \mathsf{wt}_{j_k}/8$ immediately. Now we assume that (i_k, j_k) is obtained in Item (C). We will show that both i_k and j_k are within range $[n/2 - \sqrt{n}, n/2 + \sqrt{n}]$, from which we have the desired relation:

$$\frac{\mathsf{wt}_{i_k}}{\mathsf{wt}_{j_k}} = \frac{\binom{n}{i_k}}{\binom{n}{j_k}} \geq \frac{\binom{n}{n/2-\sqrt{n}}}{\binom{n}{n/2}} \geq \frac{2^{n \cdot \left(\mathcal{H}(1/2 - 1/\sqrt{n}) - 1\right)}}{\sqrt{8 \cdot (1 - 4/n)/\pi}} \qquad \text{(by Fact 3.7 and the range of } i_k, j_k\text{)}$$

$$\geq \frac{2^{-4}}{\sqrt{8/\pi}} \geq \frac{1}{30}. \qquad \text{(by Fact 3.8)}$$

If $i_k < n/2 - \sqrt{n}$, then $j_k > i_k$ since otherwise we should be in Item (B). In addition, every even $m \in [n/2 - \sqrt{n}, n/2]$ is either selected already or contained in J_k. To see this, if there exists $m \in [n/2 - \sqrt{n}, n/2] \cap I_k$ and $m > j_k$, then we should be Item (B); if there exists $m \in [n/2 - \sqrt{n}, n/2] \cap I_k$ and $m < j_k$, then we cannot pick $i_k < n/2 - \sqrt{n} \leq m$ since $|i_k - j_k| > |m - j_k|$ in Item (C). Now observe that there are at least $\sqrt{n}/2 - 1$ even numbers in $[n/2 - \sqrt{n}, n/2]$. Since we assumed $|J| < \sqrt{n}/9$, at least $7\sqrt{n}/18 - 1$ of the even numbers in $[n/2 - \sqrt{n}, n/2]$ must be in I and they are all selected before. By Claim 4.17, this means

$$\sum_{\ell \in [k-1]} \mathsf{wt}_{i_\ell} \geq \left(\frac{7\sqrt{n}}{18} - 1\right) \cdot \frac{\delta_1 \gamma \sqrt{\log(1/(\delta_1 \gamma))}}{16\sqrt{n}} \geq \frac{\delta_1 \gamma}{60},$$

which means that we should have terminated already. This gives a contradiction.

21

If $j_k < n/2 - \sqrt{n}$, then every even number $m \in [n/2 - \sqrt{n}, n/2]$ is either selected already or contained in J_k, since otherwise we would be in Item (B). Then the same argument follows. The possibility that $i_k > n/2 + \sqrt{n}$ or $j_k > n/2 + \sqrt{n}$ can also be ruled out analogously. $\qquad\square$

Given Claim 4.18, we always have $\sum_{\ell \in [k]} \mathsf{wt}_{i_\ell} \geq \delta_1 \gamma / 60$ upon termination. This, combined with (10), establishes the first half of Item 2 and completes the proof. $\qquad\square$

By identical reasoning, we can also handle odd numbers. In addition, by relaxing the bounds, we can remove 2^{-100} from the assumption to get a cleaner statement.

Corollary 4.19. *Assume $n \geq 2^{10}$ and $\delta_1, \delta_2 \in (0, 1)$ satisfying $\delta_1 \leq \delta_2$. Let $\Gamma \subset \{0, 1, \ldots, n\}$ (resp., $\Xi \subset \{0, 1, \ldots, n\}$) be non-empty and contain only even (resp., odd) numbers. If $\|\mathcal{D}_\Gamma - \mathsf{evens}\|_{\mathsf{TV}} \geq \delta_2$ (resp., $\|\mathcal{D}_\Xi - \mathsf{odds}\|_{\mathsf{TV}} \geq \delta_2$), then there exist $i_1, \ldots, i_K, j_1, \ldots, j_K$ such that*

1. *$i_1, \ldots, i_K, j_1, \ldots, j_K$ are distinct even (resp., odd) numbers in $\{0, 1, \ldots, n\}$,*
2. *$\mathbf{Pr}\left[|X| \in \{i_1, \ldots, i_K\}\right] \geq \delta_1 / 2^{106}$ and $\mathbf{Pr}\left[|X| \in \{j_1, \ldots, j_K\}\right] = 0$ where $X \sim \mathcal{D}_\Gamma$ (resp., $X \sim \mathcal{D}_\Xi$),*
3. *$K \leq 1 + 8\sqrt{n / \left(100 + \log\left(\frac{1}{\delta_1(1 - \delta_2)}\right)\right)}$ and $|i_\ell - j_\ell| \leq 4K$ holds for each $\ell \in [K]$.*

Proof. Simply apply Lemma 4.16 with $\delta_1' = \delta_1 \cdot 2^{-100}$ then $\delta_1'(1 - \delta_2) \leq \delta_1' \leq 2^{-100}$. $\qquad\square$

We are now in the position of proving Proposition 4.5.

Proof of Proposition 4.5. Let Γ and Ξ be the set of even and odd numbers in Ψ respectively. By Lemma 4.14,

$$\mathcal{D}_\Psi = \frac{\gamma}{\gamma + \xi} \cdot \mathcal{D}_\Gamma + \frac{\xi}{\gamma + \xi} \cdot \mathcal{D}_\Xi \quad \text{where} \quad \gamma = |\mathsf{supp}\left(\mathcal{D}_\Gamma\right)| / 2^{n-1} \quad \text{and} \quad \xi = |\mathsf{supp}\left(\mathcal{D}_\Xi\right)| / 2^{n-1}.$$

In addition, either

$$\frac{\gamma \cdot \|\mathcal{D}_\Gamma - \mathsf{evens}\|_{\mathsf{TV}}}{\gamma + \xi} \geq \frac{\delta}{12} \quad \text{where} \quad \|\mathcal{D}_\Gamma - \mathsf{evens}\|_{\mathsf{TV}} = 1 - \gamma, \tag{14}$$

or $\frac{\xi}{\gamma + \xi} \cdot \|\mathcal{D}_\Xi - \mathsf{odds}\|_{\mathsf{TV}} \geq \delta/12$.

Assume we are in the former case and the latter can be handled analogously. To construct the function ϕ, we apply Corollary 4.19 with $\delta_1 = \delta_2 = 1 - \gamma$ and obtain $i_1, \ldots, i_K, j_1, \ldots, j_K$. Then we define $\phi \colon \{0, 1\}^n \to \mathbb{R}$ by

$$\phi(x) = \begin{cases} 1 & |x| \in \{i_1, \ldots, i_K\}, \\ -C & |x| \in \{j_1, \ldots, j_K\}, \\ 0 & \text{otherwise,} \end{cases} \quad \text{where} \quad C = \exp\left\{2^{100dt^3} \cdot (n/r)^{4t^2}\right\}.$$

The range of ϕ is obviously from $-C$ to 1 and

$$\mathbb{E}_{X \sim \mathcal{D}_\Psi}[\phi(X)] = \frac{\gamma}{\gamma + \xi} \cdot \left(\mathbf{Pr}_{X \sim \mathcal{D}_\Gamma}\left[|X| \in \{i_1, \ldots, i_K\}\right] - C \cdot \mathbf{Pr}_{X \sim \mathcal{D}_\Gamma}\left[|X| \in \{j_1, \ldots, j_K\}\right]\right)$$

$$\text{(by Item 1 of Corollary 4.19)}$$

$$\geq \frac{\gamma}{\gamma + \xi} \cdot \left((1 - \gamma)/2^{106} - 0\right) \qquad \text{(by Item 2 of Corollary 4.19)}$$

22

$$= \frac{\gamma(1-\gamma)}{\gamma+\xi} \cdot 2^{-106}. \tag{15}$$

In addition, we have

$$K \leq 1 + 8\sqrt{n/\left(100 + \log\left(\frac{1}{\gamma(1-\gamma)}\right)\right)} \leq \sqrt{n}. \tag{16}$$

Now we turn to the (d, r, t)-local function g. Recall that by definition, it has at least r non-connected neighborhoods of size at most t. Set $\varepsilon = 2^{-3t-d-3}$ as the distance threshold for classifying Type-1 and Type-2 neighborhoods. If at least $r' = \lceil r/2 \rceil$ of them are Type-1, then we apply Lemma 4.12 and obtain

$$\|\mathcal{P}_g - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - 4n^{1.5} \cdot \exp\left\{-r/2^{6t+2d+10}\right\} - 8 \cdot \exp\left\{-n/2^{d+4}\right\} \geq 1 - n^2 \cdot \exp\left\{-r/2^{20dt}\right\},$$

where we used the fact that $d, t \geq 1$, $r \leq n$, and $n \geq 2^{10}$. Otherwise at least $r' = \lceil r/2 \rceil$ of them are Type-2. In this case, we evaluate the expectation of ϕ under $\mathcal{P}_g$ by first decomposing the Hamming weight distribution of $\mathcal{P}_g$ by Lemma 4.13, then applying Lemma 3.9 to each typical decomposed distribution. Formally, by the definition of ϕ, we first observe that

$$\mathop{\mathbb{E}}_{X \sim \mathcal{P}_g} [\phi(X)] = \mathop{\mathbf{Pr}}_{X \sim \mathcal{P}_g} [|X| \in \{i_1, \ldots, i_K\}] - C \cdot \mathop{\mathbf{Pr}}_{X \sim \mathcal{P}_g} [|X| \in \{j_1, \ldots, j_K\}]. \tag{17}$$

Then by Lemma 4.13, the Hamming weight distribution of $X \sim \mathcal{P}_g$ is $\sum_\rho \lambda_\rho \cdot \mathcal{P}_\rho$ where each $\mathcal{P}_\rho = X_\rho + \sum_{k \in [r']} X_{\rho,k}$ is a sum of independent random variables $X_{\rho,1}, \ldots, X_{\rho,r'}$ in $\{0, 1, \ldots, t\}$ together with a constant X_ρ. Recall that we say ρ is good if

$$\sum_{k \in [r']} \left(1 - \max_{x \in \mathbb{Z}} \mathbf{Pr}\left[X_{\rho,k} \equiv x \pmod{q}\right]\right) > 2^{-7t-d} \cdot r' \quad \text{holds for all integer } q \geq 3, \tag{18}$$

and bad otherwise. Then Lemma 4.13 also proves

$$\sum_{\text{bad } \rho} \lambda_\rho \leq t \cdot \exp\left\{-2^{-7t-d-2} \cdot r'\right\} \leq t \cdot \exp\left\{-2^{-20dt} \cdot r\right\}. \tag{19}$$

Since $C \geq 0$, by (17) and (19), we have

$$\mathop{\mathbb{E}}_{X \sim \mathcal{P}_g} [\phi(X)] \leq t \cdot \exp\left\{-2^{-20dt} \cdot r\right\} + \sum_{\text{good } \rho} \lambda_\rho \cdot \sum_{\ell \in [K]} G_{\rho,\ell}, \tag{20}$$

where

$$G_{\rho,\ell} = \mathbf{Pr}\left[\sum_{k \in [r']} X_{\rho,k} = i_\ell - X_\rho\right] - C \cdot \mathbf{Pr}\left[\sum_{k \in [r']} X_{\rho,k} = j_\ell - X_\rho\right].$$

Putting all the parameters into Lemma 3.9, we can show that each $G_{\rho,\ell}$ is small. This is formalized in Claim 4.20 and will be proved later.

Claim 4.20. For any good ρ and $\ell \in [K]$, we have $G_{\rho,\ell} \leq \frac{\gamma(1-\gamma)}{K \cdot (\gamma+\xi)} \cdot 2^{-200}$.

Putting Claim 4.20 into (20), we obtain

$$\mathop{\mathbb{E}}_{X \sim \mathcal{P}_g} [\phi(X)] \leq t \cdot \exp\left\{-r/2^{20dt}\right\} + K \cdot \frac{\gamma(1-\gamma)}{K \cdot (\gamma+\xi)} \cdot 2^{-200}.$$

23

Therefore

$$\underset{X \sim \mathcal{D}_\Psi}{\mathbb{E}}[\phi(X)] - \underset{X \sim \mathcal{P}_g}{\mathbb{E}}[\phi(X)] \geq \frac{\gamma(1-\gamma)}{\gamma+\xi} \cdot 2^{-110} - t \cdot \exp\left\{-r/2^{20dt}\right\} \qquad \text{(by (15))}$$

$$\geq \delta \cdot 2^{-120} - t \cdot \exp\left\{-r/2^{20dt}\right\} \qquad \text{(by (14))}$$

as desired. $\qquad\qquad\square$

Finally we prove Claim 4.20, which is a direct calculation based on Lemma 3.9.

Proof of Claim 4.20. Following the notation in Lemma 3.9, let $n = r'$, $x = i_\ell - X_\rho$, and $\Delta = |i_\ell - j_\ell|$, where Δ is an even number by Item 1 of Corollary 4.19 as needed by Lemma 3.9. By (18), we set $L = 2^{-10dt} \cdot r'$. Since $d, t \geq 1$, $r' \geq r/2$, and $r \geq 2^{100dt}$, we have

$$m = \left\lfloor \frac{2^{-10dt} \cdot r'}{32t^4} \right\rfloor \geq 2^{-30dt} \cdot r \quad \text{and} \quad \alpha = \left(\frac{2^{-10dt} \cdot r'}{4n(t+1)}\right)^{2t^2} \geq 2^{-30dt^3} \cdot (r/n)^{2t^2}.$$

Hence

$$\alpha \cdot m \geq 2^{-60dt^3} \cdot n \cdot (r/n)^{1+2t^2} \geq 2^{-60dt^3} \cdot n \cdot (r/n)^{3t^2}, \qquad (21)$$

where we used the fact that $r \leq n$ and $t \geq 1$. Let

$$\kappa_1 = \frac{\Delta^2}{4\alpha m},$$

which already satisfies $|\Delta| \leq 2\sqrt{\kappa_1 \cdot \alpha m}$ as demanded by Lemma 3.9. Now we expand Δ^2 and obtain a simpler formula to work with:

$$\Delta^2 \leq 16K^2 \qquad \text{(by Item 3 of Corollary 4.19)}$$

$$\leq 16 \cdot \left(1 + 8\sqrt{n/\left(100 + \log\left(\frac{1}{\gamma(1-\gamma)}\right)\right)}\right)^2 \qquad \text{(by (16))}$$

$$\leq 32 \cdot \left(1 + 64n/\left(100 + \log\left(\frac{1}{\gamma(1-\gamma)}\right)\right)\right) \qquad \text{(since } (a+b)^2 \leq 2 \cdot (a^2+b^2))$$

$$= 32 \cdot \left(1 + 64n/\left(100 + \log\left(\frac{1}{\gamma+\xi}\right) + \log\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right)\right)\right)$$

$$\leq 32 \cdot \left(1 + 64n/\left(99 + \log\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right)\right)\right) \qquad \text{(since } \gamma, \xi \in [0,1])$$

$$\leq 2^{15} \cdot \min\left\{n, 1 + n/\left(1 + \ln\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right)\right)\right\}, \qquad (22)$$

where for the last inequality we used the fact that $\ln\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right) \geq \ln\left(\frac{\gamma+\xi}{\gamma}\right) > 0$.

Obviously $\kappa_1 \geq 0$. Now we verify $\kappa_1 \leq \alpha \cdot m/128$ as follows:

$$\kappa_1 \cdot \frac{128}{\alpha \cdot m} = \frac{2^5 \Delta^2}{\alpha^2 m^2} \leq \frac{2^{15} n}{\alpha^2 m^2} \leq \frac{2^{15} \cdot n}{2^{-120dt^3} \cdot n^2 \cdot (r/n)^{6t^2}} \qquad \text{(by (22) and (21))}$$

$$\leq \frac{2^{180dt^3} \cdot (n/r)^{6t^2}}{n} \leq 1. \qquad \text{(since } d, t \geq 1 \text{ and } n \geq \left(2^{30dt} \cdot n/r\right)^{6t^2})$$

24

Define
$$\kappa_2 = \left(1 + \ln\left(\frac{\gamma + \xi}{\gamma(1-\gamma)}\right)\right) \cdot 2^{10dt^3} \cdot (n/r)^{t^2},$$

which is at least 1 since $n \geq r$ and $\ln\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right) \geq \ln\left(\frac{\gamma+\xi}{\gamma}\right) > 0$. Now we verify $\kappa_2 \leq \alpha \cdot m/128$:

$$
\begin{aligned}
\kappa_2 \cdot \frac{128}{\alpha \cdot m} &\leq \frac{2^7 \cdot \left(1 + \ln\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right)\right) \cdot 2^{10dt^3} \cdot (n/r)^{t^2}}{2^{-60dt^3} \cdot n \cdot (r/n)^{3t^2}} && \text{(by (21))} \\
&\leq \frac{2^7 \cdot (1 + \ln(4/\delta)) \cdot 2^{10dt^3} \cdot (n/r)^{t^2}}{2^{-60dt^3} \cdot n \cdot (r/n)^{3t^2}} && \text{(by (14))} \\
&\leq \frac{(1 + \log(1/\delta)) \cdot 2^{20dt^3} \cdot (n/r)^{t^2}}{2^{-60dt^3} \cdot n \cdot (r/n)^{3t^2}} && \text{(since } d, t \geq 1) \\
&\leq 1. && \text{(since } n \geq r \text{ and } \log(1/\delta) \leq n \cdot \left(2^{30dt} \cdot n/r\right)^{-3t^2})
\end{aligned}
$$

To verify $C \geq 2 \cdot e^{12 \cdot (\sqrt{\kappa_1 \kappa_2} + \kappa_1)}$, we first bound $\kappa_1 \kappa_2$:

$$
\begin{aligned}
\kappa_1 \kappa_2 = \frac{\Delta^2 \kappa_2}{4\alpha m} &\leq \frac{2^{13} \cdot \left(1 + n/\left(1 + \ln\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right)\right)\right) \cdot \left(1 + \ln\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right)\right) \cdot 2^{10dt^3} \cdot (n/r)^{t^2}}{2^{-60dt^3} \cdot n \cdot (r/n)^{3t^2}} \\
&\hspace{8cm} \text{(by (22) and (21))} \\
&= \frac{2^{13} \cdot \left(1 + n + \ln\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right)\right)}{2^{-70dt^3} \cdot n \cdot (r/n)^{4t^2}} \leq \frac{2^{13} \cdot (1 + n + \ln(4/\delta))}{2^{-70dt^3} \cdot n \cdot (r/n)^{4t^2}} && \text{(by (14))} \\
&\leq \frac{2^{20} \cdot n}{2^{-70dt^3} \cdot n \cdot (r/n)^{4t^2}} && \text{(since } \log(1/\delta) \leq n \cdot \left(2^{30dt} \cdot n/r\right)^{-3t^2} \leq n) \\
&= 2^{20} \cdot 2^{70dt^3} \cdot (n/r)^{4t^2}. && (23)
\end{aligned}
$$

Since $\kappa_2 \geq 1$, we also have $\kappa_1 \leq \kappa_1 \kappa_2 \leq 2^{20} \cdot 2^{70dt^3} \cdot (n/r)^{4t^2}$ by (23). Therefore

$$2 \cdot e^{12 \cdot (\sqrt{\kappa_1 \kappa_2} + \kappa_1)} \leq 2 \cdot \exp\left\{12 \cdot 2 \cdot 2^{20} \cdot 2^{70dt^3} \cdot (n/r)^{4t^2}\right\} \leq \exp\left\{2^{100dt^3} \cdot (n/r)^{4t^2}\right\} = C$$

as desired. Finally we conclude

$$
\begin{aligned}
G_{\rho, \ell} \cdot K &\leq \sqrt{\frac{32}{\alpha \cdot m}} \cdot K \cdot e^{-2\kappa_2} && \text{(by Lemma 3.9)} \\
&\leq \sqrt{\frac{32}{2^{-60dt^3} \cdot n \cdot (r/n)^{3t^2}}} \cdot \sqrt{n} \cdot \exp\left\{-2 \cdot \left(1 + \ln\left(\frac{\gamma+\xi}{\gamma(1-\gamma)}\right)\right) \cdot 2^{10dt^3} \cdot (n/r)^{t^2}\right\} \\
&\hspace{9cm} \text{(by (16) and (21))} \\
&\leq \frac{\gamma(1-\gamma)}{\gamma+\xi} \cdot \sqrt{\frac{32}{2^{-60dt^3} \cdot (r/n)^{3t^2}}} \cdot \exp\left\{-2 \cdot 2^{10dt^3} \cdot (n/r)^{t^2}\right\} \\
&\leq \frac{\gamma(1-\gamma)}{\gamma+\xi} \cdot \sqrt{\frac{32}{2^{-60dt^3} \cdot (r/n)^{3t^2}}} \cdot \exp\left\{-300dt^3 - 2t^2 \ln(n/r)\right\} \quad \text{(since } d, t \geq 1 \text{ and } n \geq r) \\
&\leq \frac{\gamma(1-\gamma)}{\gamma+\xi} \cdot 2^{40dt^3} \cdot e^{-300dt^3} \leq \frac{\gamma(1-\gamma)}{\gamma+\xi} \cdot 2^{-200} && \text{(since } d, t \geq 1)
\end{aligned}
$$

as claimed. $\qquad\square$

4.1.2 Extremely Small Error Regime

In this section, we handle the case when $f(\mathcal{U}^m)$ is extremely close to $\mathcal{D}_\Psi$ and prove Theorem 4.9.

Theorem (Theorem 4.9 Restated). *Let* $f \colon \{0,1\}^m \to \{0,1\}^n$ *be a d-local function where* $n \geq \text{tow}(900(d+1))$. *Assume* $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq \varepsilon$ *for some* Ψ *in the central regime and*

$$\varepsilon \leq \exp\left\{ -\frac{n}{\text{tow}(800(d+1))} \right\}.$$

Then

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq \text{tow}(850(d+1)) \cdot \varepsilon$$

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

Given the assumption in Theorem 4.9, Theorem 4.8 guarantees that $f(\mathcal{U}^m)$ is already exponentially close to $\texttt{evens}$, $\texttt{odds}$, or $\texttt{all}$. Since f is d-local and thus every output bit is of granularity 2^{-d}, this implies that the distribution of any small number of output bits is *exactly* uniform. For our purposes, we only need to prove it for every pair of two as in the following Lemma 4.21, which naturally generalizes.

Lemma 4.21. *Let* $f \colon \{0,1\}^m \to \{0,1\}^n$ *be a d-local function where* $n \geq 3$. *Let* $X \sim f(\mathcal{U}^m)$ *and assume* $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq 2^{-2d-1}$ *for some* $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}, \texttt{all}\}$. *Then* $\mathbb{E}[X_i] = 1/2$ *and* $\mathbb{E}[X_i X_j] = 1/4$ *hold for any distinct* $i, j \in [n]$.

Proof. Assume without loss of generality $d \geq 1$, since otherwise $f(\mathcal{U}^m)$ is a point distribution and its distance to $\mathcal{D}$ is at least $1 - 2^{-(n-1)} > 1/2$.

Since f is d-local, the probability density function of X_i has granularity 2^{-d}. Hence $\mathbb{E}[X_i]$ is a multiple of 2^{-d}. Since $n \geq 2$, the marginal distribution of the i-th coordinate of $\mathcal{D}$ is unbiased. If $\mathbb{E}[X_i] \neq 1/2$, then $|\mathbb{E}[X_i] - 1/2| \geq 2^{-d}$, which, by Fact 3.1, implies $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \geq 2^{-d}$ and contradicts the assumption.

Similarly, $\mathbb{E}[X_i X_j]$ is a multiple of 2^{-2d} and the joint distribution of the i-th and j-th coordinates of $\mathcal{D}$ is unbiased since $n \geq 3$. If $\mathbb{E}[X_i X_j] \neq 1/4$, then $|\mathbb{E}[X_i X_j] - 1/4| \geq 2^{-2d}$, which implies $\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \geq 2^{-2d}$ and contradicts the assumption. $\qquad\square$

We will also need the following lemma, which shows that the support of $f(\mathcal{U}^m)$ has a consistent parity provided that it is sufficiently close to $\texttt{evens}$ or $\texttt{odds}$.

Lemma 4.22. *Let* $f \colon \{0,1\}^m \to \{0,1\}^n$ *be a d-local function. Assume for some* $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}\}$ *we have*

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} < 2^{-d}. \tag{24}$$

Then $\mathsf{supp}\,(f(\mathcal{U}^m)) \subseteq \mathsf{supp}\,(\mathcal{D})$.

Proof. Assume $\mathcal{D} = \texttt{evens}$ and the argument is identical for the $\mathcal{D} = \texttt{odds}$ case. Define function $g \colon \{0,1\}^m \to \{0,1\}$ as the parity of f's output, i.e., $g(x) = |f(x)| \pmod 2$ where $|f(x)|$ is the Hamming weight of $f(x)$. To show $\mathsf{supp}\,(f(\mathcal{U}^m)) \subseteq \mathsf{supp}\,(\mathcal{D})$, it suffices to show that g is the constant zero function.

Assume towards a contradiction that g is non-zero, and assume without loss of generality that $x_1 x_2 \cdots x_\ell$ is a maximal monomial in g. Note that we consider the constant 1 as a monomial with degree 0. Since f is d-local, g has degree d over $\mathbb{F}_2$. This means $\ell \leq d$. Hence

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \geq \Pr_{x \sim \mathcal{U}^m} [f(x) \text{ has odd Hamming weight}] \qquad \text{(Fact 3.1)}$$

$$\begin{aligned}
&= \mathop{\mathbb{E}}_{x \sim \mathcal{U}^m}[g(x)] && \text{(by the definition of } g) \\
&= \mathop{\mathbb{E}}_{x_{\ell+1},\dots,x_m}\left[\mathop{\mathbb{E}}_{x_1,\dots,x_\ell}[g(x) \mid x_{\ell+1},\dots,x_m]\right] \\
&\geq \mathop{\mathbb{E}}_{x_{\ell+1},\dots,x_m}\left[2^{-\ell}\right] && \text{(since } g(x) \text{ is non-zero degree } \ell \text{ conditioned on } x_{\ell+1},\dots,x_m) \\
&= 2^{-d},
\end{aligned}$$

which contradicts (24). $\qquad\square$

Now we prove Theorem 4.9.

Proof of Theorem 4.9. Note that the strong distance bound implies a weaker distance bound:

$$\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \leq \varepsilon \leq \exp\left\{-\frac{n}{\mathrm{tow}(750(d+1))}\right\}.$$

Then we can apply Theorem 4.8 and obtain a primitive bound

$$\|f(\mathcal{U}^m) - \mathcal{D}\|_{\mathsf{TV}} \leq \frac{\mathrm{tow}(750(d+1))}{\exp\{n/\mathrm{tow}(750(d+1))\}} \leq \exp\left\{-\frac{n}{\mathrm{tow}(800(d+1))}\right\} \tag{25}$$

for some $\mathcal{D} \in \{\mathtt{evens}, \mathtt{odds}, \mathtt{all}\}$.

Define a potential function $\phi \colon \{0,1\}^n \to \mathbb{R}$ by

$$\phi(x) = \left(|x| - \frac{n}{2}\right)^2 = \frac{n^2}{4} + (1-n) \cdot \sum_{i \in [n]} x_i + 2 \cdot \sum_{1 \leq i < j \leq n} x_i x_j.$$

Now we do a case analysis and boost the bound in (25).

The $\mathcal{D} = \mathtt{all}$ Case. In this setting, we will use $\mathbb{E}_{X \sim \mathcal{D}_\Psi}[\phi(X)]$ as an intermediate to establish Theorem 4.9. First we present a lower bound:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X \sim \mathcal{D}_\Psi}[\phi(X)] &\geq \mathop{\mathbb{E}}_{X \sim f(\mathcal{U}^m)}[\phi(X)] - \frac{n^2}{4} \cdot \|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} && \text{(since } 0 \leq \phi \leq n^2/4) \\
&= \frac{n}{4} - \frac{n^2}{4} \cdot \|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} && \text{(by Lemma 4.21 and (25))} \\
&\geq \frac{n}{4} - \frac{n^2}{4} \cdot \varepsilon. \tag{26}
\end{aligned}$$

To derive an upper bound, we first use triangle inequality to obtain

$$\|\mathcal{D}_\Psi - \mathtt{all}\|_{\mathsf{TV}} \leq \varepsilon + \frac{\mathrm{tow}(750(d+1))}{\exp\{n/\mathrm{tow}(750(d+1))\}} \leq \exp\left\{-\frac{n}{\mathrm{tow}(800(d+1))}\right\}. \tag{27}$$

Since $\mathcal{D}_\Psi$ is $\mathtt{all}$ conditioned on the Hamming weight being in Ψ, by Fact 3.2 and (27), we have

$$\|\mathcal{D}_\Psi - \mathtt{all}\|_{\mathsf{TV}} = \sum_{m \notin \Psi} 2^{-n} \cdot \binom{n}{m} \leq \exp\left\{-\frac{n}{\mathrm{tow}(800(d+1))}\right\}.$$

This implies that $|m - n/2| \geq n/\mathrm{tow}(800(d+1))$ for every $m \notin \Psi$, since otherwise we have

$$\exp\left\{-\frac{n}{\mathrm{tow}(800(d+1))}\right\} \geq 2^{-n} \cdot \binom{n}{n/2 - n/\mathrm{tow}(800(d+1))}$$

$$\geq \frac{1}{\sqrt{8n}} \cdot 2^{n \cdot (1 - \mathcal{H}(1/2 - 1/\text{tow}(800(d+1))))} \qquad \text{(by Fact 3.7)}$$

$$\geq \frac{1}{\sqrt{8n}} \cdot 2^{-4n/\text{tow}(800(d+1))^2} > \exp\left\{-\frac{n}{\text{tow}(800(d+1))}\right\},$$

which is a contradiction. Therefore, we get an upper bound estimate:

$$\mathop{\mathbb{E}}_{X \sim \mathcal{D}_\Psi}[\phi(X)] = \sum_{m \in \Psi} \left(m - \frac{n}{2}\right)^2 \cdot \frac{\binom{n}{m}}{|\text{supp}(\mathcal{D}_\Psi)|} \qquad \text{(by the definition of } \mathcal{D}_\Psi)$$

$$= \frac{2^n}{|\text{supp}(\mathcal{D}_\Psi)|} \cdot \left(\sum_m \left(m - \frac{n}{2}\right)^2 \cdot \frac{\binom{n}{m}}{2^n} - \sum_{m \notin \Psi} \left(m - \frac{n}{2}\right)^2 \cdot \frac{\binom{n}{m}}{2^n}\right)$$

$$\leq \frac{2^n}{|\text{supp}(\mathcal{D}_\Psi)|} \cdot \left(\sum_m \left(m - \frac{n}{2}\right)^2 \cdot \frac{\binom{n}{m}}{2^n} - \frac{n^2}{\text{tow}(800(d+1))^2} \sum_{m \notin \Psi} \frac{\binom{n}{m}}{2^n}\right)$$

$$= \frac{1}{1 - \|\mathcal{D}_\Psi - \texttt{all}\|_{\text{TV}}} \cdot \left(\frac{n}{4} - \frac{n^2}{\text{tow}(800(d+1))^2} \cdot \|\mathcal{D}_\Psi - \texttt{all}\|_{\text{TV}}\right).$$

Combining this with (26) and $n \geq \text{tow}(900(d+1))$, we have

$$\|\mathcal{D}_\Psi - \texttt{all}\|_{\text{TV}} \leq \frac{\|\mathcal{D}_\Psi - \texttt{all}\|_{\text{TV}}}{1 - \|\mathcal{D}_\Psi - \texttt{all}\|_{\text{TV}}} \leq \frac{n^2 \varepsilon/4}{n^2/\text{tow}(800(d+1))^2 - n/4} \leq \varepsilon \cdot \text{tow}(850(d+1))/4,$$

which implies

$$\|f(\mathcal{U}^m) - \texttt{all}\|_{\text{TV}} \leq \varepsilon + \varepsilon \cdot \text{tow}(850(d+1))/4 \leq \text{tow}(850(d+1)) \cdot \varepsilon$$

as desired.

The $\mathcal{D} \in \{\texttt{evens}, \texttt{odds}\}$ Case. We prove for the evens case and the other one is analogous. By Lemma 4.22 and (25), we know that $\text{supp}(f(\mathcal{U}^m)) \subseteq \text{supp}(\texttt{evens})$. Let $\Gamma \subseteq \Psi$ be the even numbers in Ψ. Then both $\text{supp}(f(\mathcal{U}^m))$ and $\text{supp}(\mathcal{D}_\Gamma)$ are subsets of $\text{supp}(\texttt{evens})$. We will use $\mathbb{E}_{X \sim \mathcal{D}_\Gamma}[\phi(X)]$ as the intermediate.

By Fact 3.1, we have $\|f(\mathcal{U}^m) - \mathcal{D}_\Gamma\|_{\text{TV}} \leq \|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\text{TV}} \leq \varepsilon$ and hence the lower bound calculation in (26) works for Γ as well:

$$\mathop{\mathbb{E}}_{X \sim \mathcal{D}_\Gamma}[\phi(X)] \geq \frac{n}{4} - \frac{n^2}{4} \cdot \varepsilon.$$

Then similar to the $\mathcal{D} = \texttt{all}$ case, we have an analogous (27):

$$\|\mathcal{D}_\Gamma - \texttt{evens}\|_{\text{TV}} \leq \exp\left\{-\frac{n}{\text{tow}(800(d+1))}\right\},$$

which, combined with Fact 3.2, implies

$$\|\mathcal{D}_\Gamma - \texttt{evens}\|_{\text{TV}} = \sum_{\text{even } m \notin \Gamma} 2^{-n+1} \cdot \binom{n}{m} \leq \exp\left\{-\frac{n}{\text{tow}(800(d+1))}\right\}.$$

An almost identical calculation using Fact 3.7 shows that $|m - n/2| \geq n/\text{tow}(800(d+1))$ for every even $m \notin \Gamma$. Hence we have a similar upper bound estimate

$$\mathop{\mathbb{E}}_{X \sim \mathcal{D}_\Gamma}[\phi(X)] \leq \frac{1}{1 - \|\mathcal{D}_\Gamma - \texttt{evens}\|_{\text{TV}}} \cdot \left(\frac{n}{4} - \frac{n^2}{\text{tow}(800(d+1))^2} \cdot \|\mathcal{D}_\Gamma - \texttt{evens}\|_{\text{TV}}\right),$$

which also implies the final bound by rearranging. $\qquad\square$

Acknowledgments

References

[ALWZ21] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795–815, 2021.

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

[BCW21] Tolson Bell, Suchakree Chueluecha, and Lutz Warnke. Note on sunflowers. *Discrete Mathematics*, 344(7):112367, 2021.

[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.

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

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

[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.

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

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence, 2013.

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

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

[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.

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

[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.

[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.

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

[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.

[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.

[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.

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

[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.

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

[Hoe94] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.

[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.

[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.

[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.

[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).

[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.

[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.

[Pet12] Valentin V Petrov. *Sums of independent random variables*, volume 82. Springer Science & Business Media, 2012.

[Rao19] Anup Rao. Coding for sunflowers. *arXiv preprint arXiv:1909.04774*, 2019.

[Ros14] Benjamin Rossman. The monotone complexity of k-clique on random graphs. *SIAM Journal on Computing*, 43(1):256–279, 2014.

[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.

[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.

[SW22] Zbigniew Szewczak and Michel Weber. Classical and almost sure local limit theorems. *arXiv preprint arXiv:2208.02700*, 2022.

[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.

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

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

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

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

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

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

[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.

[Wik23a] Wikipedia. Binary entropy function — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Binary%20entropy%20function&oldid=107150795 2023. [Online; accessed 04-December-2023].

[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].

[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].

[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.

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

A Density Comparison of Sum of Integral Random Variables

The goal of this section is to prove Theorem A.1, which generalizes Lemma 3.9.

Theorem A.1. *Let $t \geq 1$ 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 $r \geq 1$, define $p_{r,i} = \max_{x \in \mathbb{Z}} \mathbf{Pr}\left[X_i \equiv x \pmod{r}\right]$ and assume[6]

$$\sum_{i \in [n]} (1 - p_{r,i}) \geq L > 0 \quad \text{holds for all } r \in \Phi. \tag{28}$$

Let $m = \lfloor L/(16t^4\phi) \rfloor$ and $\alpha = \left(\frac{L}{4n(t+1)}\right)^{t^2\phi}$. Then for any $x \in \mathbb{Z}$ and $0 \leq \kappa_1, \kappa_2 \leq \alpha \cdot m/128$,

$$\mathbf{Pr}\left[\sum_{i \in [n]} X_i = x\right] - C \cdot \mathbf{Pr}\left[\sum_{i \in [n]} X_i = x + \Delta\right] \leq \sqrt{\frac{32}{\alpha \cdot m}} \cdot e^{-2\kappa_2}$$

holds for any $\Delta \in \mathbb{Z}$ and $C \in \mathbb{R}$ satisfying

$$|\Delta| \leq \phi\sqrt{\kappa_1 \cdot \alpha m} \text{ is a multiple of } \phi \quad \text{and} \quad C \geq 2 \cdot e^{12 \cdot (\sqrt{\kappa_1 \kappa_2} + \kappa_1)}.$$

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. We emphasize that Theorem A.1 is qualitatively tight in various aspects.

---

[6]Note that if (28) holds for some r, then it also holds for r' that is a multiple of r as $\mathbf{Pr}\left[X_i \equiv x \pmod{r'}\right] \leq \mathbf{Pr}\left[X_i \equiv x \pmod{r}\right]$. Hence we may assume that Φ contains all the multiples of r (up to t) if $r \in \Phi$.

- The assumption of Δ being a multiple of ϕ is necessary. If X_i's have some joint congruence relation which does not share with Δ, the bound can fail. Consider the case where n is even and each X_i is uniform in $\{1, 3\}$, which violates (28) 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 $\mathbf{Pr}\left[\sum_i X_i = x + \Delta\right] = 0$. Hence the final bound does not hold unless κ_2 is only constant, in which case we should use Littlewood-Offord-type anticoncentration results (e.g. [Ush86, Theorem 3]).

- The assumption on κ_1, κ_2 is necessary. Consider the case where each X_i is an unbiased coin.

  - If $\kappa_1 \gg n$, we can set $x = n/2$ and $\Delta = n$. Then $\mathbf{Pr}\left[\sum_i X_i = x\right] \approx 1/\sqrt{n}$ but $\mathbf{Pr}\left[\sum_i X_i = x + \Delta\right] = 0$. Hence the final bound does not hold unless κ_2 is constant, in which case we should use standard anticoncentration results (e.g. [Ush86, Theorem 3]).
  - If $\kappa_2 \gg n$, we can set $x = n$ and $\Delta = 1$. Then $\mathbf{Pr}\left[\sum_i X_i = x\right] = 2^{-n} \gg 2^{-\kappa_2}$ but $\mathbf{Pr}\left[\sum_i X_i = x + \Delta\right] = 0$. Hence the final bound simply cannot hold.

- The trade-off between C and the final bound is essentially optimal. Consider the case where each X_i is an unbiased coin.

  - If $\kappa_1 \gg \kappa_2$, we set $x = n/2$ and $\Delta = \sqrt{\kappa_1 n}$. Then $\mathbf{Pr}\left[\sum_i X_i = x\right] \approx 1/\sqrt{n}$ and $\mathbf{Pr}\left[\sum_i X_i = x + \Delta\right] \approx 2^{-\kappa_1}/\sqrt{n}$, which means the final bound holds only if $C \gg 2^{\kappa_1}$.
  - If $\kappa_2 \gg \kappa_1$, we set $x = n/2 + \sqrt{\kappa_2 n} - \sqrt{\kappa_1 n}$ and $\Delta = \sqrt{\kappa_1 n}$. Then $\mathbf{Pr}\left[\sum_i X_i = x\right] \approx 2^{-\kappa_2 + \sqrt{\kappa_1 \kappa_2}}/\sqrt{n}$ and $\mathbf{Pr}\left[\sum_i X_i = x + \Delta\right] \approx 2^{-\kappa_2}/\sqrt{n}$, which means the final bound holds only if $C \gg 2^{\sqrt{\kappa_1 \kappa_2}}$.

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 use the following standard concentration inequalities.

Fact A.3 (Hoeffding's Inequality). *Assume $X_1, \ldots, X_n$ are independent random variables such that $a \leq X_i \leq b$ holds for all $i \in [n]$. Then for all $\delta \geq 0$, we have*

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

Fact A.4 (Chernoff's Inequality). *Assume $X_1, \ldots, X_n$ 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*

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

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 itself 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 above quantity is in fact negative given the presence of C.

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.5. *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*

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

Then for any $y \in \mathbb{Z}$ *and* $0 \leq \kappa_1, \kappa_2 \leq \alpha \cdot m / 128$,

$$\mathbf{Pr}\left[\sum_{i \in [m]} Y_i = y\right] - C \cdot \mathbf{Pr}\left[\sum_{i \in [m]} Y_i = y + \Delta\right] \leq \sqrt{\frac{32}{\alpha \cdot m}} \cdot e^{-2\kappa_2} \tag{29}$$

holds for any $\Delta \in \mathbb{Z}$ *and* $C \in \mathbb{R}$ *satisfying*

$$|\Delta| \leq \phi \sqrt{\kappa_1 \cdot \alpha m} \text{ is a multiple of } \phi \quad and \quad C \geq 2 \cdot e^{12 \cdot (\sqrt{\kappa_1 \kappa_2} + \kappa_1)}.$$

Proof. If $\Delta < 0$, then we work with negated Y_i's. Hence we assume $\Delta \geq 0$. 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 A.4 with $\delta = 1/2$ and $\mu = \alpha \cdot m$, we have

$$\mathbf{Pr}[\mathcal{E}] \leq e^{-\alpha \cdot m / 8}. \tag{30}$$

For fixed $W = (W_1, \ldots, W_m)$ that $\mathcal{E}$ 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 (29) equals

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

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 $\Delta' = \Delta/\phi$. Then the above quantity equals $2^{-k} \cdot \left(\binom{k}{b'} - C \cdot \binom{k}{b' + \Delta'}\right)$. We will show that

$$\binom{k}{b'} - C \cdot \binom{k}{b' + \Delta'} \leq \frac{2}{\sqrt{k}} \cdot 2^k \cdot e^{-2\kappa_2}, \tag{31}$$

which, combined with (30), establishes (29):

$$\text{LHS of (29)} \leq \mathbf{Pr}[\mathcal{E}] + \mathbb{E}\left[\mathbb{1}_{\phi \text{ divides } b} \cdot 2^{-k} \cdot \left(\binom{k}{b'} - C \cdot \binom{k}{b' + \Delta'}\right)\,\middle|\, \neg\mathcal{E}\right]$$

$$\leq e^{-\alpha \cdot m / 8} + \mathbb{E}\left[\frac{2}{\sqrt{k}} \cdot e^{-2\kappa_2} \,\middle|\, \neg\mathcal{E}\right] \qquad \text{(by (30) and (31))}$$

$$\leq e^{-\alpha \cdot m / 8} + \sqrt{\frac{8}{\alpha \cdot m}} \cdot e^{-2\kappa_2} \qquad \text{(since } k \geq \alpha \cdot m / 2\text{)}$$

34

$$\leq \sqrt{\frac{32}{\alpha \cdot m}} \cdot e^{-2\kappa_2} \qquad\qquad\qquad \text{(since } \kappa_2 \leq \frac{\alpha \cdot m}{128})$$
$$= \text{RHS of (29)}.$$

To prove (31), we first observe that our assumption on κ_1, κ_2 guarantees that $\kappa_1, \kappa_2 \leq k/64$. Then we divide into the following cases:

- If $b' \leq 0$, then LHS of (31) ≤ 1, which is smaller than the RHS of (31) due to $\kappa_2 \leq k/64$.
- If $1 \leq b' \leq k/2 - \Delta'/2$, then $\binom{k}{b'} \leq \binom{k}{b'+\Delta'}$ and thus (31) holds due to $C \geq 1$.
- If $k/2 - \Delta'/2 \leq b' \leq k/2 + \sqrt{\kappa_2 \cdot k}$, then let $\delta_1 = b/k$ and $\delta_2 = (b+\Delta')/k$. Define $x_1 = 2\delta_1 - 1$ and $x_2 = 2\delta_2 - 1$. Then

$$-\frac{1}{4} \leq -\frac{\Delta'}{k} \leq x_1 \leq 2\sqrt{\frac{\kappa_2}{k}} \leq \frac{1}{4} \quad \text{and} \quad 0 \leq x_1 + \frac{2\Delta'}{k} = x_2 \leq \frac{3}{4}, \qquad (32)$$

where we used the fact that $0 \leq \Delta' \leq \sqrt{\kappa_1 \cdot \alpha m} \leq \sqrt{2\kappa_1 \cdot k} \leq k/4$ and $\sqrt{\kappa_2 \cdot k} \leq k/8$. Hence $\frac{3}{8} \leq \delta_1 \leq \frac{5}{8}$, $\frac{1}{2} \leq \delta_2 \leq \frac{7}{8}$, and

$$\text{LHS of (31)} \leq \frac{2^{k \cdot \mathcal{H}(\delta_1)}}{\sqrt{\pi k \cdot \delta_1 (1-\delta_1)}} - C \cdot \frac{2^{k \cdot \mathcal{H}(\delta_2)}}{\sqrt{8k \cdot \delta_2 (1-\delta_2)}} \qquad \text{(by Fact 3.7)}$$

$$\leq \frac{2^{k \cdot \mathcal{H}(\delta_1)}}{\sqrt{\pi k \cdot 15/64}} - C \cdot \frac{2^{k \cdot \mathcal{H}(\delta_2)}}{\sqrt{8k \cdot 1/4}} \qquad \text{(by (32))}$$

$$\leq \frac{2^{k \cdot \mathcal{H}(\delta_1)}}{\sqrt{\pi k \cdot 15/64}} \cdot \left(1 - \frac{C}{2} \cdot 2^{k \cdot (\mathcal{H}(\delta_2) - \mathcal{H}(\delta_1))}\right). \qquad (33)$$

Since $\delta_1 = \frac{1+x_1}{2}$ and $\delta_2 = \frac{1+x_2}{2}$, then by Fact 3.8, we have

$$\mathcal{H}(\delta_2) - \mathcal{H}(\delta_1) = -\frac{1}{2\ln(2)} \sum_{n=1}^{+\infty} \frac{x_2^{2n} - x_1^{2n}}{n \cdot (2n-1)} = -\frac{x_2^2 - x_1^2}{2\ln(2)} \sum_{n=1}^{+\infty} \sum_{i=0}^{n-1} \frac{x_1^{2i} x_2^{2(n-1-i)}}{n \cdot (2n-1)}$$

$$= -\frac{2\Delta' \cdot (x_1 + \Delta'/k)}{k \ln(2)} \sum_{n=1}^{+\infty} \sum_{i=0}^{n-1} \frac{x_1^{2i} x_2^{2(n-1-i)}}{n \cdot (2n-1)} \qquad \text{(since } x_2 = x_1 + 2\Delta/k)$$

$$\geq -\frac{2\Delta' \cdot (2\sqrt{\kappa_2/k} + \Delta'/k)}{k \ln(2)} \sum_{n=1}^{+\infty} \sum_{i=0}^{n-1} \frac{(1/16)^{2i}(3/4)^{2(n-1-i)}}{n \cdot (2n-1)} \qquad \text{(by (32))}$$

$$\geq -\frac{4\Delta' \cdot (2\sqrt{\kappa_2/k} + \Delta'/k)}{k \ln(2)}$$

$$\geq -\frac{4\sqrt{2\kappa_1} \cdot (2\sqrt{\kappa_2} + \sqrt{2\kappa_1})}{k \ln(2)} \qquad \text{(since } \Delta' \leq \sqrt{2\kappa_1 \cdot k})$$

$$\geq -\frac{12 \cdot (\sqrt{\kappa_1 \kappa_2} + \kappa_1)}{k \ln(2)}.$$

Putting this back to (33), we have

$$2^{k \cdot (\mathcal{H}(\delta_2) - \mathcal{H}(\delta_1))} \geq e^{-12 \cdot (\sqrt{\kappa_1 \kappa_2} + \kappa_1)} \geq 2/C,$$

which implies that the LHS of (31) is at most zero.

- If $b' \geq k/2 + \sqrt{\kappa_2 \cdot k}$, then let $x = 2\sqrt{\kappa_2/k}$. Since $\sqrt{\kappa_2 \cdot k} \leq k/8$, we have $0 \leq x \leq 1/4$ and

$$\text{LHS of } (31) \leq \binom{k}{b'} \leq \binom{k}{(1+x)\cdot k/2} \leq \frac{2^{k\cdot\mathcal{H}((1+x)/2)}}{\sqrt{\pi k \cdot (1-x^2)/4}} \qquad \text{(by Fact 3.7)}$$

$$\leq \frac{2}{\sqrt{k}} \cdot 2^{k\cdot\mathcal{H}((1+x)/2)} \leq \frac{2}{\sqrt{k}} \cdot 2^k \cdot e^{-k\cdot x^2/2} \qquad \text{(by Fact 3.8)}$$

$$= \frac{2}{\sqrt{k}} \cdot 2^k \cdot e^{-2\kappa_2}. \qquad\qquad\qquad\qquad \square$$

We remark that it is possible to prove Lemma A.5 by a local limit theorem for unbiased coins, removing the case analysis and explicit estimates on binomial coefficients. However, standard local limit results (e.g., [Pet12, SW22]) have a uniform error term which will be scaled by C in the LHS of (29), this makes the RHS of (29) dependent also on C. To avoid such a complication, we choose to work on the binomials directly.

Now we prove Theorem A.1 by reducing it to Lemma A.5. 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 $k = |\Phi| \leq t$. For each r_j, let $S_{r_j} \subseteq [n]$ be the set of X_i's with $1 - p_{r_j,i} \geq L/(2n)$, i.e.,

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

Since $0 \leq 1 - p_{r_j,i} \leq 1$ and by (28), we have

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

which implies $|S_{r_j}| \geq L/2$. 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 $k \leq t$ and originally $|S_{r_j}| \geq L/2$, each S_{r_j} contains enough number of elements when we keep only n' of them.

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 $\mathbf{Pr}\left[X_i \equiv c_i \pmod{r_j}\right] \geq \frac{1}{r_j}$. 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

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

where we used the fact that $0 < L \leq n$. Since $i \in S_{r_j}$, we also have $\mathbf{Pr}\left[X_i \equiv c_i \pmod{r_j}\right] \leq 1 - \frac{L}{2n}$ and hence, by an averaging argument, there exists $c_i' \in \mathbb{Z}_{r_j}/\mathbb{Z}$ such that $c_i' \neq c_i$ and $\mathbf{Pr}\left[X_i \equiv c_i' \pmod{r_j}\right] \geq \frac{L}{2n\cdot(r_j-1)}$. 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

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

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

36

1. $z_{r_j}, z'_{r_j} \in \{0, 1, \ldots, t\}$ and $z_{r_j} \not\equiv z'_{r_j} \pmod{r_j}$,

2. at least $1/\binom{t+1}{2} \geq 1/t^2$ fraction of $i \in S_{r_j}$ satisfies $\mathbf{Pr}\left[X_i = z_{r_j}\right], \mathbf{Pr}\left[X_i = z'_{r_j}\right] \geq \frac{L}{4n(t+1)}$.

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 satisfy 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.5. Formally, let $m = \lfloor n''/(t\phi) \rfloor$ and 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 rest of 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.5 by conditioning on Y_0 and observing $m = \lfloor \lceil \lfloor L/(4t) \rfloor /t^2 \rceil /(t\phi) \rfloor \geq \lfloor L/(16t^4\phi) \rfloor$.

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 divider 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

$$\sum_{j \in [k]} s_j \cdot w_j = \phi. \tag{34}$$

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 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

$$u_\ell + 1 = \sum_{j \in [k]:\, s_j < 0} \left(z_{r_j} \cdot t\phi + s_j \cdot w_j\right) + \sum_{j \in [k]:\, s_j \geq 0} \left(z'_{r_j} \cdot t\phi + s_j \cdot w_j\right) \qquad \text{(by (34))}$$

$$= \sum_{j \in [k]:\, s_j < 0} \left(z'_{r_j} \cdot |s_j| + z_{r_j} \cdot (t\phi - |s_j|)\right) + \sum_{j \in [k]:\, s_j \geq 0} \left(z_{r_j} \cdot |s_j| + z'_{r_j} \cdot (t\phi - |s_j|)\right).$$

$$\text{(since } w_j = z_{r_j} - z'_{r_j}\text{)}$$

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. $\qquad \square$

B Exposition for the Tail Regime

For completeness, we prove in this section the following version of Theorem 4.3, establishing a large distance bound between $f(\mathcal{U}^m)$ and *every* $\mathcal{D}_\Psi \notin \{\texttt{zeros}, \texttt{ones}, \texttt{zerones}\}$ in the tail regime. Note that Theorem 4.3 follows immediately from Lemma B.1. Recall that we define $\iota(\Psi) = \arg\min_{s \in \Psi} |s - n/2|$, breaking ties arbitrarily.

Lemma B.1. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a d-local function. Assume $n \geq 2^{2^{8 \cdot (d+1)^2}}$ and*

1. *$\iota(\Psi) \leq n/2^{d+2}$ or $\iota(\Psi) \geq n - n/2^{d+2}$,*
2. *and $\Psi \cap \{1, 2, \ldots, n-1\} \neq \emptyset$.*

Then $\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} > 1/2$.

We remark that similar results were proved by [Vio20] and [FLRS23]. Here our argument is analogous to the degree reduction approach presented in [Vio20] (see also [FLRS23, Section 1.1.1]). It is possible to improve Lemma B.1 via the approach in [FLRS23] based on the robust sunflower lemmas [Ros14, ALWZ21, BCW21, Rao19]. Since doing so does not improve the final Theorem 4.1, we choose to proceed with simpler arguments.

Our proof iteratively finds maximal non-connected output bits and then fixes the input bits that they depend on. If at some point we found many output bits that are not constantly zero or one, then we prove a distance bound via arguments like [KOW24, Lemma 4.2]. Otherwise upon the termination of the process, we fixed all the output bits which is a point distribution far from $\mathcal{D}_\Psi$. This is, in spirit, a graph elimination result like the ones that are used extensively in [KOW24].

The argument here is significantly simpler due to the fact that $\mathcal{D}_\Psi$ is far from the middle layers and any non-constant output bit will deviate a lot from its expected marginal distribution. Indeed, the marginals of $\mathcal{D}_\Psi$ are $(\iota(\Psi)/n)$-biased, $(1 - \iota(\Psi)/n)$-biased, or the mixture of the two. Since $|1/2 - \iota(\Psi)/n| \geq 1/2 - 2^{-d-2}$, where 2^{-d-2} is much smaller than the granularity 2^{-d} of each output bit of a d-local function, we expect large distance error analogous to the cases in Subsection 2.1.

Now we present the formal proof.

Proof of Lemma B.1. For convenience, we set up some notation. For every function $g\colon \{0,1\}^m \to \{0,1\}^n$, we define $A_g = \{i \in [n]\colon I_g(i) = \emptyset\}$ to be the set of constant output bits, where we recall that $I_g(i)$ is the set of input bits that the i-th output bit of g depends on. We also define $B_g \subseteq [n] \setminus A_g$ to be an arbitrary maximal set of non-connected non-constant output bits, i.e., $I_g(i) \cap I_g(i') = \emptyset$ for any distinct $i, i' \in B_g$ and each $i \in B_g$ is *not* a constant output.

We will represent the iterative conditioning sketched above by a rooted tree $\mathcal{T}$ as follows:

- Each node of $\mathcal{T}$ is labeled by a function from $\{0,1\}^m$ to $\{0,1\}^n$, where the root is f with depth zero.
- For a node $g \in \mathcal{T}$ of depth k,
    - if $B_g = \emptyset$ (i.e., g is a constant function), then we say g is a constant leaf,
    - if $|B_g| \geq C_k$ where C_k is a parameter to be tuned later, then we say g is a highly independent leaf,
    - otherwise $1 \leq |B_g| < C_k$ and define $T_g = \bigcup_{i \in B_g} I_g(i) \subseteq [m]$. Then g is an internal node with $2^{|T_g|}$ many child nodes labeled by g_ρ for all $\rho \in \{0,1\}^{T_g}$, where g_ρ is g after fixing input bits in T_g by ρ.

Since each B_g is maximal, T_g influences every output bit in $[n] \setminus A_g$. Thus the locality of g_ρ is at least decreased by one from g. This means each node at depth k is $(d-k)$-local and $\mathcal{T}$ has depth at most d. Later we will prove the following claim, which establishes distance bounds for every node in $\mathcal{T}$.

Claim B.2. Assume $n \geq \exp\{C_k \cdot 4^k\}$ holds for all $k \in \{0, 1, \ldots, d\}$. If $C_{k+1} \geq C_k \cdot 8^{d-k+1}$ for all $k \in \{0, 1, \ldots, d-1\}$, then for any node $g \in \mathcal{T}$ of depth k, we have

$$\|g(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - 3 \cdot \exp\left\{-\frac{C_k \cdot 4^{-(d-k)}}{8}\right\}. \tag{35}$$

We now complete the proof of Lemma B.1 assuming Claim B.2. Set each $C_k = 2 \cdot 8^{(k+1) \cdot (d+1)}$. Since we additionally assumed $n \geq 2^{2^{8 \cdot (d+1)^2}}$ in Lemma B.1, the conditions in Claim B.2 are satisfied. Then by Claim B.2 with $k = 0$, we have

$$\|f(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - 3 \cdot e^{-2^{d+1}} > 1/2$$

as desired. $\qquad\square$

Proof of Claim B.2. We proceed by induction on $\mathcal{T}$ in a bottom-up fashion.

Constant Leaves. We first analyze the base case where g is a constant leaf of depth k. Since $\mathcal{D}_\Psi$ is the uniform distribution over its support, the distance of $\mathcal{D}_\Psi$ to $g(\mathcal{U}^m)$, a point distribution, is at least $1 - 1/n$ by Item 2, which verifies (35) assuming $n \geq \exp\{C_k \cdot 4^k\}$.

Highly Independent Leaves. Now we turn to the other base case where g is a highly independent leaf of depth k, i.e., g has $s \geq C_k$ many non-connected non-constant output bits, which we assume without loss of generality is $[s]$. For each $i \in [n]$, let X_i denotes the i-th output bit of g. By the definition of B_g, random variables $X_1, \ldots, X_s$ are not constantly zero or one. Since g is $(d-k)$-local, we know $2^{-(d-k)} \leq \mathbb{E}[X_i] \leq 1 - 2^{-(d-k)}$ for each $i \in [s]$, which implies that

$$\min\left\{\sum_{i \in [s]} \mathbb{E}[X_i], \sum_{i \in [s]} (1 - \mathbb{E}[X_i])\right\} \geq 2^{-(d-k)} \cdot s. \tag{36}$$

For $Y = (Y_1, \ldots, Y_n) \in \{0,1\}^n$, define $\mathcal{E}(Y)$ to be the event that $2^{-(d-k)}/2 \leq \frac{1}{s} \sum_{i \in [s]} Y_i \leq 1 - 2^{-(d-k)}/2$. Since $X_1, \ldots, X_s$ are independent due to the non-connectivity, by Fact A.4 and (36), we have

$$\mathbf{Pr}[\mathcal{E}(g(\mathcal{U}^m))] = 1 - \mathbf{Pr}\left[\frac{1}{s} \sum_{i \in [s]} X_i < 2^{-(d-k)}/2\right] - \mathbf{Pr}\left[\frac{1}{s} \sum_{i \in [s]} (1 - X_i) < 2^{-(d-k)}/2\right]$$

$$\geq 1 - 2 \cdot \exp\left\{-\frac{s \cdot 2^{-(d-k)}}{8}\right\}. \tag{37}$$

On the other hand, for any $0 \leq s \leq n/2^{d+2}$ and $Z = (Z_1, \ldots, Z_n) \sim \mathcal{D}_{\{s\}} = \mathcal{D}_s$ (i.e., a uniformly random string of Hamming weight s), we have

$$\mathbf{Pr}[\mathcal{E}(Z)] \leq \mathbf{Pr}\left[\frac{1}{s} \sum_{i \in [s]} Z_i \geq 2^{-(d-k)}/2\right] \leq \mathbf{Pr}\left[\frac{1}{s} \sum_{i \in [s]} \left(Z_i - \frac{s}{n}\right) \geq 2^{-(d-k)}/4\right]$$

$$\text{(since } \mathbb{E}[Z_i] = s/n \leq 2^{-d-2} \leq 2^{-(d-k)}/4)$$

39

$$\leq \exp\left\{-\frac{s \cdot 4^{-(d-k)}}{8}\right\}, \tag{38}$$

where for the last inequality we used Fact A.3 and the fact that Fact A.3 holds for draw-without-replacement experiments as well (see e.g., [Hoe94, BLM13]). Similarly, for any $n - n/2^{d+2} \leq s \leq n$ and $Z' = (Z'_1, \ldots, Z'_n) \sim \mathcal{D}_s$, we have

$$\mathbf{Pr}[\mathcal{E}(Z')] \leq \mathbf{Pr}\left[\frac{1}{s}\sum_{i \in [s]}\left(Z'_i - \frac{s}{n}\right) \leq -2^{-(d-k)}/4\right] \leq \exp\left\{-\frac{s \cdot 4^{-(d-k)}}{8}\right\}. \tag{39}$$

By Item 1, $\mathcal{D}_\Psi$ is the mixture of $\mathcal{D}_s$ for $s \leq n/2^{d+2}$ or $s \geq n - n/2^{d+2}$. Hence, combining (38) and (39), we have $\mathbf{Pr}[\mathcal{E}(\mathcal{D}_\Psi)] \leq \exp\left\{-s \cdot 4^{-(d-k)}/8\right\}$. Then by (37) and Fact 3.1, since $s \geq C_k$, we obtain

$$\|g(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - 3 \cdot \exp\left\{-\frac{C_k \cdot 4^{-(d-k)}}{8}\right\},$$

which verifies (35).

Internal Nodes. Finally we proceed to the inductive case where g is an internal node of depth k. Since $|B_g| \leq C_k$ and g is $(d-k)$-local, we know that $|T_g| \leq (d-k) \cdot |B_g| \leq (d-k) \cdot C_k$. Hence $g(\mathcal{U}^m)$ is the convex combination of $2^{|T_g|} \leq 2^{(d-k) \cdot C_k}$ many $g_\rho(\mathcal{U}^m)$'s of depth $k+1$, which, by induction hypothesis, satisfies

$$\|g_\rho(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - 3 \cdot \exp\left\{-\frac{C_{k+1} \cdot 4^{-(d-k-1)}}{8}\right\}.$$

Then by Lemma 3.3, we verify (35) as follows:

$$\|g(\mathcal{U}^m) - \mathcal{D}_\Psi\|_{\mathsf{TV}} \geq 1 - \left(1 + 2^{(d-k) \cdot C_k}\right) \cdot 3 \cdot \exp\left\{-\frac{C_{k+1} \cdot 4^{-(d-k-1)}}{8}\right\}$$

$$\geq 1 - 3 \cdot \exp\left\{-\frac{C_{k+1} \cdot 4^{-(d-k)}}{2} + 2 \cdot (d-k) \cdot C_k\right\}$$

$$\geq 1 - 3 \cdot \exp\left\{-\frac{C_k \cdot 4^{-(d-k)}}{8}\right\},$$

where we used the assumption that $C_{k+1} \geq C_k \cdot 8^{d-k+1}$. □

C Missing Proofs in Section 4

Here, we put omitted proofs from Section 4. We will require the following bounds on the sum of binomial coefficients.

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

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

Proof of Lemma 4.7

Proof of Lemma 4.7. The proof is similar to the proof of Lemma 3.3. Let $T \subseteq [\ell]$ be the set of distributions such that $\|\mathcal{P}_i - \mathcal{Q}\|_{\mathsf{TV}} \geq 1 - \eta_1$. By Fact 3.1, for each $i \in T$ there exists an event $\mathcal{E}_i$ such that $\mathcal{P}_i(\mathcal{E}_i) - \mathcal{Q}(\mathcal{E}_i) \geq 1 - \eta_1$. This means

$$\mathcal{P}_i(\mathcal{E}_i) \geq 1 - \eta_1 \quad \text{and} \quad \mathcal{Q}(\mathcal{E}_i) \leq \eta_1 \quad \text{for } i \in T. \tag{40}$$

Define the function f to be the indicator function of the event $\bigvee_{i \in T} \mathcal{E}_i$, i.e., $f(x) = 1$ if some $\mathcal{E}_i$ happens on sample x; and $f(x) = 0$ if otherwise.

By subtracting ϕ by a, we assume $a = 0$ and $b > 0$. By multiplying ϕ and η_2 by $1/b$, we assume $a = 0$ and $b = 1$, i.e., ϕ ranges in $[0, 1]$. Define function $g = \max\{f, 1 - \phi\}$, which ranges in $[0, 1]$. We remark that g becomes the indicator function of $(\neg \mathcal{E}) \vee \bigvee_{i \in T} \mathcal{E}_i$ if ϕ is the indicator function of an event $\mathcal{E}$ as in Lemma 3.3. Let $\mathcal{P} = \sum_{i \in [\ell]} \alpha_i \cdot \mathcal{P}_i$ be the convex combination and let $\eta' = \mathbb{E}_{X \sim \mathcal{Q}}[\phi(X)]$. Then

$$0 \leq \mathbb{E}_{X \sim \mathcal{P}_i}[\phi(X)] \leq \eta' - \eta_2 \quad \text{for } i \notin T. \tag{41}$$

Hence

$$\begin{aligned}
\mathbb{E}_{X \sim \mathcal{P}}[g(X)] &= \sum_{i \in T} \alpha_i \cdot \mathbb{E}_{X \sim \mathcal{P}_i}[g(X)] + \sum_{i \notin T} \alpha_i \cdot \mathbb{E}_{X \sim \mathcal{P}_i}[g(X)] \\
&\geq \sum_{i \in T} \alpha_i \cdot \mathbb{E}_{X \sim \mathcal{P}_i}[f(X)] + \sum_{i \notin T} \alpha_i \cdot \mathbb{E}_{X \sim \mathcal{P}_i}[1 - \phi(X)] && \text{(by the definition of } g) \\
&\geq \sum_{i \in T} \alpha_i \cdot \mathcal{P}_i(\mathcal{E}_i) + \sum_{i \notin T} \alpha_i \cdot \mathbb{E}_{X \sim \mathcal{P}_i}[1 - \phi(X)] && \text{(by the definition of } f) \\
&\geq (1 - \eta_1) \cdot \sum_{i \in T} \alpha_i + (1 - (\eta' - \eta_2)) \cdot \sum_{i \notin T} \alpha_i && \text{(by (40) and (41))} \\
&\geq 1 + \eta_2 - \eta' - \eta_1. && \text{(by (41) and since } \sum_{i \in [\ell]} \alpha_i = 1 \text{ and each } \alpha_i \geq 0)
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbb{E}_{X \sim \mathcal{Q}}[g(X)] &\leq \mathbb{E}_{X \sim \mathcal{Q}}[f(X) + 1 - \phi(X)] && \text{(since both } f \text{ and } 1 - \phi \text{ are non-negative)} \\
&\leq 1 - \mathbb{E}_{X \sim \mathcal{Q}}[\phi(X)] + \sum_{i \in T} \mathcal{Q}(\mathcal{E}_i) && \text{(by the definition of } f) \\
&\leq 1 - \eta' + \ell \cdot \eta_1. && \text{(by (40))}
\end{aligned}$$

Hence

$$\begin{aligned}
\eta_2 - (\ell + 1) \cdot \eta_1 \leq \mathbb{E}_{X \sim \mathcal{P}}[g(X)] - \mathbb{E}_{X \sim \mathcal{Q}}[g(X)] &= \sum_x g(x) \cdot (\mathcal{P}(x) - \mathcal{Q}(x)) \\
&\leq \sum_{x \,:\, \mathcal{P}(x) \geq \mathcal{Q}(x)} g(x) \cdot (\mathcal{P}(x) - \mathcal{Q}(x)) \\
&\leq \sum_{x \,:\, \mathcal{P}(x) \geq \mathcal{Q}(x)} (\mathcal{P}(x) - \mathcal{Q}(x)) && \text{(since } g(x) \in [0, 1]) \\
&= \|\mathcal{P} - \mathcal{Q}\|_{\mathsf{TV}} && \text{(by Fact 3.1)}
\end{aligned}$$

as desired. $\qquad \square$

Proof of Claim 4.10

Proof of Claim 4.10. Since Ψ is in the central regime, by definition $n/2^{d+2} \leq \iota(\Psi) \leq n - n/2^{d+2}$. Hence $\iota(\Psi) \in \overline{\Psi}$ by construction. By Fact C.1, we have

$$A := \sum_{s \in \Psi:\, s < n/2^{d+3} \text{ or } s > n - n/2^{d+3}} \binom{n}{s} \leq 2 \cdot \sum_{0 \leq s < n/2^{d+3}} \binom{n}{s} \leq 2 \cdot 2^{n \cdot \mathcal{H}(1/2^{d+3})}. \tag{42}$$

Then by Fact 3.7, we have

$$B := \binom{n}{\iota(\Psi)} \geq \binom{n}{n/2^{d+2}} \geq \frac{2^{n \cdot \mathcal{H}(1/2^{d+2})}}{\sqrt{8n/2^{d+2}}}. \tag{43}$$

Since $\mathcal{D}_{\overline{\Psi}}$ is simply $\mathcal{D}_{\Psi}$ conditioned on the Hamming weight being at least $n/2^{d+3}$ and at most $n - n/2^{d+3}$. By Fact 3.2, we have

$$\begin{aligned}
\left\| \mathcal{D}_{\Psi} - \mathcal{D}_{\overline{\Psi}} \right\|_{\mathsf{TV}} &= \Pr_{X \sim \mathcal{D}_{\Psi}} \left[|X| < n/2^{d+3} \text{ or } |X| > n - n/2^{d+3} \right] \\
&= \frac{A}{|\mathsf{supp}\,(\mathcal{D}_{\Psi})|} \leq \frac{A}{A+B} \leq \frac{A}{B} \qquad \text{(by the definition of } \mathcal{D}_{\Psi}) \\
&\leq 2\sqrt{8n/2^{d+2}} \cdot 2^{-n \cdot \left(\mathcal{H}(1/2^{d+2}) - \mathcal{H}(1/2^{d+3}) \right)}. \qquad \text{(by (42) and (43))}
\end{aligned}$$

By Fact 3.8, we have

$$\begin{aligned}
\mathcal{H}(1/2^{d+2}) - \mathcal{H}(1/2^{d+3}) &= \frac{1}{2\ln(2)} \sum_{m \geq 1} \frac{\left(-1 + 2^{-(d+2)} \right)^{2m} - \left(-1 + 2^{-(d+1)} \right)^{2m}}{m \cdot (2m-1)} \\
&\geq \frac{\left(1 - 2^{-(d+2)} \right)^2 - \left(1 - 2^{-(d+1)} \right)^2}{2\ln(2)} \geq \frac{2^{-(d+3)}}{\ln(2)}.
\end{aligned}$$

Hence

$$\left\| \mathcal{D}_{\Psi} - \mathcal{D}_{\overline{\Psi}} \right\|_{\mathsf{TV}} \leq 8\sqrt{n \cdot 2^{-(d+3)}} \cdot \exp\left\{ -n \cdot 2^{-(d+3)} \right\} \leq 8 \cdot \exp\left\{ -n \cdot 2^{-(d+4)} \right\},$$

where we used $\sqrt{x} \leq e^{x/2}$ for the last inequality. $\qquad \square$

Proof of Lemma 4.13

Proof of Lemma 4.13. The proof follows closely with the proofs of [KOW24, Lemmas 5.15 & 5.22]. By rearranging indices, we assume without loss of generality that $N(1), \ldots, N(r')$ are non-connected Type-2 neighborhoods of sizes $1 \leq s_1, \ldots, s_{r'} \leq t$. Sample $Z \sim \mathcal{U}^m$ and set $X = (X_1, \ldots, X_n) = g(Z)$. Define

$$K = \sum_{i \notin N(1) \cup \cdots \cup N(r')} X_i \quad \text{and} \quad \Delta_j = \sum_{i \in N(j)} X_i \text{ for each } j \in [r'].$$

Then $|X| = K + \sum_{j \in [r']} \Delta_j$.

Let $R = [m] \setminus (I(1) \cup \cdots \cup I(r'))$ be the set of input bits that do not affect the first r' output bits. Define $\rho \in \{0,1\}^R$ as the entries of Z in R. For each ρ, we define $\lambda_\rho = 2^{-|R|}$, X_ρ being K conditioned on ρ, and $X_{\rho,j}$ being Δ_j conditioned on ρ. Then Item 1 trivially holds. Now we verify Item 2: first observe that for every $i \notin N(1) \cup \cdots \cup N(r')$, X_i depends on $I(i) \subseteq [m]$ which is

contained in R and fixed by ρ. Hence X_ρ is indeed a fixed integer. In addition, since $N(1), \ldots, N(r')$ are non-connected neighborhoods, $\Delta_1, \ldots, \Delta_{r'}$ depend on disjoint subsets of $[m]$. Since ρ simply fixes some input bits in $[m]$, they remain independent conditioned on ρ, which means $X_{\rho,1}, \ldots, X_{\rho,r'}$ are independent. Finally, each $X_{\rho,j}$ is an integer ranging between 0 and $|N(j)|$, where the latter is at most t by our assumption.

To prove Item 3, for each ρ, j and $q \geq 3$, we define

$$p_{\rho,q,j} = \max_{x \in \mathbb{Z}} \mathbf{Pr}\left[X_{\rho,j} \equiv x \pmod{q}\right] = \max_{x \in \mathbb{Z}} \mathbf{Pr}\left[\Delta_j \equiv x \pmod{q} \mid \rho\right].$$

By a similar argument to [KOW24, Claim 5.16], we obtain the following claim.

Claim C.2. $\mathbb{E}_\rho[(p_{\rho,q,j})^2] \leq 1 - 2^{-7t-d+2}$ holds for any $j \in [r']$ and $q \geq 3$.

Proof. We only highlight the difference from the proof of [KOW24, Claim 5.16]. Since $N(j)$ is Type-2, by definition, $\mathcal{P}_g|_{N(j)}$ is ε-close to the γ-biased distribution where $\gamma = s/n$ for some $n/2^{d+3} \leq s \leq n - n/2^{d+3}$. Hence $2^{-(d+3)} \leq \gamma \leq 1 - 2^{-(d+3)}$.

Then we apply Lemma 3.4 with γ and modulus $\bar{q} = \min\{q, t+1\}$ here. Let $\gamma^* = \min\{\gamma, 1 - \gamma\}$. Since $2^{-(d+3)} \leq \gamma^* \leq 1/2$ and $2 \leq \bar{q} \leq t+1$, we have

$$\frac{\gamma^*}{4\bar{q}} \cdot 2^{-50\gamma^*(t-1)/\bar{q}^2} \geq \frac{2^{-(d+3)}}{4(t+1)} \cdot 2^{-7(t-1)} = \frac{2^{-7t-d+2}}{t+1} \geq 2^{-7t-d+1} \geq \varepsilon,$$

which implies $\mathbb{E}_\rho[(p_{\rho,q,j})^2] \leq 1 - 2^{-7t-d+2}$ by Lemma 3.4. $\qquad\square$

Given Claim C.2, we have $\mathbb{E}_\rho[1 - p_{\rho,q,j}] \geq 2^{-7t-d+1}$. Since $p_{\rho,q,j}$'s are independent over random ρ for fixed q, by Fact A.4 we have

$$\mathbf{Pr}\left[\sum_{j \in [r']} (1 - p_{\rho,q,j}) \leq 2^{-7t-d} \cdot r'\right] \leq \exp\left\{-2^{-7t-d-2} \cdot r'\right\}.$$

Recall that we say ρ is bad if for some $q \geq 3$ the above event happens. By Item 2, we can additionally assume $q \leq t+1$ since $p_{\rho,q,j} = p_{\rho,t+1,j}$ for all $q \geq t+1$. By the union bound, we have

$$\mathbf{Pr}\left[\rho \text{ is bad}\right] \leq t \cdot \exp\left\{-2^{-7t-d-2} \cdot r'\right\}$$

as stated in Item 3. $\qquad\square$

Proof of Claim 4.15

Proof of Claim 4.15. We first handle the case $\gamma + \xi \leq 1$. Then (8) and (9) become $\gamma \leq 1 - \delta'$ and $\xi \leq 1 - \delta'$. Hence $\frac{\gamma^2 + \xi^2}{\gamma + \xi} \leq \max\{\gamma, \xi\} \leq 1 - \delta'$ as desired.

Now we assume $\gamma + \xi > 1$. Then the constraints are

$$0 \leq \gamma, \xi \leq 1, \quad 1 < \gamma + \xi \leq 2 \cdot (1 - \delta'), \quad \xi \geq \frac{\delta' \cdot \gamma}{1 - \delta'}, \quad \gamma \geq \frac{\delta' \cdot \xi}{1 - \delta'}.$$

By symmetry, assume without loss of generality $\gamma \leq \xi$. Observe that if $\xi < 1$ and $\gamma + \xi < 2 \cdot (1 - \delta')$, then we can increase γ, ξ by a small multiple that still satisfies the constraints but increases the optimization objective. Hence we safely assume that either $\xi = 1$ or $\gamma + \xi = 2 \cdot (1 - \delta')$.

- If $\xi = 1$, then $\gamma + \xi \leq 2 \cdot (1 - \delta')$ implies $\gamma \leq 1 - 2\delta'$ and $\gamma \geq \frac{\delta' \cdot \xi}{1 - \delta'}$ implies $\gamma \geq \frac{\delta'}{1 - \delta'}$. Then the objective is

$$\frac{\gamma^2 + \xi^2}{\gamma + \xi} = \frac{\gamma^2 + 1}{\gamma + 1} = \gamma + 1 + \frac{2}{\gamma + 1} - 2 \leq \frac{1 - 2\delta' + 2\delta'^2}{1 - \delta'}$$
$$\leq 1 - \delta'/2. \qquad \text{(since } \delta' \leq 1/3\text{)}$$

- If $\gamma + \xi = 2 \cdot (1 - \delta')$, then $\xi \leq 1$ implies $\gamma \geq 1 - 2\delta'$. In addition, $\gamma \leq 1 - \delta'$ since we assumed $\gamma \leq \xi$. Then the objective is

$$\frac{\gamma^2 + \xi^2}{\gamma + \xi} = \frac{4 \cdot (1 - \delta')^2 - 2\gamma(2 - 2\delta' - \gamma)}{2 \cdot (1 - \delta')} = \frac{\gamma^2}{1 - \delta'} - 2\gamma + 2 \cdot (1 - \delta')$$
$$\leq \frac{(1 - 2\delta')^2}{1 - \delta'} - 2 \cdot (1 - 2\delta') + 2 \cdot (1 - \delta') = \frac{1 - 2\delta' + 2\delta'^2}{1 - \delta'}$$
$$\leq 1 - \delta'/2$$

as well. $\qquad\qquad\square$

Proof of Claim 4.17

We first prove the following estimate on $\mathsf{wt}_{\leq m}$.

Lemma C.3. *For all* $m \in \{0, 1, \ldots, n\}$*, we have* $\mathsf{wt}_{\leq m} = \mathsf{wt}_{\geq n-m}$*. Assume* $n \geq 35$ *and* $m = (n - c)/2$ *for some* $0 \leq c \leq n$*. Then*

$$2^{-17} \cdot 4^{-c^2/n} \leq \mathsf{wt}_{\leq m} \leq \mathsf{wt}_m \cdot \frac{n + 1}{c + 1}.$$

Proof. The first equality relation follows from the definition of $\mathsf{wt}_{\leq m}$ and $\mathsf{wt}_{\geq n-m}$. For the upper bound, we first note that the bound trivially holds for $m = 0$ (i.e., $c = n$). Hence we assume $m \geq 1$ and observe that

$$\mathsf{wt}_{\leq m} = 2^{-n+1} \sum_{i \geq 0} \binom{n}{m - 2i} \leq 2^{-n+1} \sum_{i \leq m} \binom{n}{i} \leq 2^{-n+1} \cdot \binom{n}{m} \cdot \frac{n - m + 1}{n - 2m + 1} \qquad \text{(by Fact C.1)}$$
$$= \mathsf{wt}_m \cdot \frac{n - m + 1}{n - 2m + 1} \leq \mathsf{wt}_m \cdot \frac{n + 1}{n - 2m + 1} = \mathsf{wt}_m \cdot \frac{n + 1}{c + 1}. \qquad \text{(since } m = (n - c)/2\text{)}$$

For the lower bound, we first observe that, if $0 \leq c \leq n - 10$, we have $5 \leq m \leq n/2$ and thus $m - 2\lfloor\sqrt{m}\rfloor \geq 1$ and $m - 2\sqrt{m} \geq 0$. Then

$$\mathsf{wt}_{\leq m} \geq 2^{-n+1} \sum_{i \geq 0}^{\lfloor\sqrt{m}\rfloor} \binom{n}{m - 2i} \geq 2^{-n+1} \cdot \sqrt{m} \cdot \binom{n}{m - 2\lfloor\sqrt{m}\rfloor}$$
$$\geq 2^{-n+1} \cdot \sqrt{m} \cdot \frac{2^{n \cdot \mathcal{H}\left(\frac{m - 2\sqrt{m}}{n}\right)}}{\sqrt{8m}} \qquad \text{(by Fact 3.7 and } 1 \leq m - 2\lfloor\sqrt{m}\rfloor \leq n/2, m - 2\sqrt{m} \geq 0\text{)}$$
$$= \frac{2^{-n}}{\sqrt{2}} \cdot 2^{n \cdot \mathcal{H}\left(\frac{1 - (c + 4\sqrt{m})/n}{2}\right)} \qquad \text{(since } m = (n - c)/2\text{)}$$
$$\geq \frac{1}{\sqrt{2}} \cdot 2^{-(c + 4\sqrt{m})^2/n} \qquad \text{(by Fact 3.8)}$$

44

$$\geq \frac{1}{\sqrt{2}} \cdot 2^{-\left(c+4\sqrt{n/2}\right)^2/n} \qquad\qquad \text{(since } m \leq n/2 \text{ and } c \geq 0\text{)}$$

$$\geq 2^{-17} \cdot 4^{-c^2/n}. \qquad\qquad \text{(since } (a+b)^2 \leq 2a^2 + 2b^2\text{)}$$

Now if $n - 10 < c \leq n$, then $\mathsf{wt}_{\leq m} \geq \mathsf{wt}_0 = 2^{-n+1}$. On the other hand, $c^2 \geq n^2/2$ as $n \geq 35$. Hence the bound $2^{-17} \cdot 4^{-c^2/n} \leq 2^{-17} \cdot 2^{-n} \leq \mathsf{wt}_{\leq m}$ also holds. $\qquad\square$

Now we prove Claim 4.17.

Proof of Claim 4.17. We first show that

$$m_{\mathsf{L}} \leq n/2 - \sqrt{n \log(1/(\delta_1\gamma))}/8. \tag{44}$$

Since $\gamma \leq 1 - \delta_2$ from (11), this implies that $m_{\mathsf{L}} \leq n/2 - \sqrt{n}$ as $\delta_1\gamma \leq \delta_1(1 - \delta_2) \leq 2^{-100}$. Let $M = \left\lfloor n/2 - \sqrt{n \log(1/(\delta_1\gamma))}/8 \right\rfloor$ and let $m^* \in \{M, M-1\}$ be the largest even number at most M. It is equivalent to show $\mathsf{wt}_{\leq m^*} \geq \delta_1\gamma/4$. Define $c = n - 2m^*$ and observe that

$$0 \leq \sqrt{n \log(1/(\delta_1\gamma))}/4 \leq c \leq 4 + \sqrt{n \log(1/(\delta_1\gamma))}/4 \leq \sqrt{n \log(1/(\delta_1\gamma))}/3 \leq n$$

since $\delta_1\gamma > 2^{-n+1}$ and $n \geq 2^{10}$. Hence by Lemma C.3, we have

$$\mathsf{wt}_{\leq m^*} \geq 2^{-17} \cdot 4^{-c^2/n} \geq 2^{-17} \cdot (\delta_1\gamma)^{2/3} \geq \delta_1\gamma/4$$

since $\delta_1\gamma \leq \delta_1(1 - \delta_2) \leq 2^{-100}$. Therefore for any $m_{\mathsf{L}} \leq m \leq n/2$, we have

$$\mathsf{wt}_m \geq \mathsf{wt}_{m_{\mathsf{L}}} \geq \frac{n - 2 \cdot m_{\mathsf{L}} + 1}{n+1} \cdot \mathsf{wt}_{\leq m_{\mathsf{L}}} \qquad\qquad \text{(by Lemma C.3)}$$

$$\geq \frac{n - 2 \cdot m_{\mathsf{L}} + 1}{n+1} \cdot \frac{\delta_1\gamma}{4} \geq \frac{\sqrt{n \log(1/(\delta_1\gamma))}/4}{n} \cdot \frac{\delta_1\gamma}{4} \qquad \text{(by the definition of } m_{\mathsf{L}} \text{ and (44))}$$

$$= \frac{\delta_1\gamma\sqrt{\log(1/(\delta_1\gamma))}}{16\sqrt{n}}.$$

The bound on m_{R} and the case $n/2 \leq m \leq m_{\mathsf{R}}$ can be proved in an analogous way. $\qquad\square$