Electronic Colloquium on Computational Complexity, Report No. 31 (2024)

# Locality Bounds for Sampling Hamming Slices

Daniel M. Kane[*] Anthony Ostuni[†] Kewen Wu[‡]

## Abstract

Spurred by the influential work of Viola (Journal of Computing 2012), the past decade has witnessed an active line of research into the complexity of (approximately) *sampling* distributions, in contrast to the traditional focus on the complexity of *computing* functions.

We build upon and make explicit earlier implicit results of Viola to provide superconstant lower bounds on the locality of Boolean functions approximately sampling the uniform distribution over binary strings of particular Hamming weights, both exactly and modulo an integer, answering questions of Viola (Journal of Computing 2012) and Filmus, Leigh, Riazanov, and Sokolov (RANDOM 2023). Applications to data structure lower bounds and quantum-classical separations are discussed.

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

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

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

ISSN 1433-8092

# Contents

# 1 Introduction

Historically, complexity theory has been dominated by research determining the complexity of *computing* particular functions. Following the seminal work of Viola [Vio12b], the past decade has seen a rise in study on the complexity of *sampling* particular distributions [Vio12b, LV11, BIL12, DW12, Vio16, Vio20, GW20, CGZ22, Vio23, FLRS23]. In this setting, the goal is to construct a circuit whose input is an infinite string of unbiased, independent random bits and whose output is a distribution close in total variation distance to a specified distribution.

As a standard motivating example, consider the parity function. Håstad's flagship result [Hås86] shows that $\mathsf{AC}^0$ circuits need exponentially many gates to compute parity. However, one can easily sample pairs of the form $(X, \textsc{parity}(X))$, where each output bit depends on just two input bits: output $(x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_{n-1} \oplus x_n, x_n \oplus x_1)$ on input $x_1, x_2, \ldots$ [Bab87, BL87]. In fact, $\mathsf{AC}^0$ circuits can even sample $(X, f(X))$ for more complicated functions, such as inner product [IN96] and symmetric functions [Vio12b].

Despite these examples illustrating the difficulty of sampling lower bounds, the challenge has been rewarded with results providing intuition for and applications to succinct data structures [Vio12b, LV11, BIL12, Vio20, CGZ22, Vio23], pseudorandom generators [Vio12a, LV11, BIL12], and extractors [Vio12c, DW12, Vio14, CZ16, CS16].

Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a Boolean function of $n$ output bits. Additionally, let $\mathcal{U}^m$ be the uniform distribution over $\{0,1\}^m$ and $f(\mathcal{U}^m)$ be the output distribution of $f$ given a uniform input. We say $f$ is $d$-local if every output bit of $f$ depends on at most $d$ input bits. For constant $d$, this captures $\mathsf{NC}^0$ circuits, and more generally, $d$-local functions encompass circuits of depth $\log(d)$ and bounded fan-in. In his influential paper, Viola [Vio12b] considered the problem of determining locality lower bounds for Boolean functions $f$ where $f(\mathcal{U}^m)$ approximates one of the following distributions:

1. $\mathcal{D}_k$ – the uniform distribution over $x \in \{0,1\}^n$ of Hamming weight $k = \Theta(n)$.
2. $\mathcal{M}_n$ – the uniform distribution over $x \in \{0,1\}^n$ conditioned on $\textsc{modmaj}_p(x) = 0$ where

$$\textsc{modmaj}_p(x) \coloneqq \mathbb{1}[(x_1 + \cdots + x_n) \bmod p \geq p/2] \quad \text{for some prime } p = \Theta(\log(n)).$$

In particular, an $\Omega(\log(n))$ locality lower bound was proved for both distributions[1], conditional on $f$'s input domain being sufficiently small ($m = (1 + o(1)) \cdot n$). Additionally, Viola gave results without this restriction, but with a worse distance bound decaying exponentially in the locality $d$.

Towards a full locality-distance trade-off, Viola asked whether lower bounds could be obtained with strong error bounds and no input length restriction. Later, the similar bound $\widetilde{\Omega}(\log(n/k))$ was discovered for $\mathcal{D}_k$ by Filmus, Leigh, Riazanov, and Sokolov [FLRS23], answering the question for small $k$. To complement this result, Viola proved an $\Omega(\log(n))$ bound in the case of non-dyadic[2] $k/n$ [Vio23].[3] However, Viola's question in the general regime remained open.

More recently, building on Viola's results on the distribution $\mathcal{M}_n$, Watts and Parham [WP23] proved an input-independent separation between $\mathsf{QNC}^0$ and $\mathsf{NC}^0$ circuits of restricted domain size. Such domain conditions boil down to the exact domain conditions of the locality bounds for $\mathcal{M}_n$ in Viola's analysis, rendering their result only a partial separation.

[1] Viola actually considered the uniform distribution over $(X, \textsc{modmaj}_p(X))$ pairs, but these are essentially equivalent (see, e.g., [Vio12b, Lemma 4.3]).

[2] Recall a number is *dyadic* if it can be expressed as a fraction whose denominator is a power of two.

[3] The bound was originally implicit in [Vio23] with many of the ideas appearing earlier in [Vio12a]. Very recently and after the submission of our paper, Viola posted an update making the bound explicit – see Section 9 of https://eccc.weizmann.ac.il/report/2021/073/. All our bounds, including the non-dyadic case, were proven independently.

### 1.1 Our Results

We resolve Viola's question in the affirmative by providing nontrivial locality lower bounds on $\mathcal{D}_k$ in the $k = \Theta(n)$ regime, as well as for $\mathcal{M}_n$, with no restrictions on the domain size.

Formally, we say a distribution $\mathcal{P}$ is $\delta$-far (resp., $\delta$-close) from a distribution $\mathcal{Q}$ if the total variation distance is at least $\delta$ (resp., at most $\delta$). Our results provide a wide range of trade-offs between locality and distance. For readability, we present them here with some particular parameter choices.

**Theorem 1.1** (Consequence of Theorem 5.10)**.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function, and let $1 \le k \le n-1$. If $n$ is sufficiently large and*

$$d \le \frac{\log^*(\log^*(n))}{60} \quad \text{and} \quad \frac{1}{\log^*(\log^*(n))} \le \frac{k}{n} \le 1 - \frac{1}{\log^*(\log^*(n))},$$

*then $f(\mathcal{U}^m)$ is $\left(1 - \frac{\log^*(n)}{\sqrt{n}}\right)$-far from $\mathcal{D}_k$, where $\log^*(\cdot)$ is the iterated logarithm with base 2.*

By a different application of Theorem 5.10, we obtain the following sharp bound. The tightness of Theorem 1.2 is discussed in Subsection 5.3.

**Theorem 1.2** (Consequence of Theorem 5.10)**.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function and $1 \le k \le n-1$. If both $d$ and $k/n$ are constant, then $f(\mathcal{U}^m)$ is $(1 - O(1/\sqrt{n}))$-far from $\mathcal{D}_k$.*

We remark that the above theorems hold in a stronger setting where $f$ is fed with some arbitrary binary[4] product distribution as an input. Taking advantage of the fact that the input is actually unbiased coins, we prove the following bound.

**Theorem 1.3** (Consequence of Theorem 5.7)**.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function, where $n$ is a multiple of 3. Then $f(\mathcal{U}^m)$ is $\left(1 - \exp\left\{-n \cdot 2^{-O(d^2)}\right\}\right)$-far from $\mathcal{D}_{n/3}$.*

The above theorem (as well as a version with adaptivity) appears with the bound $1-2\cdot 2^{-\sqrt{n}/2^{O(d)}}$ in [Vio23] (see Footnote 3), which for sufficiently large $d$ eclipses our result (as well as Theorem 1.1 and Theorem 1.2 when $k/n$ is non-dyadic). However, the proof in [Vio23] notes the $\sqrt{n}$ term can be optimized, so the exact trade-off between bounds depends on the extent this optimization is possible. Given these similar bounds, we view our primary contribution here to be the generality of our statements, as they have no dyadic restrictions.

Towards lower bounds for the distribution $\mathcal{M}_n$, we introduce the following notation. Let $q$ be an integer and let $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$ be a non-empty set, where $\mathbb{Z}/q\mathbb{Z} = \{0, 1, \ldots, q-1\}$. We define the distribution $\mathcal{D}_{q,\Lambda}$ to be the uniform distribution over $x \in \{0,1\}^n$ conditioned on $(x_1 + \cdots + x_n) \bmod q \in \Lambda$.

**Theorem 1.4** (Consequence of Theorem 5.17)**.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function. Let $3 \le q \le \sqrt{n/\log^*(n)}$ be an integer, and let $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$ be a non-empty set. If $n$ is sufficiently large and $d \le \log^*(\log^*(n))/20$, then $f(\mathcal{U}^m)$ and $\mathcal{D}_{q,\Lambda}$ have distance at least*

$$1 - \exp\left\{-\frac{n}{q^2 \cdot \log^*(n)}\right\} - \begin{cases} |\Lambda|/q & q \text{ is odd,} \\ 2 \cdot \max\left\{|\Lambda_{\mathsf{even}}|, |\Lambda_{\mathsf{odd}}|\right\}/q & q \text{ is even,} \end{cases}$$

*where $\Lambda_{\mathsf{even}}$ is the set of even numbers in $\Lambda$ and $\Lambda_{\mathsf{odd}}$ is the set of odd numbers in $\Lambda$.*

[4] In fact, it works even in the case where the product distribution is not binary, as long as the alphabet is a constant (or slightly superconstant). This will be clear in the proof, and we will discuss it in Section 2.

Taking $q = p$ and $\Lambda = \{c \in \mathbb{Z}/q\mathbb{Z} \colon c \geq p/2\}$, we recover $\mathcal{D}_{q,\Lambda} = \mathcal{M}_n$. By setting $\Lambda = \{0\}$, Theorem 1.4 also answers an open problem in [FLRS23] for locality lower bounds for the uniform distribution over binary strings of Hamming weight 0 mod $q$.

We highlight that Theorem 1.4 is essentially tight for *any* choice of $q$ and $\Lambda$, and direct interested readers to the discussion in Subsection 5.4.

### 1.1.1 Data Structure Lower Bounds

It is an active line of research to determine optimal bounds for *succinct data structures*: structures that store their data close to the information theoretic limit, while still including sufficient redundancy to allow for efficient and meaningful queries [GM07, Vio12a, Vio12b, LV11, BIL12, Vio20, PY20, LPPZ23]. We will focus on the following setting of a binary alphabet and bit probes.

**Definition 1.5** (Dictionary Problem)**.** Let $\mathcal{H} \subseteq \{0,1\}^n$ and $s, q \in \mathbb{N}$. The dictionary problem of $\mathcal{H}$ with parameters $s$ and $q$ asks for a pair of algorithms $\mathcal{A}$ and $\mathcal{B}$ such that the following holds:

- Given an arbitrary $a \in \mathcal{H}$, $\mathcal{A}$ produces a data structure $\mathsf{str}_a \in \{0,1\}^s$.
- Given access to $\mathsf{str} \in \{0,1\}^s$, for every query $i \in [n]$, $\mathcal{B}$ produces an answer $b_i \in \{0,1\}$ with $q$ (adaptive / non-adaptive) bit probes (i.e., number of bits read) to $\mathsf{str}$.
- When $\mathsf{str} = \mathsf{str}_a$, we have $b_i = a_i$ for all $i \in [n]$.

We remark that this setting is static, in contrast to the dynamic setting where the data structure needs to support updates to the underlying input $a$. As a weaker model, proving static lower bounds has traditionally been much more difficult than dynamic lower bounds. The locality-distance trade-off provides a useful tool in establishing trade-offs between parameters in the static dictionary problem.

**Claim 1.6** ([Vio12b, Claim 1.8])**.** Suppose we can solve the dictionary problem of $\mathcal{H} \subseteq \{0,1\}^n$ with parameters $s, q$ and non-adaptive (resp., adaptive) queries. Then there exists a $q$-local (resp., $(2^q - 1)$-local) function $f \colon \{0,1\}^s \to \{0,1\}^n$ such that $f(\mathcal{U}^s)$ is $(1 - |\mathcal{H}|/2^s)$-close to the uniform distribution over $\mathcal{H}$.

Combining Claim 1.6 with our results, we obtain a number of lower bounds. Here we highlight the most interesting ones, and interested readers are encouraged to instantiate more on their own.

**Corollary 1.7** (Via Theorem 1.4)**.** *Let $\mathcal{H} = \{a \in \{0,1\}^n \colon (a_1 + \cdots + a_n) \bmod r = 0\}$ where $r \geq 3$ is an odd constant. The dictionary problem of $\mathcal{H}$ needs either $s = n$ bits of storage or $q = \omega(1)$ bit probes per query.*

Note that the information theoretic limit is $\lceil \log(|\mathcal{H}|) \rceil = n - \lfloor \log(r) \rfloor$. On the other hand, the trivial data structure that simply stores $a$ in $n$ bits can support every membership query by a single bit probe. Hence Corollary 1.7 shows that the *only* efficient data structure using constant probes is the trivial one. We can obtain a similar sharp trade-off for even $r$ by adding a modulus to reflect Theorem 1.4's different quantitative behavior for odd and even moduli.

**Corollary 1.8** (Via Theorem 1.4)**.** *Let $\mathcal{H} = \{a \in \{0,1\}^n \colon (a_1 + \cdots + a_n) \bmod r \in \{0,1\}\}$ where $r \geq 3$ is an even constant. The dictionary problem of $\mathcal{H}$ needs either $s = n$ bits of storage or $q = \omega(1)$ bit probes per query.*

In the case of $\mathcal{H} = \{a \in \{0,1\}^n \colon a_1 + \cdots + a_n = n/k\}$ where $n/k$ is non-dyadic, the results of [Vio12a] are superior to those we can obtain via our techniques. Specifically, they show that the dictionary problem of $\mathcal{H}$ with $q$ (adaptive) queries requires at least $\log(|\mathcal{H}|) + n2^{-O(q)} - \log(n)$ bits of storage. However, our generality allows us to prove bounds in the dyadic setting.

**Corollary 1.9** (Via Theorem 1.2)**.** *Let $\mathcal{H} = \{a \in \{0,1\}^n : a_1 + \cdots + a_n = n/2\}$ where $n$ is a multiple of two. The dictionary problem of $\mathcal{H}$ needs either $s = n - O(1)$ bits of storage or $q = \omega(1)$ bit probes per query.*

The previous best result in the setting of Corollary 1.9 is [Vio12b], which gives an $s = n - 0.01\log(n)$ versus $q = \Omega(\log(n))$ trade-off. Our Corollary 1.9 improves the storage bound to optimal (ignoring the hidden constant in $O(1)$) at the cost of a worse bit probe bound. It remains an interesting question whether one can get the best of the two results that further improves our bit probe bound to $\Omega(\log(n))$ without weakening the $n - O(1)$ storage bound.

Meanwhile we note that it is impossible to get an $n$-vs-$\omega(1)$ trade-off as in Corollary 1.7 and Corollary 1.8. Here is a simple data structure of $s = (n-1)$-bit of storage and using $q = 2$ bit probes per query for $\mathcal{H}$ in Corollary 1.9: for each $i \in [n-1]$, store the prefix sum $\mathsf{str}_a[i] = a_1 \oplus \cdots \oplus a_i$. For $i = n$, the prefix sum is precisely $n/2 \bmod 2$ that we do not need to store. Then every query $i$ can be answered by the parity of the prefix sum up to $i$ and $i-1$. It would be interesting to determine if a similar structure exists with fewer than $(n-1)$-bits of storage while maintaining $O(1)$ bit probes per query.

The results compared here are by no means a complete list of data structure lower bounds for the dictionary problem. In particular, there are many results on cell probe lower bounds (e.g., [PY20]), the dynamic dictionary setting (e.g., [LLYZ23]), and other natural choices of $\mathcal{H}$ (e.g., [Vio12a]). We refer interested readers to [Vio12b] for a detailed discussion.

### 1.1.2 Input-Independent Quantum-Classical Separation

A driving research direction in quantum computing is exhibiting separations between quantum and classical complexity. In the theme of our paper, we consider the problem of devising distributions that quantum circuits can efficiently sample, whereas classical circuits cannot. Note that such a separation does not rely on a particular input. Instead, the quantum circuit is fed with a fixed initial state (ideally $|0\rangle^n$), and each qubit is measured in the computational basis at the end to produce the desired distribution over $\{0,1\}^n$. Meanwhile, a classical circuit, which has $n$ output bits, has access to unbiased coins and aims to reproduce the distribution.

The problem of establishing such an *input-independent* separation between circuit classes $\mathsf{QNC}^0$ and $\mathsf{NC}^0$ was first proposed by Bravyi, Gosset, and König [BGK18], and was later found to be connected to the complexity of quantum states [WP23] as well. Using ideas from [Vio12b], Watts and Parham [WP23] gave a family of distributions over $\{0,1\}^n$ that constant-depth quantum circuits can produce within distance $1/6 + o(1)$, but any $\mathsf{NC}^0$ circuit's output is at least $(1/2 - o(1))$-far from, assuming the total number of random bits the $\mathsf{NC}^0$ circuit could use is $(1 + o(1)) \cdot n$. The exact distributions are variants of the $\mathcal{M}_n$ distribution above.

However, an ideal separation result should have no restriction on the number of classical random bits, as well as a maximal quantum-classical distance gap of $1 - o(1)$ or even $1 - e^{-\Omega(n)}$. Towards this goal, [FLRS23] suggested determining locality lower bounds for $\mathcal{M}_n$ and related distributions. Without diving into detail, we remark that our Theorem 1.4 resolves this open problem, and we can lift the domain size assumption in [WP23, Theorem 5] while still preserving the separation.

Aside from directly improving previous analysis, we note that there is a simpler distribution that produces an optimal separation. Let $\mathcal{U}^n_{1/3}$ be the 1/3-biased distribution over $n$ bits, where each bit is independently set as 1 with probability 1/3.

**Theorem 1.10** (Consequence of Theorem 5.3)**.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function. Then $f(\mathcal{U}^m)$ is $\left(1 - \exp\left\{-n \cdot 2^{-O(d^2)}\right\}\right)$-far from $\mathcal{U}^n_{1/3}$.*

Observe that, starting with $|0\rangle^n$, a $\mathsf{QNC}^0$ circuit can perfectly simulate $\mathcal{U}_{1/3}^n$ using just one layer of single-qubit gates each mapping $|0\rangle$ to $\sqrt{2/3}\,|0\rangle + \sqrt{1/3}\,|1\rangle$. Thus we obtain the following ideal input-independent quantum-classical separation.

**Corollary 1.11.** *There exists a distribution that $\mathsf{QNC}^0$ circuits of depth one can sample without error, but any $\mathsf{NC}^0$ circuit is $(1 - \exp\{-\Omega(n)\})$-far from.*

We suspect this result may be folklore, especially after a reviewer pointed out that Theorem 1.10 is implicit in [Vio12a, Vio23] (with a stronger bound in at least some parameter regimes). However, it does not seem to explicitly appear in the literature, so we hope our statement will be beneficial to future researchers.

**Remark 1.12.** The quantum-classical separation result obtained in Corollary 1.11 seems a bit dishonest, as it takes advantage of precision issues arising from the classical binary representation. One may desire a separation where the quantum circuit is also restricted to "binary operations" to rule out distributions like 1/3-biased. One natural candidate is Clifford circuits where non-Clifford gates are not allowed. However, the sampling task there sometimes can be reduced to the search task with a constant depth overhead [GS20, Section F], where the latter is trivial in the input-independent setting. Hence one must be careful in formulating such a restriction on the quantum circuit.

A different way to compensate for the precision issue is to give $\mathsf{NC}^0$ circuits access to arbitrary binary product distributions. Then $\mathsf{NC}^0$ circuits can certainly generate $\mathcal{U}_{1/3}^n$ by simply receiving 1/3-biased coins. We remark that our proof of Theorem 1.4 still holds in this setting.[5] Thus, even giving $\mathsf{NC}^0$ this extra power, we have a separation combining Theorem 1.4 and [WP23, Theorem 5]. One caveat here is that the $\mathsf{QNC}^0$ circuit needs to start with the $\mathrm{GHZ}_n$ state. If it is forced with $|0\rangle^n$ as the initial state, one may want to prove locality lower bounds (without the domain assumption) for a more complicate distribution designed in [WP23, Theorem 3]. Due to its similarity with the $\mathcal{M}_n$ distribution, we believe our techniques can be used there, and we leave this as a future work.

## 1.2 Future Directions

Beyond considering specific distributions, Filmus, Leigh, Riazanov, and Sokolov [FLRS23] conjectured a classification of when $\mathsf{NC}^0$ circuits can approximately sample $\mathcal{D}_\Lambda$, where $\mathcal{D}_\Lambda$ is the uniform distribution over binary strings with Hamming weights in $\Lambda$. They hypothesized that if $f$ is $O(1)$-local and $f(\mathcal{U}^m)$ is $\varepsilon$-close to $\mathcal{D}_\Lambda$, then $f(\mathcal{U}^m)$ is $O(\varepsilon)$-close to $\mathcal{D}_{\Lambda'}$ for $\Lambda'$ being one of the following:

$$\{0\}, \{n\}, \{0, n\}, \{0, 2, 4, \ldots\}, \{1, 3, 5, \ldots\}, [n].$$

Our Theorem 1.1, combined with their main results, rules out all the singleton $\Lambda'$ other than $\{0\}$ and $\{n\}$. In addition, our Theorem 1.4 rules out all the $q$-periodic $\Lambda'$ for $3 \le q \le n^{1/2-o(1)}$. With a number of new ideas, in an upcoming paper [KOWar] we are able to resolve this conjecture (and a strengthening of it) affirmatively.

One question we are not able to resolve concerns the quantitative bounds derived. While our distance bounds are asymptotically optimal when locality is constant, the locality-distance trade-offs deteriorate quickly when locality becomes superconstant. We believe our trade-offs can be further improved, especially in light of the best known upper bounds (see Section 6). However, in Appendix A we give examples to show the tightness of the parts in our analysis that create such inevitable blowup. This suggests that new ideas may be needed to get substantial improvements.

[5] Theorem 1.4 works in the case where the input distribution is a product distribution with constant (or slightly superconstant) alphabet. This will be clear in the proof, and we will discuss it in Section 2.

**Paper Organization.** An overview of our proofs is given in Section 2. In Section 3, we define necessary notation and list standard inequalities. In Section 4, we prove additional useful inequalities for total variation distance and bipartite graph structures. In Section 5, we prove sharp lower bounds for specific distributions including biased distributions, uniform strings of a fixed Hamming weight, and uniform strings of periodic Hamming weights. Upper bound constructions are presented in Section 6. Missing proofs can be found in the appendices. In addition, in Appendix A, we give examples to explain the barriers of improving our analysis to obtain better locality lower bounds.

## 2 Proof Overview

Let $f$ be a $d$-local function with $n$ output bits, and let $\mathcal{D}$ be a distribution over $\{0,1\}^n$. The goal is to prove that $f$, fed with uniform inputs, cannot generate a distribution close to $\mathcal{D}$. The general recipe of establishing such a bound is as follows:

1. First, we consider a simpler setting where not only does every output bit of $f$ depend on few input bits, but every input bit of $f$ influences few output bits as well.

   In this case, we can find many output bits that depend on disjoint sets of input bits. Now if the desired distribution $\mathcal{D}$ has long-range correlation (e.g., the Hamming weight must equal $k$), we would expect a large error, since these output bits are independent and cannot coordinate with each other.

2. Then, based on the error bound established in the first step, we aim to reduce the general case, where we may have popular input bits that many output bits depend on, to the structured case above.

   At this step, we shall prove certain graph elimination results, showing that the desired structure in the first step can be obtained after deleting some input bits.

The above description is an oversimplification of our analysis, and for each of our results in Subsection 1.1 we face different issues, which we elaborate on below. For convenience and simplicity, we will hide minor factors when stating bounds.

The framework of viewing $f$ as a convex combination of specific, easier-to-handle restrictions was largely developed in [Vio12b, Vio20] and applied in [FLRS23]. Thus, our primary contributions are the specific choices of structure we reduce to and the corresponding technical analysis.

**The $1/3$-Biased Distribution.** We first consider the toy example $\mathcal{D} = \mathcal{U}_{1/3}^n$, the $1/3$-biased distribution. The idea here works equally well for any $\gamma$-biased distribution where $\gamma$ is non-dyadic. Observe that $1/3$ can only be approximated up to error $\approx 2^{-d}$ using integer multiples of $2^{-d}$. Therefore the marginal distribution for every output bit of $f$ is doomed to be $2^{-d}$-far from a $1/3$-biased coin. Since total variation distance is closed under marginal projections, this already implies a $2^{-d}$ bound.

To further boost it to $1-o(1)$, we first assume that we can find $r$ non-connected output bits, i.e., they do not depend on common input bits, which means they are independent. Since each one of these output bits incurs $2^{-d}$ error, intuitively their error should accumulate. We prove (Lemma 4.2) that this is indeed the case. If we have $r$ pairs of distributions $(\mathcal{P}_1, \mathcal{Q}_1), \ldots, (\mathcal{P}_r, \mathcal{Q}_r)$ where each $\mathcal{P}_i$ is $\varepsilon$-far from $\mathcal{Q}_i$, then their products $\mathcal{P}_1 \times \cdots \times \mathcal{P}_r$ and $\mathcal{Q}_1 \times \cdots \times \mathcal{Q}_r$ are $(1-2^{-\varepsilon^2 r})$-far from each other. We briefly sketch the proof: each weak distance bound implies an event $\mathcal{E}_i$ that happens $\varepsilon$ more often in $\mathcal{P}_i$ than $\mathcal{Q}_i$. Then by independence and standard concentration, the number of total events happening in $\mathcal{P}_1 \times \cdots \times \mathcal{P}_r$ is typically $r \cdot \varepsilon/2$ larger than the number in $\mathcal{Q}_1 \times \cdots \times \mathcal{Q}_r$, thus establishing the bound. Applied here, each $\mathcal{P}_i$ corresponds to a selected output bit, and each $\mathcal{Q}_i$ is a $1/3$-biased coin. Hence we get (Proposition 5.5) a $1 - \exp\left\{-r \cdot 2^{-d}\right\}$ bound.

Now back to reality, we may not immediately find non-connected output bits, since the degrees of input bits can be unbounded. For example, there could be one input bit that all outputs depend on, and therefore no two output bits are independent. However, conditioning on this one bit would decrease the degree to $d-1$ and also fix the problem, at the cost of changing the distribution by a factor of 2. Since the distance bound above is sufficiently strong, we can indeed pay some loss to condition on input bits.

In particular, we show (Lemma 4.3) that the convex combination of distributions $\mathcal{P}_1, \ldots, \mathcal{P}_m$ is $(1 - m \cdot \varepsilon)$-far from a distribution $\mathcal{Q}$, provided that each $\mathcal{P}_i$ is $(1-\varepsilon)$-far from $\mathcal{Q}$. This is proved as follows: each distance bound implies an event $\mathcal{E}_i$ happening with probability at least $1 - \varepsilon$ in $\mathcal{P}_i$ but at most $\varepsilon$ in $\mathcal{Q}$. Then their disjunction will inherent the $1 - \varepsilon$ probability in any convex combination of $\mathcal{P}_i$'s, but still happens with at most $m \cdot \varepsilon$ probability in $\mathcal{Q}$ by the union bound, which gives the desired statement. Applied here, each $\mathcal{P}_i$ corresponds to the distribution of output bits conditioned on a specific assignment of $c = \log(m)$ input bits. This will add a $2^c$ overhead on top of the distance bound for the distribution after each conditioning. Given this observation and the $1 - \exp\left\{-r \cdot 2^{-d}\right\}$ bound above, we can afford to delete roughly $r \cdot 2^{-d}$ input bits as long as we can find $r$ non-connected output bits after this.

At this point, the problem is graph theoretic: given a bipartite graph $G$ where each left vertex (representing an output bit) has degree bounded by $d$, we are allowed to delete a few right vertices (representing input bits) to get many non-connected left vertices, where we say two left vertices are non-connected if they are not both adjacent to the same right vertex. More precisely, we are allowed to delete at most $r \cdot 2^{-d}$ right vertices to get at least $r$ non-connected left vertices. In addition, we would like to maximize $r$, since the final bound will be roughly $1 - \exp\left\{-r \cdot 2^{-d}\right\}$. It turns out that this can be achieved (Proposition 5.6) with $r = n/2^{d^2}$, which explains our bounds in Theorem 1.10. The starting point of the proof is the following naive attempt: if we remove all right vertices of degree at least $\ell$, then we obtain a bipartite graph with left degree $d$ and right degree $\ell$, which readily gives $n/(d \cdot \ell)$ non-connected left vertices. Hence if the desired bound does not hold, the number of right vertices of degree at least $\ell$ is larger than $r \cdot 2^{-d} \geq n/(d2^d \cdot \ell)$. Then summing over all $\ell$ up to roughly $2^{2^d}$, we will find the right-hand side of above will be a sum of harmonic series and larger than $d \cdot n$, whereas the left-hand side of above is still upper bounded by the number of total edges, which is at most $d \cdot n$. This forms a contradiction. By analyzing more carefully, we can improve (Corollary 4.8) the $2^{2^d}$ to $2^{d^2}$. This turns out to be sharp (Appendix A.1).

**The Hamming Slice of Weight $n/3$.** Now we move on to the single Hamming slice case $\mathcal{D} = \mathcal{D}_k$, the uniform distribution over $n$-bit binary strings of Hamming weight $k$. A simpler case here is still when $k/n$ has precision issues – think of $k = n/3$ for now. Then every bit in $\mathcal{D}_{n/3}$ is supposed to be $1/3$-biased, whereas every output bit in the produced distribution is still $2^{-d}$-far from it.

While we largely follow the analysis above, the caveat here is that being far from $\mathcal{U}_{1/3}^n$ does not imply being far from $\mathcal{D}_{n/3}$, as the distance between $\mathcal{U}_{1/3}^n$ and $\mathcal{D}_{n/3}$ is itself $1 - 1/\sqrt{n}$. More precisely, this issue arises when we try to aggregate the errors from $m$ independent output bits. In the previous argument, we compared $\mathcal{P}_i$ (representing the true output bits) and $\mathcal{Q}_i = \mathcal{U}_{1/3}$ (representing the desired marginal distributions), then showed that the weak individual error can be boosted to $1 - o(1)$ between their product distributions, where the issue kicks in as the product of $\mathcal{Q}_i$'s is $\mathcal{U}_{1/3}^n$ instead of (and actually far from) $\mathcal{D}_{n/3}$.

To get around this, we strengthen (Proposition 5.8) the above argument to use $\mathcal{Q}_i$'s as a proxy between the actual distribution and the desired distribution. Notice that, despite being far in the total variation distance metric, $\mathcal{U}_{1/3}^n$ is close to $\mathcal{D}_{n/3}$ in the pointwise multiplicative error sense. More formally, every string in the support of $\mathcal{D}_{n/3}$ has density $\binom{n}{n/3}^{-1}$, and has density $(1/3)^{n/3}(2/3)^{2n/3}$ under $\mathcal{U}_{1/3}^n$. These two quantities are only off by a $\sqrt{n}$ multiplicative factor, which means every event of probability at most $\varepsilon$ under $\mathcal{U}_{1/3}^n$ will have probability at most $\varepsilon\sqrt{n}$ under $\mathcal{D}_{n/3}$. Therefore we can modify (Lemma 4.2) the previous analysis to show that there is an event of probability at least $1 - \exp\left\{-r \cdot 2^{-d}\right\}$ under $\mathcal{P}_1 \times \cdots \times \mathcal{P}_r$ but of probability at most $\sqrt{n} \cdot \exp\left\{-r \cdot 2^{-d}\right\}$ under $\mathcal{D}_{n/3}$, thus establishing a strong distance bound between the actual and desired distributions. Since later in the graph theoretic task we will set $r \approx n/2^{d^2}$, this $\mathsf{poly}(n)$ loss is affordable when $d$ is not particularly large.

**The Hamming Slice of Weight $n/2$.** The above analysis works well when individual output bits have inevitable error against the marginal of the desired distribution. As such, we need new ideas if we want to establish lower bounds for the general case. For simplicity let us focus on the $k = n/2$ case, as the analysis will generalize to any $k$ that is not too close to 0 or $n$.

If we can find $r$ independent output bits that are $\varepsilon$-far from being unbiased, then we can use the same argument to boost them to a $1 - \exp\left\{-\varepsilon^2 r\right\}$ bound. Otherwise we need to exploit the long-range correlation of $\mathcal{D}_{n/2}$ that the Hamming weight must sum to exactly $n/2$. One possible exploitation is through anticoncentration inequalities, which have played an important role in the analysis of similar problems [Vio12b, CGZ22]. In particular, if there are $r$ independent output bits that are actually unbiased, then by Littlewood-Offord anticoncentration [LO43, Erd45], they cannot sum to any particular value with probability more than $1/\sqrt{r}$, which seemingly means the distribution is still $(1 - 1/\sqrt{r})$-far from $\mathcal{D}_{n/2}$. The issue with this argument is that the $r$ independent output bits can correlate with many other output bits, which might be able to force the total Hamming weight to a fixed value. For example, one can consider the construction $(X_1, 1 - X_1, X_2, 1 - X_2, \ldots, X_{n/2}, 1 - X_{n/2})$, where we have $n/2$ independent bits but the total sum is always $n/2$ and every individual bit is unbiased.

To address this problem, we need to take into account the neighborhood of each output bit. Define the neighborhood $N(i)$ of an output bit $i$ as the set of output bits that depend on some input bit that $i$ also depends on. We will exploit a key tension between two facts about $N(i)$'s distribution. Firstly, every small neighborhood should be unbiased, since the marginals of $\mathcal{D}_{n/2}$ restricted to any small number of bits are $1/\mathsf{poly}(n)$-close to the uniform distribution over those bits. Secondly, resampling the input bits on which $i$ depends should not change the Hamming weight of the output (and thus does not change the Hamming weight of $N(i)$). However, since the output of $i$ depends only on these inputs, the second property implies the distribution over Hamming weights of $N(i)$ conditioned on $i = 0$ would be the same as the distribution over $i = 1$, which contradicts the first property. Note this argument has no issue with the above construction.

Let $\varepsilon$ be a parameter to be optimized later. We classify each neighborhood as $\mathsf{Type\text{-}1}$ if it is $\varepsilon$-far from being unbiased, and as $\mathsf{Type\text{-}2}$ if it is $\varepsilon$-close to unbiased coins. Mimicking the previous analysis, we say two neighborhoods $N(i), N(j)$ are non-connected if all pairs $(i', j') \in N(i) \times N(j)$ are non-connected. Thus by the same argument (Lemma 5.14), if we have $r$ non-connected neighborhoods of $\mathsf{Type\text{-}1}$, then our distribution is at distance $1 - \sqrt{n} \cdot \exp\left\{-\varepsilon^2 r\right\}$ from $\mathcal{D}_{n/2}$.

Now suppose we have $r$ non-connected neighborhoods of $\mathsf{Type\text{-}2}$, each of size at most $t$. We would like to use anticoncentration inequalities to argue that with high probability the Hamming weight does not sum up to $n/2$. Assume the neighborhoods are $N(1), \ldots, N(r)$ and $I(i)$ is the set of input bits the $i$-th output bit depends on. We fix all the input bits outside $I(1) \cup \cdots \cup I(r)$ as $\rho$. Then all the output bits outside $N(1) \cup \cdots \cup N(r)$ are fixed, and moreover, the neighborhoods

$N(1), \ldots, N(r)$ are independent to each other. At this point, if the Hamming sum of each $N(i)$ is still not fixed, we can apply anticoncentration (Fact 3.4) to obtain the desired bound.

To this end, we use the property that $N(i)$ is Type-2, i.e., it is roughly unbiased under random $\rho$. Say $N(i)$ has size $t$. Then under a uniform random input, the Hamming sum of $N(i)$ is distributed like a binomial distribution of $t$ coins. If we resample the input bits in $I(i)$, with half probability the $i$-th output bit is flipped, whereas the Hamming sum of $N(i) \setminus i$ is a binomial distribution of $t-1$ coins. This implies that such an experiment has $1/\sqrt{t} - \varepsilon$ probability of changing the Hamming sum of $N(i)$, where $1/\sqrt{t}$ comes from the total variation distance between a binomial distribution of $t$ coins and its shift, and $\varepsilon$ comes from the error between the actual distribution of $N(i)$ and $\mathcal{U}^t$. Meanwhile, since $\rho$ does not touch $I(i)$, we cannot change the Hamming sum by simply resampling $I(i)$ if the Hamming sum is already fixed by $\rho$. Hence as long as $\varepsilon \le 1/\sqrt{t}$, we show (Claim 5.16) that the Hamming sum of $N(i)$ is not fixed under random (and thus a typical) $\rho$. Since these neighborhoods are independent, by standard concentration many neighborhoods will enjoy this property simultaneously for a typical $\rho$. Then we can apply anticoncentration and obtain a bound of roughly $1 - 1/\sqrt{r}$ (Lemma 5.15).

Set $\varepsilon = 1/(2\sqrt{t})$. To summarize (Proposition 5.12), if we have $r$ non-connected neighborhoods of size at most $t$, then

- either $r/2$ of them are Type-1, which implies a distance bound of $1 - \sqrt{n} \cdot \exp\{-r/t\}$;
- or $r/2$ of them are Type-2, which implies a distance bound of $1 - 1/\sqrt{r}$.

Following the previous argument, this means that we can afford conditioning on $\min\{r/t, \log r\}$ input bits to get the above structure. This seems too stringent and impossible, even without considering the undesirable loss in the final bound. Instead, we observe that the distance bound from the second case actually tells more; it is proved in the stronger sense that our output distribution hits any point in the support of $\mathcal{D}_{n/2}$ with probability at most $1/\sqrt{r}$. Hence we can refine (Lemma 4.3) the previous analysis as follows: any convex combination of $\mathcal{P}_1, \ldots, \mathcal{P}_m$ is $(1 - m \cdot \varepsilon_1 - \varepsilon_2)$-far from $\mathcal{Q}$, provided that each $\mathcal{P}_i$ is either $(1-\varepsilon_1)$-far from $\mathcal{Q}$, or hits the support of $\mathcal{Q}$ with probability at most $\varepsilon_2$. The proof is not much different, and we simply merge the event "not hitting the support" into the previous union bound. Therefore we can remove $r/t$ input bits now.

Finally we need to handle the graph theoretic task: given a bipartite graph $G$ with left degree at most $d$, show we can obtain $r$ non-connected left neighborhoods (representing the neighborhoods of output bits) of size $t$ by removing $r/t$ right vertices. The left neighborhood of a left vertex is the set of left vertices reachable from it with two edges. Two left neighborhoods are non-connected if they do not connect to common right vertices. In addition, we aim to maximize $r$ and minimize $t$, since the final distance bound will be $1 - 1/\sqrt{r} - \sqrt{n} \cdot \exp\{-r/t\}$. This task is significantly more challenging than the previous one, as now we need to eliminate the dependency of the neighborhoods too. Consequently, we only get a bound with tower-type dependence on $d$. That is, we show (Proposition 5.6) that the problem can be solved with $r = n/\mathrm{tow}_2(d)$ and $t = \mathrm{tow}_2(d)$.[6] Perhaps surprisingly, this tower-type dependency is in fact necessarily (Appendix A.2).

Here we briefly sketch the proof. As before, assume towards contradiction it is false. Then we follow the previous approach and argue that we will have too many right vertices of large degree, which will imply the following structural result (Lemma 4.10): if we have removed $n/\alpha$ right vertices from the graph where $\alpha \ge C$ is sufficiently large, we can additionally remove $n/\log(\alpha)$ right vertices to shave $n$ edges from the graph. Then we arrive at a contradiction, as the graph has at most $d \cdot n$ edges and thus can support the elimination process up to $d$ times. However repeating $d$ times

[6] $\mathrm{tow}_2(d) = 2^{2^{2^{\cdot^{\cdot^{\cdot}}}}}$ is the tower of 2's of height $d$.

only removes roughly $n/\log^{(d)}(n)$ many right vertices in total, which means the elimination process should continue if $\log^{(d)}(n) \geq C$.[7]

**Hamming Slices of Weight 0 Modulo 3.** We note that the analysis for $\mathcal{D}_{n/2}$ also works for the union of multiple Hamming slices, since the main place where we use $n/2$ is that it is *one* fixed value and thus has $1/\sqrt{n}$ bound via anticoncentration. Beyond a single slice, the $1/\sqrt{n}$ bound simply scales with the number of slices. Nevertheless, this does not go beyond $\sqrt{n}$ slices. Here we demonstrate that our framework is robust enough to handle $\Omega(n)$ periodic slices.

For simplicity, we consider the case where $\mathcal{D}$ equals the uniform distribution over $n$-bit binary strings with Hamming weight 0 modulo 3. Note that this distribution consists of roughly $n/3$ Hamming slices and has marginal distribution almost unbiased. We follow the proof of $\mathcal{D}_{n/2}$. Let $\varepsilon$ be a parameter measuring the distance between the marginal distribution of neighborhoods and the unbiased distribution. Similarly we classify each neighborhood as Type-1 if the distance is at least $\varepsilon$, and as Type-2 if otherwise. Once we have $r$ non-connected neighborhoods of Type-1, we readily get a $1 - \exp\left\{-\varepsilon^2 r\right\}$ distance bound following the same argument.

On the other hand, if we have $r$ non-connected Type-2 neighborhoods of size at most $t$, then we use anticoncentration inequalities (in fact, a local limit theorem) to show that with certain probability we cannot have Hamming weight equal to 0 modulo 3. Recall that in the single Hamming slice case, we argue that a Type-2 neighborhood $N(i)$ is not fixed after a typical restriction $\rho$ which does not touch $I(i)$ (the input bits that the $i$-th output bit depends on). This is proved via a thought experiment where we resample the input bits in $I(i)$ and compare the binomial distribution of $t$ coins with its shift. Here we need a similar statement (Claim 5.23) that a Type-2 neighborhood is not fixed *modulo* 3 after a typical restriction $\rho$. The only difference is that now we need to compare the binomial-modulo-3 distribution with its shift. Since 3 does not divide 2, the binomial-modulo-3 distribution can never be uniform over $\{0, 1, 2\}$. In fact, by granularity, it is $2^{-t}$-far from its shift, which means that the Hamming sum modulo 3 of $N(i)$ is typically not fixed as long as $\varepsilon \leq 2^{-t}/2$. Then using a local limit theorem (which is an almost tight Littlewood-Offord-type anticoncentration) on the additive group modulo 3 (Lemma 3.7), we obtain that under typical $\rho$, the Hamming sum modulo 3 is roughly uniform over $\{0, 1, 2\}$, thus it hits any particular value with probability $1/3 + o(1)$.

Set $\varepsilon = 2^{-t}/2$. To summarize (Proposition 5.19), if we have $r$ non-connected neighborhoods of size at most $t$, then

- either $r/2$ of them are Type-1, which implies a distance bound of $1 - \exp\left\{-r/2^t\right\}$;
- or $r/2$ of them are Type-2, then hitting 0 modulo 3 has probability at most $1/3 + o(1)$.

By the same reasoning, we seek the above structure at the cost of removing at most $r/2^t$ input bits, while simultaneously maximizing $r$ and minimizing $t$. It turns out that this (Proposition 5.20) is still manageable with a tower-type loss on $d$ via a similar graph elimination argument.

At last, we mention that the local limit theorem used for analyzing Type-2 neighborhoods holds generally for all modulus (Theorem B.1) including 2. However, the comparison between the binomial-modulo-$q$ distribution and its shift can only be done for modulus $q \geq 3$. This is because the binomial-modulo-2 distribution is indeed uniform over $\{0, 1\}$. Thus for even $q$'s (i.e., $q$'s not coprime with 2), there will be an additional contributing factor (Lemma 3.7), which results in a different bound for even $q$'s in Theorem 1.4.

[7] $\log^{(d)}(n) = \log(\log(\log(\cdots(n))))$ is the iterated logarithm of order $d$.

**More General Input Distributions.** Now we briefly discuss how to modify our analysis to prove similar lower bounds when the input distribution changes from unbiased coins to general product distributions. While this is not true for the 1/3-biased distribution (or any $\gamma$-biased in general), it works for the Hamming slices setting. Since it is standard that a Boolean circuit takes unbiased coins as input, we focus on this case and leave the following more general treatment for interested readers as an exercise.

Recall that our analysis starts with a simpler setting where we can find many small non-connected neighborhoods. In this case, we prove distance lower bounds by comparing the marginal distributions of these non-connected neighborhoods with the desired marginal distribution (unbiased or 1/3-biased coins). Then we classify them into Type-1 and Type-2 and argue the final distance bound separately. The analysis in this part has nothing to do with the *input distribution*, since the only property we need is the non-connectivity of output neighborhoods in the input-output relation, which generalizes trivially when we view the input "bits" as taking values in a larger alphabet. Then we reduce the general setting to the above simpler setting by removing a few input bits. This part is also oblivious to the alphabet of the input as it works in a purely graph theoretic sense where the input-output dependency is defined in an abstract way regardless the alphabet.

The only problematic part is where we put the above two steps together (Lemma 4.3). There, we have to pay a union bound of $m$ for all the possible conditioning (or equivalently, the number of different distributions after conditioning), since the true output distribution is a convex combination of them. If the alphabet of the input is $\Sigma$ and we need to remove $t$ input bits, we will need to set $m = |\Sigma|^t$. To compensate this loss, the graph theoretic problem needs to be slightly reformulated, but it will still be manageable if $|\Sigma|$ is a constant or even slightly superconstant. For example, in the setting of $\mathcal{D} = \mathcal{D}_{n/2}$, previously we needed to obtain $r$ non-connected neighborhoods of size $t$ after removing $r/t$ input bits; now we need to obtain $r$ non-connected neighborhoods of size $t$ after removing $r/(t \cdot \log(|\Sigma|))$ input "bits".

Finally we remark that an extremely general result, where the bounds have no restriction on the alphabet, is simply not true. One can use a 1-local function to sample any distribution if the input alphabet is large enough to include all possible outcomes and is dubbed just with the desired distribution.

## 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 positive integer $q$, we use $\mathbb{Z}/q\mathbb{Z} = \{0, 1, \ldots, q-1\}$ to denote the additive group modulo $q$. 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. We use $\log^*(x)$ to denote the iterated logarithm with base $e$:

$$\log^*(x) = \begin{cases} 0 & 0 \le x \le 1, \\ 1 + \log^*(\log(x)) & x > 1. \end{cases}$$

For $a > 0$ and $b \in \mathbb{N}$, we use $\mathrm{tow}_a(b)$ to denote the power tower of base $a$ and order $b$, where

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

Note that $\log^*(\mathrm{tow}_2(b)) = b$ and $x \le \mathrm{tow}_2(\log^*(x)) \le 2^x$.

**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 $\gamma$-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 $\mathsf{supp}(\mathcal{P}) = \{x \colon \mathcal{P}(x) > 0\}$ 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.[8] We say $\mathcal{P}$ is $\varepsilon$-close to $\mathcal{Q}$ if $\|\mathcal{P}(x) - \mathcal{Q}(x)\|_{\mathsf{TV}} \le \varepsilon$, and $\varepsilon$-far otherwise.

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

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

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 $S$, we use $\mathcal{P}^S$ to emphasize that coordinates of $\mathcal{P}^{|S|}$ are indexed by elements in $S$. 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 $\mathcal{P} = \sum_{i \in [t]} \alpha_i \cdot \mathcal{P}_i$.

**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 $|I_f(i)| \le d$ 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 $|I_G(i)| \le d$ 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 $i$.

[8] 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.

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 $\deg_G(j) = |\{i \in V_1 : j \in I_G(i)\}|$ 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 $I_G(i), N_G(i), \deg_G(j)$.

**Concentration and Anticoncentration.** We will use the following standard concentration inequalities.

**Fact 3.2** (Hoeffding's Inequality)**.** *Assume $X_1, \ldots, X_n$ are independent random variables such that $a \le X_i \le b$ holds for all $i \in [n]$. Then for all $\delta \ge 0$, we have*

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

**Fact 3.3** (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*

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

We also need the following version of the Littlewood-Offord-type anticoncentration inequality, which uniformly bounds the probability density function of the sum of independent random variables.

**Fact 3.4** ([Ush86, Theorem 3])**.** *Assume $X_1, \ldots, X_n$ are independent random variables in $\mathbb{R}$. For each $i \in [n]$, define $p_i = \max_{x\in\mathbb{R}} \mathbf{Pr}[X_i = x]$. Then there exists a universal constant $C > 0$ such that*

$$\mathbf{Pr}\left[\sum_{i\in[n]} X_i = x\right] \le \frac{C}{\sqrt{\sum_{i\in[n]}(1-p_i)}} \qquad \textit{holds for any } x \in \mathbb{R}.$$

**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 3.5** (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 3.6** (See e.g., [Wik23])**.** *For any $x \in [-1,1]$, we have*

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

**Local Limit Theorems.** Local limit theorems provide sharp estimates for the probability density function of the sum of independent random variables, strengthening the usual (anti-)concentration inequalities and central limit theorems. We refer interested readers to a recent survey by Szewczak and Weber [SW22].

We will require the following local limit result in the additive group modulo $q$, which is a special case of the more general statement Theorem B.1. The proof of Lemma 3.7 is deferred to Appendix B, where we also discuss its tightness.

**Lemma 3.7.** *Let $q \geq 3$ be an integer, and let $X_1, \ldots, X_n$ be independent random variables in $\mathbb{Z}$. For each $i \in [n]$ and $r \geq 1$, define $p_{r,i} = \max_{x \in \mathbb{Z}} \mathbf{Pr}\left[X_i \equiv x \pmod r\right]$ and assume*

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

*Then for any $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$, we have*

$$\mathbf{Pr}\left[\sum_{i \in [n]} X_i \mod q \in \Lambda\right] \leq q \cdot e^{-2L/q^2} + \begin{cases} |\Lambda|/q & q \text{ is odd,} \\ 2 \cdot \max\left\{|\Lambda_{\mathsf{even}}|, |\Lambda_{\mathsf{odd}}|\right\}/q & q \text{ is even,} \end{cases}$$

*where $\Lambda_{\mathsf{even}} = \{$even numbers in $\Lambda\}$ and $\Lambda_{\mathsf{odd}} = \{$odd numbers in $\Lambda\}$.*

# 4 Useful Lemmas

In this section, we prove additional useful lemmas which will appear multiple times with various parameter choices in later sections. In such generality, they may be of independent interest elsewhere.

## 4.1 Total Variation Bounds

Here we prove various lemmas to control total variation bounds.

The following fact is standard, showing that two distributions close to each other remain close after conditioning. For completeness, we include a proof in Appendix C.

**Fact 4.1.** *Assume $\mathcal{P}$ is $\varepsilon$-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 $f$,*

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

Intuitively if the marginals of two product distributions do not match, the two distributions in general should be extremely far apart. This intuition is generalized and formalized as the following lemma, where we actually prove a strengthening of the above intuition that works even for non-product distributions.

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

- *$\mathcal{P}|_S$ and $\mathcal{W}|_S$ are two product distributions,*
- *$\left\|\mathcal{P}|_{\{i\}} - \mathcal{W}|_{\{i\}}\right\|_{\mathsf{TV}} \geq \varepsilon$ holds for all $i \in S$, and*

- *$\mathcal{W}(x) \geq \eta \cdot \mathcal{Q}(x)$ holds for some $\eta > 0$ and all $x$.*

*Then*

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

*Proof.* By Fact 3.1, for each $i \in S$ there exists an event $\mathcal{E}_i$ such that $\mathcal{P}|_{\{i\}}(\mathcal{E}_i) - \mathcal{W}|_{\{i\}}(\mathcal{E}_i) \geq \varepsilon$. Let $\mathbb{1}_{\mathcal{E}_i} \in \{0,1\}$ be the indicator of event $\mathcal{E}_i$. Now define event $\mathcal{E}$ such that

$$\mathcal{E} \text{ happens if and only if } \tfrac{1}{s}\textstyle\sum_{i\in S}\left(\mathbb{1}_{\mathcal{E}_i} - \mathcal{P}|_{\{i\}}(\mathcal{E}_i)\right) \geq -\varepsilon/2.$$

Then

$$\begin{aligned}
\mathcal{P}(\mathcal{E}) = \mathcal{P}|_S(\mathcal{E}) &= \Pr_{\mathcal{P}|_S}\left[\frac{1}{s}\sum_{i\in S}\left(\mathbb{1}_{\mathcal{E}_i} - \mathcal{P}|_{\{i\}}(\mathcal{E}_i)\right) \geq -\varepsilon/2\right] \\
&= 1 - \Pr_{\mathcal{P}|_S}\left[\frac{1}{s}\sum_{i\in S}\left(\mathbb{1}_{\mathcal{E}_i} - \mathcal{P}|_{\{i\}}(\mathcal{E}_i)\right) < -\varepsilon/2\right] \\
&\geq 1 - e^{-\varepsilon^2 s/2}. \qquad \text{(since } \mathcal{P}|_S \text{ is a product distribution and by Fact 3.2)}
\end{aligned}$$

We also have

$$\begin{aligned}
\mathcal{W}(\mathcal{E}) = \mathcal{W}|_S(\mathcal{E}) &= \Pr_{\mathcal{W}|_S}\left[\frac{1}{s}\sum_{i\in S}\left(\mathbb{1}_{\mathcal{E}_i} - \mathcal{P}|_{\{i\}}(\mathcal{E}_i)\right) \geq -\varepsilon/2\right] \\
&\leq \Pr_{\mathcal{W}|_S}\left[\frac{1}{s}\sum_{i\in S}\left(\mathbb{1}_{\mathcal{E}_i} - \mathcal{W}|_{\{i\}}(\mathcal{E}_i)\right) \geq \varepsilon/2\right] \qquad \text{(since } \mathcal{P}_i(\mathcal{E}_i) \geq \mathcal{W}|_{\{i\}}(\mathcal{E}_i) + \varepsilon\text{)} \\
&\leq e^{-\varepsilon^2 s/2}. \qquad \text{(since } \mathcal{W}|_S \text{ is a product distribution and by Fact 3.2)}
\end{aligned}$$

Since $\mathcal{W}(x) \geq \eta \cdot \mathcal{Q}(x) > 0$ for all $x$, we have

$$\mathcal{W}(\mathcal{E}) = \sum_{x:\mathcal{E}\text{ happens}} \mathcal{W}(x) \geq \sum_{x:\mathcal{E}\text{ happens}} \mathcal{Q}(x) \cdot \eta = \eta \cdot \mathcal{Q}(\mathcal{E}),$$

which then implies $\mathcal{Q}(\mathcal{E}) \leq e^{-\varepsilon^2 s/2}/\eta$. Therefore by Fact 3.1, we obtain the desired bound. □

Suppose we can prove distance bounds from a distribution to a set of distributions. This should establish distance bounds from the former distribution to any distribution inside the convex hull of the latter set of distributions. This is characterized by Lemma 4.3, a special case of which appears in [Vio20, Section 4.1].

**Lemma 4.3.** *Let $\mathcal{P}_1, \ldots, \mathcal{P}_t$ and $\mathcal{Q}$ be distributions. Assume there exists an event $\mathcal{E}$ and values $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 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*

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

*Proof.* Let $T \subseteq [t]$ be the set of distributions such that $\|\mathcal{P}_i - \mathcal{Q}\|_{\mathsf{TV}} \ge 1-\varepsilon_1$. By Fact 3.1, for each $i \in T$ there exists an event $\mathcal{E}_i$ such that $\mathcal{P}_i(\mathcal{E}_i) - \mathcal{Q}(\mathcal{E}_i) \ge 1-\varepsilon_1$. This means

$$\mathcal{P}_i(\mathcal{E}_i) \ge 1-\varepsilon_1 \quad \text{and} \quad \mathcal{Q}(\mathcal{E}_i) \le \varepsilon_1 \quad \text{for } i \in T.$$

Now define the event $\mathcal{E}' = (\neg\mathcal{E}) \vee \bigvee_{i\in T} \mathcal{E}_i$. Assume $\mathcal{P} = \sum_{i\in[t]} \alpha_i \cdot \mathcal{P}_i$ is the convex combination. Then

$$\mathcal{P}(\mathcal{E}') \ge \sum_{i\in T} \alpha_i \cdot \mathcal{P}_i(\mathcal{E}_i) + \sum_{i\notin T} \alpha_i \cdot \mathcal{P}_i(\neg\mathcal{E}) \ge (1-\varepsilon_1)\cdot\sum_{i\in T}\alpha_i + (1-\varepsilon_2)\cdot\sum_{i\notin T}\alpha_i \ge 1-\varepsilon_1-\varepsilon_2,$$

since $\sum_{i\in[t]} \alpha_i = 1$. In addition,

$$\mathcal{Q}(\mathcal{E}') \le \mathcal{Q}(\neg\mathcal{E}) + \sum_{i\in T} \mathcal{Q}(\mathcal{E}_i) \le \varepsilon_3 + t\cdot\varepsilon_1.$$

Then the desired bound follows from Fact 3.1. □

The next lemma shows that if two coupled random vectors are both individually $\gamma$-biased, they will still have Hamming weight mismatch (even modulo an integer) as long as parts of their entries are independent.

**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 $q \ge \min\{3, t+1\}$ be an integer.*[9] *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 are $\varepsilon$-close to $\mathcal{U}_\gamma^t$ for some $\gamma \in (0, 1/2]$*[10] *and*
$$\varepsilon \le \frac{\gamma}{4q} \cdot 2^{-50\gamma(t-1)/q^2}.$$

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

*Proof.* If $t = 1$ then we observe that $\mathbf{Pr}\left[X + |Y| = Z + |W|\right] = \mathbf{Pr}\left[X = Z\right]$ as $q \ge 2$. Since $X$ and $Z$ are independent and of the same distribution $\varepsilon$-close to $\mathcal{U}_\gamma^1$, we have $\mathbf{Pr}\left[X = 1\right] = \mathbf{Pr}\left[Z = 1\right] \in [\gamma-\varepsilon, \gamma+\varepsilon]$. Hence

$$\mathbf{Pr}\left[X = Z\right] = \mathbf{Pr}\left[X = 1\right]^2 + (1 - \mathbf{Pr}\left[X = 1\right])^2 \le (\gamma-\varepsilon)^2 + (1-\gamma+\varepsilon)^2 \le 1 - \gamma/2, \tag{1}$$

where we use the fact that $\gamma \in (0, 1/2]$ and $\varepsilon \le \gamma/2$.

Now we assume $t \ge 2$ and $q \ge 3$. Expand $\mathbf{Pr}\left[X + |Y| \equiv Z + |W| \pmod q\right]$ as

$$\sum_{x,z\in\{0,1\}} \mathbf{Pr}\left[X = x, Z = z\right] \mathbf{Pr}\left[x + |Y| \equiv z + |W| \pmod q \,\middle|\, X = x, Z = z\right]. \tag{2}$$

For fixed $x$ and $z$, consider the distribution of $x + |Y| \bmod q$ conditioned on $X = x, Z = z$. Since $Z$ is independent from $(X, Y)$, it is the same as the distribution, denoted by $\mathcal{P}_x$, of $x + |Y| \bmod q$

[9] If $q \ge t+1$, then one may instead apply Lemma 4.4 with modulus $t+1$, since $X + |Y| \equiv Z + |W| \pmod q$ is equivalent to $X + |Y| = Z + |W|$ for $q \ge t+1$.

[10] Lemma 4.4 holds for $\gamma \in [1/2, 1)$ as well, with $\gamma$ 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 $\varepsilon$-closeness to $\mathcal{U}_{1-\gamma}^t$ and preserves the congruence.

conditioned on $X = x$. Similarly define $\mathcal{Q}_z$ as the distribution of $z + |W| \bmod q$ conditioned on $Z = z$ (or equivalently, conditioned on $Z = z, X = x$).

Since $(X, Y)$ is $\varepsilon$-close to $\mathcal{U}_\gamma^t$, by Fact 4.1, $\mathcal{P}_0$ is $\frac{2\varepsilon}{1-\gamma}$-close to $\mathcal{D}_0$, the distribution of $|V| \bmod q$ for $V \sim \mathcal{U}_\gamma^{t-1}$. Similarly, $\mathcal{Q}_1$ is $\frac{2\varepsilon}{\gamma}$-close to $\mathcal{D}_1$, the distribution of $1 + |V| \bmod q$ for $V \sim \mathcal{U}_\gamma^{t-1}$. Hence

$$\begin{aligned}
\mathbf{Pr}\left[|Y| \equiv 1 + |W| \pmod q \mid X = 0, Z = 1\right] &\le 1 - \|\mathcal{P}_0 - \mathcal{Q}_1\|_{\mathsf{TV}} && \text{(by Fact 3.1)}\\
&\le 1 + \frac{2\varepsilon}{\gamma} + \frac{2\varepsilon}{1-\gamma} - \|\mathcal{D}_0 - \mathcal{D}_1\|_{\mathsf{TV}} && \text{(by Fact 4.1)}\\
&\le 1 + \frac{4\varepsilon}{\gamma} - \|\mathcal{D}_0 - \mathcal{D}_1\|_{\mathsf{TV}} && \text{(since } \gamma \le 1/2)\\
&\le 1 + \frac{4\varepsilon}{\gamma} - \frac{2}{q} \cdot 2^{-50\gamma(t-1)/q^2},
\end{aligned}$$

where we apply the following claim for the last inequality. Claim 4.5 is proved in Appendix C by Fourier analysis.

**Claim 4.5.** $\|\mathcal{D}_0 - \mathcal{D}_1\|_{\mathsf{TV}} \ge \frac{2}{q} \cdot 2^{-50\gamma(t-1)/q^2}$ for any $q \ge 3$.

By our assumption on $\varepsilon$, we now have

$$\mathbf{Pr}\left[|Y| \equiv 1 + |W| \pmod q \mid X = 0, Z = 1\right] \le 1 - \frac{1}{q} \cdot 2^{-50\gamma(t-1)/q^2}.$$

The same bound holds for $\mathbf{Pr}\left[1 + |Y| \equiv |W| \pmod q \mid X = 1, Z = 0\right]$. Plugging back into (2) and using (1), we can upper bound $\mathbf{Pr}\left[X + |Y| \equiv Z + |W| \pmod q\right]$ by

$$\mathbf{Pr}[X = Z] + \mathbf{Pr}[X \ne Z] \cdot \left(1 - \frac{1}{q} \cdot 2^{-50\gamma(t-1)/q^2}\right) \le 1 - \frac{\gamma}{2q} \cdot 2^{-50\gamma(t-1)/q^2}$$

as desired. □

We remark that Lemma 4.4 does not hold when $q = 2$ and $t \ge 2$, even if we assume $(X, Y) = (Z, W) = \mathcal{U}^t$ and $t = 2$: let $B$ be an unbiased coin independent from $X$ and $Z$. Then define $(X, Y, Z, W) = (X, X \oplus B, Z, Z \oplus B)$, and one can verify that this distribution satisfies all the conditions yet has $X + Y \equiv Z + W \pmod 2$ always.

### 4.2 Graph Elimination: Non-Connected Vertices

In this section we prove the graph theoretic results mentioned in Section 2 that aim to reduce a general $d$-local function to a more structured one: a $d$-local function with many non-connected output bits.

Recall the notation and terminology for bipartite graphs from Section 3. In particular, recall that $d$-left-bounded means each of the left vertices has degree at most $d$, and two left vertices are non-connected if they are not both adjacent to the same right vertex. We show that a $d$-left-bounded bipartite graph $G = ([n], [m], E)$ has many non-connected left vertices after removing few right vertices.

Let $\beta, \lambda \ge 1$ be parameters (not necessarily constant). We formalize the desired property as the following Property 4.6 with parameters $\beta, \lambda$.

**Property 4.6.** *There exists $S \subseteq [m]$ such that deleting those right vertices (and their incident edges) produces a bipartite graph with $r$ non-connected left vertices satisfying*

$$|S| \leq \frac{r}{\beta} \quad \textit{and} \quad r \geq \frac{n}{\lambda}.$$

Assuming Property 4.6 is false, we prove the following graph elimination result to show that we can remove many edges by deleting few right vertices. Later we will iteratively apply this with a proper choice of relation between $\beta$ and $\lambda$ to show that actually Property 4.6 always holds.

**Lemma 4.7.** *Assume Property 4.6 does not hold for a particular choice of (not necessarily constant) parameters $\beta, \lambda \geq 1$ and $d$-left-bounded bipartite graph $G = ([n], [m], E)$ with $d \geq 1$. Let $U \subseteq [m]$ be of size at most $n/\alpha$. Define*

$$s = \min\left\{n, \frac{\lambda}{2d}, \frac{\alpha}{2d\beta}\right\}. \tag{3}$$

*If $s \geq 1$, then there exists $V \subseteq [m] \setminus U$ of size at most $n/s$ and $\sum_{j \in V} \deg_G(j) \geq n/2$.*

The proof exploits that unless such a $V$ exists, there are many small neighborhoods. If so, we can find many non-connected left vertices by a simple greedy argument, contradicting to the assumption that Property 4.6 is false.

*Proof of Lemma 4.7.* We first remove right vertices (and their incident edges) in $U$ to obtain graph $G'$. Note that $\deg_G(j) = \deg_{G'}(j)$ holds for any $j \in [m] \setminus U$. Hence for simplicity we use $\deg(j)$ to denote both of them.

For each $i \in [n]$, we say $N_{G'}(i)$ is a small left neighborhood if every $j \in I_{G'}(i)$ satisfies $\deg(j) < s$. Since $G$ is $d$-left-bounded, each small left neighborhood has size less than $d \cdot s$. Let $A = \{j \in [m] \setminus U \colon \deg(j) \geq s\}$. Then each $j \in A$ prevents $\deg(j)$ left neighborhoods from being small, which means that the number of small left neighborhoods is at least $n - \sum_{j \in A} \deg(j)$. If $\sum_{j \in A} \deg(j) \geq n/2$, then we have the following cases:

- if $|A| \leq n/s$, then Lemma 4.7 follows by setting $V = A$,
- otherwise $|A| > n/s$. Then we pick an arbitrary set $V \subset A$ of size $\lfloor n/s \rfloor$. Since $\deg(j) > s$ for all $j \in A \supset V$, we have $\sum_{j \in V} \deg(j) > s \cdot \lfloor n/s \rfloor \geq n/2$ for any $0 < s \leq n$.

Now we assume $\sum_{j \in A} \deg(j) < n/2$, which means that we have at least $n/2$ small left neighborhoods. We will show that this cannot happen.

Observe that two left vertices in $G'$ are non-connected if and only if one is not in the left neighborhood of the other. Since $d, s \geq 1$, among the left vertices with small left neighborhoods, we can find at least

$$r = \underbrace{\frac{n}{2}}_{\#i\colon N_{G'}(i) \text{ is small}} \cdot \underbrace{\frac{1}{d \cdot s}}_{|N_{G'}(i)| < d \cdot s}$$

of them that are not connected to each other. If Property 4.6 is false, we must have

$$\frac{n}{\alpha} \geq |U| > \frac{r}{\beta} = \frac{n}{2ds\beta} \quad \text{or} \quad r = \frac{n}{2ds} < \frac{n}{\lambda}.$$

However by our choice (3) of $s$, we now have a contradiction. □

We now show that Property 4.6 always holds if $\lambda$ is not too small with respect to $\beta$ and $d$.

**Corollary 4.8.** *Let $\beta, \lambda \geq 1$ be parameters (not necessarily constant), and let $G = ([n], [m], E)$ be a $d$-left-bounded bipartite graph with $d \geq 1$. If*

$$\lambda \geq 2d \cdot (2d\beta + 1)^{2d},$$

*then Property 4.6 holds for $G$.*

*Proof.* If $n \leq \lambda$, then we can simply pick an arbitrary left vertex in $G$ and set $S = \emptyset, r = 1$ for Property 4.6. Now we assume $n \geq \lambda \geq 2d \cdot (2d\beta + 1)^{2d}$ and Corollary 4.8 is false.

We will apply Lemma 4.7 iteratively. For convenience, we define

$$\alpha_i = (2d\beta + 1)^{2d-i} \cdot 2d\beta \quad \text{and} \quad s_i = (2d\beta + 1)^{2d-i} \quad \text{for } i = 0, 1, \ldots, 2d.$$

Notice that for each $i = 0, 1, \ldots, 2d$, we have

$$s_i \geq 1 \quad \text{and} \quad \min\left\{n, \frac{\lambda}{2d}, \frac{\alpha_i}{2d\beta}\right\} = \frac{\alpha_i}{2d\beta} = s_i. \tag{4}$$

Let $U_0 = \emptyset$. For each $i = 0, 1, \ldots, 2d$, we apply Lemma 4.7 to $U_i$ with $\alpha_i$ to obtain $V_i$, then set $U_{i+1} = U_i \cup V_i$. Now we prove by induction that $|U_i| \leq n/\alpha_i$ and $|V_i| \leq n/s_i$, which establishes the validity of the above process. The base case $i = 0$ is $|U_i| = 0 \leq n/\alpha_i$ and $|V_i| \leq n/s_i$ by (4). For the inductive case $i \geq 1$, we have

$$|U_i| = |U_{i-1}| + |V_{i-1}| \leq \frac{n}{\alpha_{i-1}} + \frac{n}{s_{i-1}} = \frac{n}{\alpha_i}$$

by the induction hypothesis and our choice of $\alpha_{i-1}, s_{i-1}, \alpha_i$. Then the size bound on $|V_i|$ follows again from (4). This completes the induction.

Note that by Lemma 4.7, we have

$$\sum_{j \in U_{2d+1}} \deg_G(j) = \sum_{i=0}^{2d} \sum_{j \in V_i} \deg_G(j) \geq (2d+1) \cdot n/2,$$

contradicting the fact that $G$ has at most $d \cdot n$ edges, as it is $d$-left-bounded. Hence Corollary 4.8 must be true. □

We will apply this result in the proof of Proposition 5.6 to show that we can find many independent output bits of a $d$-local function by conditioning on only a few input bits. In addition, in Appendix A.1, we will show that the bound in Corollary 4.8 is essentially tight.

## 4.3 Graph Elimination: Non-Connected Neighborhoods

Similar to Subsection 4.2, here we aim to reduce a general $d$-local function to one having many non-connected neighborhoods of small size by deleting a few input bits. However the situation here is much more complicated than the one in Subsection 4.3, particularly because in later applications, we will impose different constraints between the number of input bits and the number of non-connected neighborhoods.

Let $\lambda, \kappa \geq 1$ be parameters (not necessarily constant) and $G = ([n], [m], E)$ be a $d$-left-bounded bipartite graph. Let $F(\cdot)$ be a function to be chosen based on later applications. We will require that $G$ has many non-connected left neighborhoods after removing few right vertices, formulated as the following property.

**Property 4.9.** *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*

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

Similar to the previous section, we prove the following graph elimination result, which shows, under the condition that Property 4.9 is false, we can remove many edges by deleting few right vertices.

**Lemma 4.10.** *Assume Property 4.9 does not hold for a particular choice of (not necessarily constant) parameters $\lambda, \kappa \geq 1$ and $d$-left-bounded bipartite graph $G = ([n], [m], E)$ with $d \geq 1$. Let $U \subseteq [m]$ be of size at most $n/\alpha$, and let $s$ be another parameter. If*

$$1 \leq s \leq \min\left\{n, \frac{\kappa}{d}\right\} \quad \text{and} \quad 1 \leq \alpha \leq 2\lambda \cdot F(d \cdot s) \quad \text{and} \quad \ln(\alpha \cdot d) \geq 8d^4 s^2 \cdot F(d \cdot s), \tag{5}$$

*then there exists $V \subseteq [m] \setminus U$ of size at most $n/s$ and $\sum_{j \in V} \deg_G(j) \geq n/2$.*

The proof is similar to the proof of Lemma 4.7 and exploits that unless such a $V$ exists, there are many small neighborhoods. If so, consider taking $S = \{v \in [m] : \deg(v) \geq \ell\}$ in Property 4.9. Since we removed the vertices of high degree, a small neighborhood cannot be connected to too many others. Hence, unless this $S$ satisfies Property 4.9, it must be that $S$ is large. However, this implies there are many right vertices of large degree, violating our total degree bound.

*Proof of Lemma 4.10.* We first remove right vertices (and their incident edges) in $U$ to obtain graph $G'$. Note that $\deg_G(j) = \deg_{G'}(j)$ holds for all $j \in [m] \setminus U$. Hence for simplicity we use $\deg(j)$ to denote both of them. For each $i \in [n]$, we say $N_{G'}(i)$ is a small left neighborhood if every $j \in I_{G'}(i)$ satisfies $\deg(j) < s$. Since $G$ is $d$-left-bounded, each small left neighborhood has size less than $d \cdot s$. By the same argument in the proof of Lemma 4.7, the lemma holds unless there are at least $n/2$ small left neighborhoods, so assume this to be the case. We will show this cannot happen.

Let $K$ be a parameter to determine later. For $1 \leq \ell \leq K$, let $B_\ell = \{j \in [m] \setminus U \colon \deg(j) \geq \ell\}$ be the set of right vertices with degree at least $\ell$. If we remove $B_\ell$ from $G'$ and obtain $H$, every small left neighborhood $N_H(i)$ is connected to

$$< \underbrace{d \cdot s}_{i' \in N_H(i)} \cdot \underbrace{d}_{j' \in I_H(i')} \cdot \underbrace{\ell}_{i'' \colon I_H(i'') \ni j'} \cdot \underbrace{d}_{j'' \in I_H(i'')} \cdot \underbrace{s}_{i''' \colon I_H(i''') \ni j''} = d^3 s^2 \ell$$

small left neighborhoods. Since there are at least $n/2$ small left neighborhoods and $d, s, \ell \geq 1$, we can find

$$r = \frac{n}{2d^3 s^2 \ell}$$

non-connected small left neighborhoods, each of which has size less than $d \cdot s$. Setting

$$K = \frac{\alpha}{4d^3 s^2 F(d \cdot s)},$$

we obtain

$$r \geq \frac{n}{2d^3 s^2 K} = \frac{2n \cdot F(d \cdot s)}{\alpha},$$

so (5) implies

$$t = d \cdot s \leq \kappa \quad \text{and} \quad r \geq \frac{n}{\lambda}. \tag{6}$$

If Property 4.9 is false, we must have

$$|U| + |B_\ell| > \frac{r}{F(t)} = \frac{n}{2d^3s^2F(d\cdot s)\cdot \ell},$$

since the other conditions are satisfied by (6). Therefore, by (5) and (6), we have

$$|B_\ell| > \frac{n}{2d^3s^2F(d\cdot s)\cdot \ell} - \frac{n}{\alpha} \geq \frac{n}{4d^3s^2F(d\cdot s)\cdot \ell}.$$

where the last inequality follows from

$$4d^3s^2F(d\cdot s)\cdot \ell \leq 4d^3s^2F(d\cdot s)\cdot K = \alpha.$$

Now we sum over[11] $1 \leq \ell \leq \lfloor K \rfloor$ and obtain

$$\sum_{1\leq \ell\leq \lfloor K\rfloor} |B_\ell| > \sum_{1\leq \ell\leq \lfloor K\rfloor} \frac{n}{4d^3s^2F(d\cdot s)\cdot \ell} \geq \frac{n}{4d^3s^2F(d\cdot s)} \int_1^K \frac{1}{\ell}\,\mathrm{d}\,\ell = \frac{n\cdot \ln(K)}{4d^3s^2F(d\cdot s)}.$$

Since $G'$ is $d$-left-bounded, we also have

$$\sum_{1\leq \ell\leq \lfloor K\rfloor} |B_\ell| \leq \sum_{\ell\geq 1} |\{j\in[m]\setminus U\colon\ \deg(j)\geq \ell\}| = \text{number of edges in } G' \leq d\cdot n.$$

At this point, we obtain the relation

$$d > \frac{\ln(K)}{4d^3s^2F(d\cdot s)} = \frac{1}{4d^3s^2F(d\cdot s)}\cdot \ln\left(\frac{\alpha}{4d^3s^2F(d\cdot s)}\right),$$

or equivalently, $K\cdot \ln(K) < \alpha d$. Since $\alpha, d \geq 1$, this implies $K < \frac{2\alpha\cdot d}{\ln(\alpha\cdot d)}$, i.e.,

$$\ln(\alpha\cdot d) < 8d^4s^2\cdot F(d\cdot s),$$

which contradicts (5). □

Similar to Corollary 4.8, we also show that Property 4.9 holds if $\lambda, \kappa$ are not too small with respect to $d$ and the function $F$.

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

*Define*

$$\widetilde{F}(x) = \frac{1}{d}\cdot \exp\left\{32d^4x^2\cdot F(2d\cdot x)\right\}. \tag{7}$$

*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 all $x \geq L$ and*

$$F(x) \geq 1 \text{ holds for all } x \geq 1 \quad \text{and} \quad \kappa \geq \lambda \geq d\cdot H^{(2d+2)}(L), \tag{8}$$

*where $H^{(k)}$ is the iterated $H$ of order $k$*[12]*, then Property 4.9 holds for $G$.*

[11] In the following inequality we do not need to assume $K \geq 1$, since otherwise LHS $= 0 >$ RHS already holds.

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

*Proof.* The proof is similar to the one for Corollary 4.8. If $n \le \lambda$, then we can simply pick an arbitrary left vertex in $G$ and set $S = \emptyset, r = 1, t = n$ for Property 4.9. Now we assume $n \ge \lambda$ and Corollary 4.11 is false.

We will apply Lemma 4.10 iteratively. For convenience, we define

$$\alpha_i = H^{(2d+2-i)}(L) \quad \text{and} \quad s_i = 2 \cdot H^{(2d+1-i)}(L) \quad \text{for } i = 0, 1, \ldots, 2d.$$

Since $H$ is increasing and $H(x) \ge 2x$ for $x \ge L$, we have

$$L \le H^{(2)}(L) = \alpha_{2d} \le \alpha_{2d-1} \le \cdots \le \alpha_0 = H^{(2d+2)}(L). \tag{9}$$

Similarly

$$2 \cdot L \le 2 \cdot H(L) = s_{2d} \le s_{2d-1} \le \cdots \le s_0 = 2 \cdot H^{(2d+1)}(L) \le H^{(2d+2)}(L). \tag{10}$$

Let $U_0 = \emptyset$. For each $i = 0, 1, \ldots, 2d$, we apply Lemma 4.10 to $U_i$ with $\alpha_i$ to obtain $V_i$ with $s_i$, then set $U_{i+1} = U_i \cup V_i$. To show the validity of the process, we first verify the following relations:

$$1 \le s_i \le \min\left\{n, \frac{\kappa}{d}\right\} \quad \text{and} \quad 1 \le \alpha_i \le 2\lambda \cdot F(d \cdot s_i) \quad \text{and} \quad \ln(\alpha_i \cdot d) \ge 8d^4 s_i^2 \cdot F(d \cdot s_i). \tag{11}$$

- The first one is due to $1 \le 2 \cdot L \le s_i \le 2 \cdot H^{(2d+1)}(L) \le \lambda \le n$ and $\kappa/d \ge H^{(2d+2)}(L) \ge s_i$ by (10) and (8).
- The second one is due to $1 \le L \le \alpha_i \le H^{(2d+2)}(L) \le \lambda$ and $F(d \cdot s_i) \ge 1$ as $d \cdot s_i \ge 1$ by (9) and (8).
- The third one is equivalent to verifying

  $$\alpha_i \ge \frac{1}{d} \cdot \exp\left\{8d^4 s_i^2 \cdot F(d \cdot s_i)\right\} = \widetilde{F}(s_i/2),$$

  where we recall the definition of $\widetilde{F}$ from (7). Since $H(x) \ge \widetilde{F}(x)$ for $x \ge L$ and $s_i/2 \ge L$ from (10), we have

  $$\widetilde{F}(s_i/2) \le H(s_i/2) = H^{(2d+2-i)}(L) = \alpha_i$$

  as desired.

Given (11), we prove by induction that $|U_i| \le n/\alpha_i$ and $|V_i| \le n/s_i$. The base case $i = 0$ is $|U_i| = 0 \le n/\alpha_i$ and $|V_i| \le n/s_i$ by (11). For the inductive case $i \ge 1$, we have

$$|U_i| = |U_{i-1}| + |V_{i-1}| \le \frac{n}{\alpha_{i-1}} + \frac{n}{s_{i-1}} = \frac{n}{H^{(2d+3-i)}(L)} + \frac{n}{2 \cdot H^{(2d+2-i)}(L)} \le \frac{n}{H^{(2d+2-i)}(L)} = \frac{n}{\alpha_i},$$

where the second inequality used the fact that $H(x) \ge 2x$ for $x \ge L$. Then the size bound on $|V_i|$ follows again from (11). This completes the induction. Finally we obtain the same contradiction from the total number of edges as in the proof of Corollary 4.8. Hence Corollary 4.11 must be true. □

Observe that in Corollary 4.11, 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 $\kappa$ and $\lambda$ will (at least) be a tower-type blowup in $d$. We emphasize that this is surprisingly inevitable and will be elaborated in Appendix A.2.

# 5 Lower Bounds

In this section, we prove lower bounds for a variety of distributions related to Hamming slices. Subsection 5.1 contains lower bounds for $\gamma$-biased distributions. Subsection 5.2 contains lower bounds for single Hamming slices of weight $\gamma n$ when $\gamma$ has binary representation error, and in Subsection 5.3, we extend the analysis to the general case of $\gamma$. We conclude by proving lower bounds for sampling from periodic Hamming slices in Subsection 5.4.

## 5.1 Biased Distributions

Recall that $\mathcal{D}_\gamma^n$ is the $\gamma$-biased distribution on $\{0,1\}^n$. In this section we show that if $\gamma$ is not close to a dyadic number, then local functions cannot produce distributions close to $\mathcal{D}_\gamma^n$. After proving this result, we learned it is implicit in [Vio12a, Vio23]. However, we do not find references explicitly giving such bounds, and the techniques used in proving this result will be generalized to other cases. Therefore we include a complete proof here.

**Definition 5.1** (Binary Representation Error)**.** For each $t \in \mathbb{N}$, we use $\mathsf{err}(\gamma, t)$ to denote the minimum distance of $\gamma$ to an integer multiple of $2^{-t}$. In particular, given a binary representation of $\gamma$ as $\gamma = \sum_{i\in\mathbb{Z}} a_i \cdot 2^i$ where each $a_i \in \{0,1\}$, we have

$$\mathsf{err}(\gamma, t) = \min\left\{\sum_{i<-t} a_i \cdot 2^i, \sum_{i<-t} (1-a_i)\cdot 2^i\right\}.$$

It is easy to see that $0 \le \mathsf{err}(\gamma, t) \le 2^{-t-1}$. A concrete non-trivial example is $\mathsf{err}(1/3, t) \ge 2^{-t-2}$ for all $t \ge 0$, since $1/3$ has binary representation $\sum_{i<0} 2^{2i}$.

**Fact 5.2.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function. Then the marginal distribution of $f(\mathcal{U}^m)$ on any single output bit is $\mathsf{err}(\gamma, d)$-far from $\mathcal{U}_\gamma^1$.*

Fact 5.2 already shows that $f(\mathcal{U}^m)$ is $\mathsf{err}(\gamma, d)$-far from $\mathcal{U}_\gamma^n$. Our goal is to boost the distance closer to 1.

**Theorem 5.3.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function, and let $0 \le \gamma \le 1$ be a parameter. If $\mathsf{err}(\gamma, d) \ge \delta$ for some $\delta > 0$, then*

$$\left\|f(\mathcal{U}^m) - \mathcal{U}_\gamma^n\right\|_{\mathsf{TV}} \ge 1 - 4\cdot\exp\left\{-n\cdot\delta^{40d}\right\}.$$

We first consider a simple case where we can find many output bits that do not correlate with each other. For example, this happens when every input bit influences few output bits.

**Definition 5.4** ($(d,r)$-Local Function)**.** We say $g\colon \{0,1\}^m \to \{0,1\}^n$ is a $(d,r)$-local function if $g$ is a $d$-local function with $r$ non-connected output bits.

**Proposition 5.5.** *Let $g\colon \{0,1\}^m \to \{0,1\}^n$ be a $(d,r)$-local function. Then*

$$\left\|g(\mathcal{U}^m) - \mathcal{U}_\gamma^n\right\|_{\mathsf{TV}} \ge 1 - 2\cdot\exp\left\{-\frac{\mathsf{err}(\gamma,d)^2\cdot r}{2}\right\}.$$

*Proof.* The bound is trivial when $r < 1$. Hence we assume $r \ge 1$. By rearranging indices, we assume without loss of generality that $1, 2, \ldots, r$ are non-connected output bits. We will apply Lemma 4.2 with

$$\mathcal{P} = g(\mathcal{U}^m), \quad \mathcal{Q} = \mathcal{W} = \mathcal{U}_\gamma^n, \quad S = [r].$$

Observe that $\mathcal{P}$ is a product distribution marginally on $S$, since $1, 2, \ldots, r$ are non-connected. Additionally by Fact 5.2, we have

$$\left\|\mathcal{P}|_{\{i\}} - \mathcal{W}|_{\{i\}}\right\|_{\mathsf{TV}} \geq \mathsf{err}(\gamma, d) =: \varepsilon.$$

Then the desired bound follows from Lemma 4.2 with $\eta = 1$. □

We next show that any $d$-local function $f$ can be made $(d, r)$-local by restricting a few input bits.

**Proposition 5.6.** *Assume* $\mathsf{err}(\gamma, d) > 0$. *Let* $f\colon \{0,1\}^m \to \{0,1\}^n$ *be a* $d$*-local function with* $d \geq 1$. *Then there exists a set* $S \subseteq [m]$ *such that any fixing of input bits in* $S$ *reduces* $f$ *to a* $(d, r)$*-local function* $g$*, where*

$$|S| \leq \frac{\mathsf{err}(\gamma, d)^2 \cdot r}{4} \quad \text{and} \quad r \geq n \cdot \left(\frac{\mathsf{err}(\gamma, d)^2}{16d}\right)^{2d+1}.$$

*Proof.* Recall the graph theoretic view of the dependency relations in $f$. We apply Corollary 4.8 with $\beta = 4/\mathsf{err}(\gamma, d)^2$ and $\lambda = (4d\beta)^{2d+1}$. □

Now we prove Theorem 5.3.

*Proof of Theorem 5.3.* Recall that $\mathsf{err}(\gamma, d) \geq \delta > 0$ and $\mathsf{err}(\gamma, d) \leq 2^{-d-1}$. We assume $d \geq 1$, as otherwise the theorem is trivial. By Proposition 5.6, we find a set $S \subseteq [m]$ such that any fixing $\rho$ of input bits in $S$ reduces $f$ to a $(d, r)$-local function $f_\rho$ where

$$|S| \leq \frac{\delta^2 \cdot r}{4} \quad \text{and} \quad r \geq \frac{n}{(16d/\delta^2)^{2d+1}}.$$

Then for each $f_\rho$, we apply Proposition 5.5 and obtain that

$$\left\|f_\rho(\mathcal{U}^{[m]\setminus S}) - \mathcal{U}_\gamma^n\right\|_{\mathsf{TV}} \geq 1 - 2 \cdot e^{-\delta^2 \cdot r/2}.$$

Note that $f(\mathcal{U}^m) = \mathbb{E}_\rho\left[f_\rho(\mathcal{U}^{[m]\setminus S})\right]$ where $\rho \sim \mathcal{U}^S$. By Lemma 4.3 with $\left\{f_\rho(\mathcal{U}^{[m]\setminus S})\right\}_\rho, \mathcal{U}_\gamma^n$, and

$$\varepsilon_1 = 2 \cdot e^{-\delta^2 \cdot r/2}, \quad \varepsilon_2 = \varepsilon_3 = 0, \quad \mathcal{E} = \emptyset,$$

we have

$$\begin{aligned}
\left\|f(\mathcal{U}^m) - \mathcal{U}_\gamma^n\right\|_{\mathsf{TV}} &\geq 1 - \left(2^{|S|} + 1\right) \cdot 2 \cdot e^{-\delta^2 \cdot r/2} \geq 1 - 4 \cdot e^{-\delta^2 \cdot r/4} \\
&\geq 1 - 4 \cdot \exp\left\{-\frac{\delta^2 \cdot n}{4 \cdot (16d/\delta^2)^{2d+1}}\right\} \\
&\geq 1 - 4 \cdot \exp\left\{-n \cdot \delta^{40d}\right\} \qquad \text{(since } d \geq 1 \text{ and } \delta \leq 2^{-d-1}\text{)}
\end{aligned}$$

as desired. □

### 5.2 A Single Hamming Slice: The Non-Dyadic Case

The argument in Subsection 5.1 works beyond $\gamma$-biased distributions. Here we generalize it to the Hamming slice setting, the proof of which introduces new ideas to handle non-product distributions and will be useful later.

Let $\mathcal{D}_k$ be the uniform distribution of binary strings of length $n$ with Hamming weight $k$. Define $\gamma = k/n$. Our goal here is to prove local function cannot sample from $\mathcal{D}_k$ when $\gamma$ has large binary representation error. This is similar to Theorem 5.3 which replaces $\mathcal{D}_k$ by the $\gamma$-biased distribution.

**Theorem 5.7.** *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function, and let $1 \le k \le n-1$ be an integer. Define $\gamma = k/n$. If $\mathsf{err}(\gamma, d) \ge \delta$ for some $\delta > 0$, then*

$$\|f(\mathcal{U}^m) - \mathcal{D}_k\|_{\mathsf{TV}} \ge 1 - 4\sqrt{2n} \cdot \exp\left\{-n \cdot \delta^{40d}\right\}.$$

Following the previous framework, we first prove the distance bound for $(d,r)$-local functions analogous to Proposition 5.5. However unlike $\mathcal{U}_\gamma^n$ there, we have $\mathcal{D}_k$ here. Though marginally every bit in $\mathcal{D}_k$ is exactly $\gamma$-biased $\mathcal{U}_\gamma^1$, the distance between $\mathcal{D}_k|_S$ and $\mathcal{U}_\gamma^S$ enlarges quickly when $|S|$ grows. Nevertheless, we can use $\mathcal{U}_\gamma^n$ as a proxy between our distribution and $\mathcal{D}_k$. The crucial point is that $\mathcal{U}_\gamma^n$ and $\mathcal{D}_k$ are not far from each other in the multiplicative sense, though they have total variation distance roughly $1 - 1/\sqrt{n}$ (for constant $\gamma$).

**Proposition 5.8.** *Let $g\colon \{0,1\}^m \to \{0,1\}^n$ be a $(d,r)$-local function. Then*

$$\|g(\mathcal{U}^m) - \mathcal{D}_k\|_{\mathsf{TV}} \ge 1 - 2\sqrt{2n} \cdot \exp\left\{-\frac{\mathsf{err}(\gamma,d)^2 \cdot r}{2}\right\}.$$

*Proof.* The bound is trivial when $r < 1$. Hence we assume $r \ge 1$. By rearranging indices, we assume without loss of generality that $1, 2, \ldots, r$ are non-connected output bits. We will apply Lemma 4.2 with

$$\mathcal{P} = g(\mathcal{U}^m), \quad \mathcal{Q} = \mathcal{D}_k, \quad \mathcal{W} = \mathcal{U}_\gamma^n, \quad S = [r].$$

Note that $\mathcal{P}$ is a product distribution on $S$. By Fact 5.2, we have

$$\left\|\mathcal{P}|_{\{i\}} - \mathcal{W}|_{\{i\}}\right\|_{\mathsf{TV}} \ge \mathsf{err}(\gamma, d) =: \varepsilon.$$

To apply Lemma 4.2, it remains to bound $\eta := \min_{x \in \mathsf{supp}(\mathcal{Q})} \mathcal{W}(x)/\mathcal{Q}(x)$. For any $x \in \mathsf{supp}\,(\mathcal{Q})$, we have

$$\mathcal{Q}(x) = \frac{1}{\binom{n}{k}} = \frac{1}{\binom{n}{\gamma n}} \quad \text{and} \quad \mathcal{W}(x) = \gamma^k \cdot (1-\gamma)^{n-k} = \gamma^{\gamma n} \cdot (1-\gamma)^{(1-\gamma)n}.$$

By Fact 3.5, we have

$$\eta = \min_{x \in \mathsf{supp}(\mathcal{Q})} \frac{\mathcal{W}(x)}{\mathcal{Q}(x)} \ge \frac{1}{\sqrt{8\gamma n(1-\gamma)}} \ge \frac{1}{\sqrt{2n}}.$$

Applying Lemma 4.2 gives the desired bound. □

**Remark 5.9.** It is possible to improve the construction of $\mathcal{W}$ in Proposition 5.8 to get a better bound of $\eta$. To see this, we can obtain $\mathcal{W}$ as follows: first we sample bits in $S$ according to $\mathcal{U}_\gamma^S$, then we complete the other coordinates $[n] \setminus S$ by the distribution of $\mathcal{Q} = \mathcal{D}_k$ conditioned on the sampled bits in $S$.

As such, $\eta$ will be the minimum ratio of $\mathcal{U}_\gamma^S(x)$ and $\mathcal{D}_k|_S(x)$ for $x \in \{0,1\}^S$. Note that if $|S|$ is not too large, then $\eta$ will not be too small. For example, if $\gamma$ is constant and $|S| \ll n$, then one can show that $\eta \geq \Omega(1)$. Since later in Proposition 5.6 we indeed have relatively small $|S|$, this is the typical case that matters to us.

We choose our current presentation for simplicity. Moreover this improvement is only a factor of $\mathsf{poly}(n)$ in terms of applications after combining everything, ultimately subsumed by $\exp\{-\Omega_d(n)\}$.

We use the same Proposition 5.6 to convert $d$-local functions to $(d, r)$-local, and apply Lemma 4.3 to put them together.

*Proof of Proposition 5.8.* The argument is almost identical to the proof of Theorem 5.3, except that now we have

$$\varepsilon_1 = 2\sqrt{2n} \cdot e^{-\delta^2 \cdot r/2}.$$

Combining Proposition 5.6 and Lemma 4.3, we have

$$\|f(\mathcal{U}^m) - \mathcal{D}_k\|_{\mathsf{TV}} \geq 1 - 4\sqrt{2n} \cdot \exp\left\{-n \cdot \delta^{40d}\right\}. \qquad \square$$

## 5.3 A Single Hamming Slice: The General Case

In the previous section, we showed that local functions cannot sample from $\mathcal{D}_k$ when $\gamma = k/n$ has large binary representation error. In particular, this shows $\mathcal{D}_{n/3}$ is not locally sampleable. In this section, we aim to address the general case of $\mathcal{D}_k$, where we do not gain advantages from the non-dyadic numbers.

A concrete example is $k = n/2$. The coordinatewise-independent version of $\mathcal{D}_{n/2}$ is simply $\mathcal{U}^n$, which can be exactly sampled by a 1-local function. However this does not seem to generalize: $\mathcal{D}_{n/2}$ is $(1 - \Theta(1/\sqrt{n}))$-far from $\mathcal{U}^n$. We will show that this is actually the best possible strategy.

We will prove the following more general statement, which works for all $o(n) \leq k \leq n/2$. To build intuition, specific instantiations can be found in Theorem 1.1 and Theorem 1.2.

**Theorem 5.10.** *There exists a universal constant $\kappa \geq 1$ such that the following holds. Let*[13] *$1 \leq k \leq n/2$ and let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function. Define $\gamma = k/n$ and let $\theta(n)$ be arbitrary. If $\theta(n) \geq \kappa$ and*

$$d \leq \log^*(\theta(n))/60 \quad \text{and} \quad \log^*(1/\gamma) \leq \log^*(\theta(n))/2,$$

*then*

$$\|f(\mathcal{U}^m) - \mathcal{D}_k\|_{\mathsf{TV}} \geq 1 - \theta(n)/\sqrt{n}.$$

We remark that the analysis in this section generalizes to multiple Hamming slices, with a loss of the union bound on top of the $\sqrt{n}$ that scales linearly with the number of slices. This will be clear in the anticoncentration analysis (Lemma 5.15) which works equally well in the generalized setting.

In addition, the bounds in Theorem 5.10 are asymptotically tight when we set $\theta(n)$ to be a fixed constant sufficiently large. This is because the $2^{-t}$-biased distribution over $n$ bits is $(1 - \Theta_t(1/\sqrt{n}))$-close to $\mathcal{D}_{n/2^t}$, where the former can be sampled by a $t$-local function.

To prove Theorem 5.10, we first consider a simpler setting where we are given a $d$-local function with many non-connected neighborhoods of small size. For intuition, one can view it as the case where every input bit influences few output bits.

[13] By flipping zero and one, sampling from $\mathcal{D}_k$ is equivalent to sampling from $\mathcal{D}_{n-k}$. Therefore Theorem 5.10 also holds for $k \geq n/2$ with $\gamma$ replaced by $1 - \gamma$.

**Definition 5.11** ($(d,r,t)$-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 $t$.

We remark that the notion of $(d,r,t)$-local generalizes the notion of $(d,r)$-local in the previous section. There, the analysis depended on individual bits having incorrect bias, but now we consider a more subtle exploitation. In particular, we need both the distribution over the neighborhood $N(i)$ to be close to uniform and resampling to not substantially change the sum. (This latter property implies the distribution of the sum conditioned on $i=0$ is approximately the same as conditioned on $i=1$.) However, this trade-off alone gives only a small error in total variation distance to amplify; we need many independent neighborhoods (rather than just independent bits). This independence follows from non-connectedness.

The following proposition concerns lower bounds for $(d,r,t)$-local functions. It is similar to Proposition 5.8 and will be proved later.

**Proposition 5.12.** *Let $g\colon \{0,1\}^m \to \{0,1\}^n$ be a $(d,r,t)$-local function and define $\mathcal{P}_g = g(\mathcal{U}^m)$. Then there exists a universal constant $C \ge 1$ such that either*

$$\|\mathcal{P}_g - \mathcal{D}_k\|_{\mathsf{TV}} \ge 1 - C\sqrt{n}\cdot\exp\left\{-\frac{\gamma^2\cdot r}{C\cdot t}\right\} \quad \textit{or} \quad \mathcal{P}_g(\mathsf{supp}\,(\mathcal{D}_k)) \le \frac{C\cdot t^{1/4}}{\sqrt{\gamma\cdot r}}.$$

Then we show that any $d$-local function $f$ can be turned into a $(d,r,t)$-local function $g$ by restricting a few input bits. Note that this is for some values of $r$ and $t$, which might depend on the function $f$. For intuition, one can think of $d,\gamma,C$ as constants, then we will obtain $r=\Omega(n)$ non-connected neighborhoods of size at most $t=O(1)$ by restricting way fewer than $r$ input bits. This result is similar to Proposition 5.6.

**Proposition 5.13.** *Let $C \ge 1$ be an integer parameter and let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function with $d \ge 1$. Then 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*

$$|S| \le \frac{\gamma^2\cdot r}{2C\cdot t} \quad \textit{and} \quad r \ge \frac{n}{\mathrm{tow}_2(20d+\log^*(1/\gamma)+C)} \quad \textit{and} \quad t \le \mathrm{tow}_2(20d+\log^*(1/\gamma)+C).$$

*Proof.* Recall the graph theoretic view of the dependency relations in $f$. We will apply Corollary 4.11. Setting $F(x) = 2C\cdot x/\gamma^2$ gives

$$\widetilde{F}(x) = \frac{1}{d}\cdot\exp\left\{\frac{128d^5C\cdot x^3}{\gamma^2}\right\}.$$

Define $H(x) = 2^{2^x}$ and let $L = 10\cdot\log(d) + 30\cdot\log(1/\gamma) + 2\cdot\log(C)$. By setting

$$\kappa = \lambda = \mathrm{tow}_2(20d+\log^*(1/\gamma)+C) \ge d\cdot H^{(2d+2)}(L),$$

the conditions in Corollary 4.11 are satisfied, where we used the fact that $\gamma \le 1/2$ and $d,C \ge 1$. This implies that Property 4.9 holds for the dependency graph of $f$ with parameter $\lambda,\kappa,F$. □

Finally, we convert lower bounds for $(d,r,t)$-local functions to $d$-local functions via Lemma 4.3 as before.

*Proof of Theorem 5.10.* Let $C \ge 1$ be the universal constant in Proposition 5.12. Without loss of generality we assume it is an integer. Define $\kappa = \mathrm{tow}_2(60C)$. If $d \le C$, then we can simply set $d = C$, since a $d'$-local function is also $d$-local if $d \geq d'$. Since we assumed that $\theta(n) \geq \kappa = \mathsf{tow}_2(60C)$, setting $d = C$ still satisfies the condition $d \leq \log^*(\theta(n))/60$. From now on we safely assume $d \geq C$.

By Proposition 5.13, we find a set $S \subseteq [m]$ such that any fixing $\rho$ of input bits in $S$ reduces $f$ to a $(d, r, t)$-local function $f_\rho$ where

$$|S| \leq \frac{\gamma^2 \cdot r}{2C \cdot t} \quad \text{and} \quad r \geq \frac{n}{\mathsf{tow}_2(20d + \log^*(1/\gamma) + C)} \quad \text{and} \quad t \leq \mathsf{tow}_2(20d + \log^*(1/\gamma) + C).$$

Since a $(d, r, t')$-local function is also $(d, r, t)$-local if $t \geq t'$, we may assume $t = \mathsf{tow}_2(20d+\log^*(1/\gamma)+C)$. Now for each $f_\rho$, we apply Proposition 5.12 and obtain that

$$\text{either} \quad \|\mathcal{P}_{f_\rho} - \mathcal{D}_k\|_{\mathsf{TV}} \geq 1 - C\sqrt{n} \cdot \exp\left\{-\frac{\gamma^2 \cdot r}{C \cdot t}\right\} \quad \text{or} \quad \mathcal{P}_{f_\rho}(\mathsf{supp}\,(\mathcal{D}_k)) \leq \frac{C \cdot t^{1/4}}{\sqrt{\gamma \cdot r}}.$$

Note that $f(\mathcal{U}^m) = \mathbb{E}_\rho\left[f_\rho(\mathcal{U}^{[m]\setminus S})\right] = \mathbb{E}_\rho[\mathcal{P}_{f_\rho}]$ where $\rho \sim \mathcal{U}^S$. By Lemma 4.3 with $\left\{\mathcal{P}_{f_\rho}\right\}_\rho, \mathcal{D}_k$, and

$$\varepsilon_1 = C\sqrt{n} \cdot \exp\left\{-\frac{\gamma^2 \cdot r}{C \cdot t}\right\}, \quad \varepsilon_2 = \frac{C \cdot t^{1/4}}{\sqrt{\gamma \cdot r}}, \quad \varepsilon_3 = 0, \quad \mathcal{E} = \mathsf{supp}\,(\mathcal{D}_k)\,,$$

we have

$$\begin{aligned}
\|f(\mathcal{U}^m) - \mathcal{D}_k\|_{\mathsf{TV}} &\geq 1 - \left(2^{|S|} + 1\right) \cdot C\sqrt{n} \cdot \exp\left\{-\frac{\gamma^2 \cdot r}{C \cdot t}\right\} - \frac{C \cdot t^{1/4}}{\sqrt{\gamma \cdot r}} \\
&\geq 1 - 2C\sqrt{n} \cdot \exp\left\{-\frac{\gamma^2 \cdot r}{2C \cdot t}\right\} - \frac{C \cdot t^{1/4}}{\sqrt{\gamma \cdot r}} && \left(\text{since } |S| \leq \tfrac{\gamma^2 \cdot r}{2C \cdot t}\right) \\
&\geq 1 - 2C\sqrt{n} \cdot \exp\left\{-\frac{n}{t^3}\right\} - \frac{t^2}{\sqrt{n}} && \left(\text{since } r \geq \tfrac{2Cn}{\gamma^2 \cdot t^2} \text{ and } r \geq \tfrac{C^2 n}{\gamma^2 \cdot t^2}\right) \\
&\geq 1 - \frac{3C \cdot t^3}{\sqrt{n}}. && (\text{since } e^{-x} \leq 1/x)
\end{aligned}$$

Observe that

$$\begin{aligned}
t &\leq \mathsf{tow}_2(21d + \log^*(1/\gamma)) && (\text{since } d \geq C) \\
&\leq \mathsf{tow}_2(\lfloor \log^*(\theta) \cdot 0.9 \rfloor) && (\text{since } d \leq \log^*(\theta)/60 \text{ and } \log^*(1/\gamma) \leq \log^*(\theta)/2) \\
&\leq \mathsf{tow}_2(\log^*(\theta) - 5) && (\text{since } \theta \geq \kappa = \mathsf{tow}_2(60C) \geq \mathsf{tow}_2(60)) \\
&= \mathsf{tow}_2(\log^*(\log^{(5)}(\theta))) \leq 2^{\log^{(5)}(\theta)} && (\text{since } \mathsf{tow}_2(\log^*(x)) \leq 2^x) \\
&\leq (\theta/3C)^{1/3}. && (\text{since } \theta \geq \mathsf{tow}_2(60C))
\end{aligned}$$

Hence $\|f(\mathcal{U}^m) - \mathcal{D}_k\|_{\mathsf{TV}} \geq 1 - \theta/\sqrt{n}$ as desired. □

Now we prove lower bounds for $(d, r, t)$-local functions.

**Proposition** (Proposition 5.12 Restated)**.** *Let $g\colon \{0,1\}^m \to \{0,1\}^n$ be a $(d, r, t)$-local function and define $\mathcal{P}_g = g(\mathcal{U}^m)$. Then there exists a universal constant $C \geq 1$ such that either*

$$\|\mathcal{P}_g - \mathcal{D}_k\|_{\mathsf{TV}} \geq 1 - C\sqrt{n} \cdot \exp\left\{-\frac{\gamma^2 \cdot r}{C \cdot t}\right\} \quad \textit{or} \quad \mathcal{P}_g(\mathsf{supp}\,(\mathcal{D}_k)) \leq \frac{C \cdot t^{1/4}}{\sqrt{\gamma \cdot r}}.$$

Recall that $\gamma = k/n \in [1/n, 1/2]$. Let $\varepsilon \in [0, 1]$ be a parameter to be optimized later. For each neighborhood $N(i) = N_g(i)$ of size $s_i = |N(i)|$, we classify it into one of the following two cases:

- $\mathsf{Type\text{-}1}$. $\mathcal{P}_g|_{N(i)}$ is not $\varepsilon$-close to $\mathcal{U}_\gamma^{s_i}$.
- $\mathsf{Type\text{-}2}$. $\mathcal{P}_g|_{N(i)}$ is $\varepsilon$-close to $\mathcal{U}_\gamma^{s_i}$.

Intuitively a $\mathsf{Type\text{-}1}$ neighborhood means the marginal $\mathcal{P}_g$ on $N(i)$ is far from the $\gamma$-biased distribution. If we find many $\mathsf{Type\text{-}1}$ neighborhoods, we can prove the distance bound analogous to Proposition 5.8.

On the other hand, the output distribution of a $\mathsf{Type\text{-}2}$ neighborhood is close to $\gamma$-biased, in which case the previous argument fails. Then we show that these neighborhoods are somewhat independent and it is unlikely for them to sum to a fixed value.

We first prove that $\mathcal{P}_g$ is far from $\mathcal{D}_k$ if there are many small non-connected $\mathsf{Type\text{-}1}$ neighborhoods. The intuition is that the local view of $\mathcal{D}_k$ should be $\gamma$-biased.

**Lemma 5.14.** *Assume there are $r' \ge 1$ non-connected $\mathsf{Type\text{-}1}$ neighborhoods. Then*

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

*Proof.* The proof is similar to the one for Proposition 5.8. The only change is to work with non-connected neighborhoods instead of non-connected output bits. By rearranging indices, we assume without loss of generality that $N(1), \ldots, N(r')$ are non-connected $\mathsf{Type\text{-}1}$ neighborhoods of sizes $s_1, \ldots, s_{r'}$.

We will apply Lemma 4.2, for which we define $\mathcal{P}, \mathcal{Q}, S, \mathcal{W}$. Let $R = [n] \setminus (N(1) \cup \cdots N(r'))$ be the rest of the output coordinates.

- Define $\mathcal{P}$ as $\mathcal{P}_g$ but grouping each $N(i)$ for $i \in [r']$ and $R$ as coordinates. That is, $\mathcal{P}$ now is a distribution over a product space of $r'+1$ coordinates, where $\mathcal{P}|_{\{i\}} = \mathcal{P}_g|_{N(i)}$ is over $\{0,1\}^{s_i}$ for $i \in [r']$ and $\mathcal{P}|_{\{r'+1\}} = \mathcal{P}_g|_R$.
- Define $\mathcal{Q}$ as $\mathcal{D}_k$ but also grouping each $N(i)$ for $i \in [r']$ and $R$ as coordinates.
- Define $S = [r']$.
- Define $\mathcal{W}$ as $\mathcal{U}_\gamma^n$ but also grouping each $N(i)$ for $i \in [r']$ and $R$ as coordinates.

Observe that $\mathcal{W}|_S = \bigtimes_{i \in S} \mathcal{U}_\gamma^{s_i}$ is a product distribution and $\mathcal{P}|_S = \bigtimes_{i \in S} \mathcal{P}_g|_{N(i)}$ is also a product distribution since $N(1), \ldots, N(r')$ are non-connected.

Since each $N(i)$ here is $\mathsf{Type\text{-}1}$, we have $\left\|\mathcal{P}|_{\{i\}} - \mathcal{W}|_{\{i\}}\right\|_{\mathsf{TV}} \ge \varepsilon$. Note that $\mathcal{W}$ is the $\gamma$-biased distribution with $\gamma \in [1/n, 1/2]$. Therefore for any $x \in \mathsf{supp}(\mathcal{Q})$, we have

$$\frac{\mathcal{W}(x)}{\mathcal{Q}(x)} = \gamma^{\gamma n} \cdot (1-\gamma)^{(1-\gamma)n} \cdot \binom{n}{\gamma n},$$

where we recall that $\mathcal{Q} = \mathcal{D}_k = \mathcal{D}_{\gamma n}$. Then by the same calculation in the proof of Proposition 5.8, we can set $\eta = 1/\sqrt{2n}$ in Lemma 4.2 and obtain the desired bound. □

We note that the same improvement idea described in Remark 5.9 also works here. Now we turn to the second case where there are many small non-connected $\mathsf{Type\text{-}2}$ neighborhoods. In this case, we show that with high probability the sampled binary string from $\mathcal{P}_g$ does not have Hamming weight $k$ via anticoncentration inequalities.

**Lemma 5.15.** *Assume there are $r' \ge 1$ non-connected $\mathsf{Type\text{-}2}$ neighborhoods of size at most $t$. If $\varepsilon \le \frac{\gamma}{128\sqrt{t}}$, then*

$$\mathcal{P}_g(\mathsf{supp}(\mathcal{D}_k)) \le O\left(\frac{t^{1/4}}{\sqrt{\gamma \cdot r'}}\right).$$

*Proof.* By rearranging indices, we assume without loss of generality that $N(1),\dots,N(r')$ are non-connected $\mathsf{Type\text{-}2}$ neighborhoods of sizes $1 \le s_1,\dots,s_{r'} \le t$. Recall that $I(i) = I_g(i)$ is the set of input bits that the $i$-th output bit depends on. We sample a random $Z \sim \mathcal{U}^m$ and let $(X_1,\dots,X_n) = g(Z)$.

Let $R = [m] \setminus (I(1) \cup \cdots \cup I(r'))$. Since $(X_1,\dots,X_n)$ has its marginal equal to $\mathcal{P}_g$, we have

$$\mathcal{P}_g(\mathsf{supp}\,(\mathcal{D}_k)) = \mathop{\mathbb{E}}_{\rho}\left[\mathbf{Pr}\left[\sum_{i\in[n]} X_i = k \,\middle|\, \rho\right]\right], \tag{12}$$

where $\rho \sim \{0,1\}^R$ and the condition on $\rho$ means that $Z_j = \rho_j$ for all $j \in R$. We will use Fact 3.4 to upper bound the above probability for most $\rho$'s. To this end, we decompose $\sum_{i\in[n]} X_i$ into $K + \sum_{\ell\in[r']} \Delta_\ell$, where

$$K = \sum_{i \notin N(1)\cup\cdots\cup N(r')} X_i \quad\text{and}\quad \Delta_\ell = \sum_{i\in N(\ell)} X_i.$$

Observe that if $i \notin N(1) \cup \cdots \cup N(r')$, then $I(i) \subseteq R$ and thus $K$ is fixed given $\rho$. For each $\ell \in [r']$ and $\rho \in \{0,1\}^R$, define the random variable

$$p_{\rho,\ell} = \max_c \mathbf{Pr}\left[\Delta_\ell = c \,|\, \rho\right].$$

We will use Lemma 4.4 to prove an upper bound on $p_{\rho,\ell}$ in expectation.

**Claim 5.16.** $\mathbb{E}_\rho\left[(p_{\rho,\ell})^2\right] \le 1 - \frac{\gamma}{64\sqrt{t}}$ holds for all $\ell \in [r']$.

We first conclude the proof of Lemma 5.15 assuming Claim 5.16. Firstly by Jensen's inequality, we have

$$\mathop{\mathbb{E}}_{\rho}\left[1 - p_{\rho,\ell}\right] \ge 1 - \sqrt{\mathop{\mathbb{E}}_{\rho}\left[(p_{\rho,\ell})^2\right]} \ge 1 - \sqrt{1 - \frac{\gamma}{64\sqrt{t}}} \ge \frac{\gamma}{128\sqrt{t}}.$$

Since $N(1),\dots,N(\ell)$ are non-connected, each $\Delta_\ell$ depends on disjoint parts of $Z$. Thus each $p_{\rho,\ell}$ depends on disjoint parts of $\rho$, which means they are independent. Since each $p_{\rho,\ell} \in [0,1]$, by Fact 3.3 with $\delta = 1/2$ we have

$$\mathop{\mathbf{Pr}}_{\rho}\left[\sum_{\ell\in[r']}(1 - p_{\rho,\ell}) \le \frac{1}{2}\cdot\frac{\gamma\cdot r'}{128\sqrt{t}}\right] \le \exp\left\{-\frac{\gamma\cdot r'}{1024\sqrt{t}}\right\} \le O\left(\frac{t^{1/4}}{\sqrt{\gamma\cdot r'}}\right). \tag{13}$$

We say $\rho$ is bad if the above event happens, and good otherwise. Then we have

$$\begin{aligned}
\mathcal{P}_g(\mathsf{supp}\,(\mathcal{D}_k)) &\le O\left(\frac{t^{1/4}}{\sqrt{\gamma\cdot r'}}\right) + \mathbf{Pr}\left[K + \sum_{\ell\in[r']}\Delta_\ell = k \,\middle|\, \rho \text{ is good}\right] && \text{(by (12) and (13))}\\
&\le O\left(\frac{t^{1/4}}{\sqrt{\gamma\cdot r'}}\right) + O\left(\mathop{\mathbb{E}}_{\rho}\left[\frac{1}{\sqrt{\sum_{\ell\in[r']}(1 - p_{\rho,\ell})}} \,\middle|\, \rho \text{ is good}\right]\right) && \text{(by Fact 3.4)}\\
&= O\left(\frac{t^{1/4}}{\sqrt{\gamma\cdot r'}}\right)
\end{aligned}$$

as desired.

Now we prove Claim 5.16.

*Proof of Claim 5.16.* Recall that $(X_1,\ldots,X_n)=g(Z)$ for a random $Z\sim\mathcal{U}^m$, and let $I_\ell=\bigcup_{i\in N(\ell)}I(i)$. Then $\Delta_\ell=\sum_{i\in N(\ell)}X_i$ depends only on bits of $Z$ in $I_\ell$. Since $N(1),\ldots,N(r')$ are non-connected, $I_\ell\cap I(i)=\emptyset$ holds for all $i\neq\ell$. This means the distribution of $\Delta_\ell$ conditioned on $Z_j$'s for $j\in[m]\setminus I(\ell)$ is the same as if we only condition on $Z_j$'s for $j\in R=[m]\setminus(I(1)\cup\cdots\cup I(r'))\supseteq I_\ell\setminus I(\ell)$. Therefore

$$\mathop{\mathbb{E}}_{\rho}\left[(p_{\rho,\ell})^2\right]=\mathop{\mathbb{E}}_{Z_j:j\notin I(\ell)}\left[\max_c\mathbf{Pr}\left[\sum_{i\in N(\ell)}X_i=c\,\middle|\,Z_j:j\notin I(\ell)\right]^2\right]. \tag{14}$$

We sample $Z'\sim\mathcal{U}^m$ conditioned on $Z'_j=Z_j$ for all $j\notin I(\ell)$. In other words, we randomly flip bits in $I(\ell)$ of $Z$ to obtain $Z'$. Define $(Y_1,\ldots,Y_n)=g(Z')$. Then for any value $c$, we have

$$\begin{aligned}\mathbf{Pr}\left[\sum_{i\in N(\ell)}X_i=c\,\middle|\,Z_j\colon j\notin I(\ell)\right]^2&=\mathbf{Pr}\left[\sum_{i\in N(\ell)}X_i=\sum_{i\in N(\ell)}Y_i=c\,\middle|\,Z_j\colon j\notin I(\ell)\right]\\&\qquad\text{($X$'s and $Y$'s are conditionally independent)}\\&\le\mathbf{Pr}\left[\sum_{i\in N(\ell)}X_i=\sum_{i\in N(\ell)}Y_i\,\middle|\,Z_j\colon j\notin I(\ell)\right].\end{aligned}$$

Putting into (14), we have $\mathbb{E}_\rho\left[(p_{\rho,\ell})^2\right]\le\mathbf{Pr}\left[\sum_{i\in N(\ell)}X_i=\sum_{i\in N(\ell)}Y_i\right]$.

By rearranging indices, we assume $N(\ell)=[\ell]$. Now we apply Lemma 4.4 to $(A,B,C,D)$ with $q=8\lceil\sqrt{\gamma t}\rceil$, where $A=X_\ell,C=Y_\ell$ and $B=(X_1,\ldots,X_{\ell-1}),D=(Y_1,\ldots,Y_{\ell-1})$. This holds since $I(\ell)$ is resampled in $Z'$, which decouples $A=X_\ell$ from $(C,D)$ and $C$ from $(A,B)$. In addition, $(A,B),(C,D)$ have the same marginal distribution of $\mathcal{P}_g|_{N(\ell)}$, which is of Type-2, i.e., $\varepsilon$-close to $\mathcal{U}_\gamma^{s_\ell}$. Since $s_\ell\le t$ and

$$\frac{\gamma}{4q}\cdot2^{-50\gamma(t-1)/q^2}\ge\frac{\gamma}{32\lceil\sqrt{\gamma t}\rceil}\cdot\frac{1}{2}\ge\frac{\gamma}{128\sqrt{t}}\ge\varepsilon,$$

Lemma 4.4 implies

$$\begin{aligned}\mathop{\mathbb{E}}_{\rho}\left[(p_{\rho,\ell})^2\right]\le\mathbf{Pr}\left[\sum_{i\in N(\ell)}X_i=\sum_{i\in N(\ell)}Y_i\right]&=\mathbf{Pr}\left[A+|B|=C+|D|\right]\\&\le\mathbf{Pr}\left[A+|B|\equiv C+|D|\pmod q\right]\le1-\frac{\gamma}{64\sqrt{t}}\end{aligned}$$

as desired. □

□

At this point, we are ready to prove Proposition 5.12.

*Proof of Proposition 5.12.* Firstly we note that the bound trivially holds when $r<1$. Hence we assume now $r\ge1$. We set $\varepsilon=\frac{\gamma}{64\sqrt{t}}$ and let $C$ be a universal constant sufficiently large. By Definition 5.11, there are $r$ non-connected neighborhoods of size at most $t$. If $\lceil r/2\rceil$ of them are Type-1, then we apply Lemma 5.14 with $r'=\lceil r/2\rceil$ and obtain

$$\|\mathcal{P}_g-\mathcal{D}_k\|_{\mathsf{TV}}\ge1-2\sqrt{2n}\cdot\exp\left\{-\varepsilon^2r/4\right\}\ge1-C\sqrt{n}\cdot\exp\left\{-\frac{\gamma^2\cdot r}{C\cdot t}\right\}.$$

Otherwise there are $\lceil r/2 \rceil$ of $\mathsf{Type}$-2, and we apply Lemma 5.15 with $r' = \lceil r/2 \rceil$ to obtain

$$\mathcal{P}_g(\mathsf{supp}\,(\mathcal{D}_k)) \le \frac{C \cdot t^{1/4}}{\sqrt{\gamma \cdot r}}. \qquad \square$$

### 5.4 Periodic Hamming Slices

In the last section, we proved lower bounds for sampling a single Hamming slice. Our technique is robust enough that it also works for uniform distributions over multiple Hamming slices. Here we illustrate with periodic Hamming slices.

Let $q \ge 3$ be an integer, and let $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$ be a non-empty set. We define the distribution $\mathcal{D}_{q,\Lambda}$ to be the uniform distribution over $x \in \{0,1\}^n$ conditioned on $|x| \bmod q \in \Lambda$. We will show that, under moderate conditions on $q$ and $\Lambda$, local functions cannot effectively sample from $\mathcal{D}_{q,\Lambda}$.

**Theorem 5.17.** *Let $q \ge 3$ be an integer, and let $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$ not empty. Define $\eta = |\mathsf{supp}\,(\mathcal{D}_{q,\Lambda})| \cdot 2^{-n}$. Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a $d$-local function. Then*

$$\|f(\mathcal{U}^m) - \mathcal{D}_{q,\Lambda}\|_{\mathsf{TV}} \ge 1 - \frac{6q}{\eta} \cdot \exp\left\{-\frac{n}{q^2 \cdot \mathrm{tow}_2(18d)}\right\} - \begin{cases} |\Lambda|/q & q \text{ is odd,} \\ 2 \cdot \max\left\{|\Lambda_{\mathit{even}}|, |\Lambda_{\mathit{odd}}|\right\}/q & q \text{ is even.}\end{cases}$$

We note the following simple lower bound for $\eta$, which implies that Theorem 5.17 gives non-trivial bounds for all $q \le O_d(\sqrt{n})$.

**Claim 5.18.** *$\eta \ge 2^{-q^2/n}/\sqrt{2n}$.*

*Proof.* The bound is trivial when $q = n$. Hence now we assume $q \le n-1$. Since $\Lambda \ne \emptyset$, at least one Hamming slice with weight in $[(n-q)/2, (n+q)/2]$ will be included in $\mathsf{supp}\,(\mathcal{D}_{q,\Lambda})$. By Fact 3.5 and Fact 3.6, we obtain

$$\eta \ge 2^{-n} \cdot \binom{n}{(n-q)/2} \ge 2^{\left(\mathcal{H}\left(\frac{1}{2} - \frac{q}{2n}\right) - 1\right) \cdot n}/\sqrt{2n} \ge 2^{-q^2/n}/\sqrt{2n}. \qquad \square$$

We also remark that the bound in Theorem 5.17 is essentially tight for $q$ not extremely large:

- If $q = 2$, then we can perfectly produce $\mathcal{D}_{2,\{0\}}$ by a 2-local function $(x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_{n-1} \oplus x_n, x_n \oplus x_1)$, and similarly for $\mathcal{D}_{2,\{1\}}$.
- If $q \ge 3$ is odd, then we can produce $\mathcal{U}^n$ by a 1-local function, which hits $\Lambda$ with probability roughly $|\Lambda|/q$.
- If $q \ge 3$ is even, then we can produce $\mathcal{D}_{2,\{0\}}$ as described above, which is roughly uniform over even numbers in $\mathbb{Z}/q\mathbb{Z}$ after modulo $q$. Thus we hit $\Lambda_{\mathsf{even}}$ with probability $|\Lambda_{\mathsf{even}}|/(q/2)$. Similar construction using $\mathcal{D}_{2,\{1\}}$ will achieve the distance bound $1 - |\Lambda_{\mathsf{odd}}|/(q/2)$.

The proof of Theorem 5.17 follows the same paradigm as the previous section. We first give bounds for $(d,r,t)$-local functions. The proof of the following proposition is presented at the end of the section.

**Proposition 5.19.** *Let $g\colon \{0,1\}^m \to \{0,1\}^n$ be a $(d,r,t)$-local function and define $\mathcal{P}_g = g(\mathcal{U}^m)$. Then either*

$$\|\mathcal{P}_g - \mathcal{D}_{q,\Lambda}\|_{\mathsf{TV}} \ge 1 - \frac{2}{\eta} \cdot \exp\left\{-2^{-28t+19} \cdot r\right\}.$$

*or*

$$\mathcal{P}_g(\mathsf{supp}\,(\mathcal{D}_{q,\Lambda})) \le 2q \cdot \exp\left\{-\frac{r \cdot 2^{-14t+10}}{q^2}\right\} + \begin{cases} |\Lambda|/q & q \text{ odd,} \\ 2 \cdot \max\left\{|\Lambda_{\mathit{even}}|, |\Lambda_{\mathit{odd}}|\right\}/q & q \text{ even,}\end{cases}$$

Then we prove the following graph elimination result tailored for the parameters in Proposition 5.19.

**Proposition 5.20.** *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*

$$|S| \le \frac{r}{2^{28t-18}} \quad \text{and} \quad r \ge \frac{n}{\mathrm{tow}_2(16d)} \quad \text{and} \quad t \le \mathrm{tow}_2(16d).$$

*Proof.* The statement is trivial when $d = 0$ since then we can set $S = \emptyset, r = n, t = 0$. For $d \ge 1$, we apply Corollary 4.11. Set $F(x) = 2^{28t-18}$. Then

$$\widetilde{F}(x) = \frac{1}{d} \cdot \exp\left\{32d^4x^2 \cdot 2^{56dx-18}\right\}.$$

Define $H(x) = 2^{2^{2^x}}$ and let $L = 10d$. By setting

$$\kappa = \lambda = \mathrm{tow}_2(16d) \ge d \cdot H^{(2d+2)}(L),$$

the conditions in Corollary 4.11 are satisfied. This implies that Property 4.9 holds for the dependency graph of $f$ with parameter $\lambda, \kappa, F$. □

Finally we use the above graph elimination results to lift the lower bounds of $(d, r, t)$-local functions to $d$-local functions.

*Proof of Theorem 5.17.* We assume $d \ge 1$, as otherwise $f$ must be constant, and one can verify the bound holds. By Proposition 5.20, we find a set $S \subseteq [m]$ such that any fixing $\rho$ of input bits in $S$ reduces $f$ to a $(d, r, t)$-local function $f_\rho$ where

$$|S| \le \frac{r}{2^{28t-18}} \quad \text{and} \quad r \ge \frac{n}{\mathrm{tow}_2(16d)} \quad \text{and} \quad t \le \mathrm{tow}_2(16d).$$

Now for each $f_\rho$, we apply Proposition 5.19 and obtain that either

$$\left\|\mathcal{P}_{f_\rho} - \mathcal{D}_{q,\Lambda}\right\|_{\mathsf{TV}} \ge 1 - \underbrace{\frac{2}{\eta} \cdot \exp\left\{-2^{-28t+19} \cdot r\right\}}_{\varepsilon_1}$$

or

$$\mathcal{P}_{f_\rho}(\mathsf{supp}\left(\mathcal{D}_{q,\Lambda}\right)) \le \varepsilon_2 := 2q \cdot \exp\left\{-\frac{r \cdot 2^{-14t+10}}{q^2}\right\} + \begin{cases} |\Lambda|/q & q \text{ is odd,} \\ 2 \cdot \max\left\{|\Lambda_{\mathsf{even}}|, |\Lambda_{\mathsf{odd}}|\right\}/q & q \text{ is even.} \end{cases}$$

Note that $f(\mathcal{U}^m)$ is a convex combination of $\mathcal{P}_{f_\rho}$'s. By Lemma 4.3 with $\left\{\mathcal{P}_{f_\rho}\right\}_\rho, \mathcal{D}_{q,\Lambda}$, and $\varepsilon_3 = 0, \mathcal{E} = \mathsf{supp}\left(\mathcal{D}_{q,\Lambda}\right)$ and $\varepsilon_1, \varepsilon_2$ defined above, we have

$$\begin{aligned} \|f(\mathcal{U}^m) - \mathcal{D}_{q,\Lambda}\|_{\mathsf{TV}} &\ge 1 - \left(2^{|S|} + 1\right) \cdot \varepsilon_1 - \varepsilon_2 \ge 1 - \frac{4}{\eta} \cdot \exp\left\{-\frac{r}{2^{28t-18}}\right\} - \varepsilon_2 \\ &\ge 1 - \frac{6q}{\eta} \cdot \exp\left\{-\frac{r}{q^2 \cdot 2^{28t-18}}\right\} - \begin{cases} |\Lambda|/q & q \text{ is odd,} \\ 2 \cdot \max\left\{|\Lambda_{\mathsf{even}}|, |\Lambda_{\mathsf{odd}}|\right\}/q & q \text{ is even.} \end{cases} \end{aligned}$$

Since $t \le \mathrm{tow}_2(16d)$ and $r \ge n/\mathrm{tow}_2(16d)$, we can bound

$$\frac{r}{2^{28t-18}} \ge \frac{n}{\mathrm{tow}_2(16d) \cdot 2^{28 \cdot \mathrm{tow}_2(16d) - 18}} \ge \frac{n}{\mathrm{tow}_2(18d)},$$

which gives the desired bound in Theorem 5.17. □

Now we prove Proposition 5.19. The road-map is similar to the proof of Proposition 5.12: we first classify each non-connected neighborhood dependent on whether its marginal distribution is far from unbiased. If most are far, we use Lemma 4.2 to show the distance bound. Otherwise we use local limit results to show that with certain probability the Hamming weight modulo $q$ cannot fall into $\Lambda$.

Let $\varepsilon \in [0,1]$ be a parameter to be optimized later. For each neighborhood $N(i) = N_g(i)$ of size $s_i = |N(i)|$, we classify it into one of the following two cases:

- Type-1. $\mathcal{P}_g|_{N(i)}$ is not $\varepsilon$-close to $\mathcal{U}^{s_i}$.
- Type-2. $\mathcal{P}_g|_{N(i)}$ is $\varepsilon$-close to $\mathcal{U}^{s_i}$.

By almost identical reasoning as Lemma 5.14 (except that we fix $\gamma = 1/2$ here), we obtain a large distance bound when there are many small Type-1 neighborhoods.

**Lemma 5.21.** *Assume there are $r' \geq 1$ non-connected Type-1 neighborhoods. Then*

$$\|\mathcal{P}_g - \mathcal{D}_{q,\Lambda}\|_{\mathsf{TV}} \geq 1 - \frac{2}{\eta} \cdot \exp\left\{-\varepsilon^2 r'/2\right\}.$$

Now we turn to the second case where we have many small Type-2 neighborhoods. The analysis for this setting is also similar to the proof of Lemma 5.15, except that we now use a local limit theorem in additive groups instead of anticoncentration inequalities over the real numbers.

**Lemma 5.22.** *Assume there are $r' \geq 1$ non-connected Type-2 neighborhoods of size at most $t$. If $\varepsilon \leq 2^{-14t+10}$, then*

$$\mathcal{P}_g(\mathsf{supp}\left(\mathcal{D}_{q,\Lambda}\right)) \leq 2q \cdot \exp\left\{-\frac{r' \cdot 2^{-14t+11}}{q^2}\right\} + \begin{cases} |\Lambda|/q & q \text{ odd,} \\ 2 \cdot \max\left\{|\Lambda_{\mathsf{even}}|, |\Lambda_{\mathsf{odd}}|\right\}/q & q \text{ even,} \end{cases}$$

*Proof.* We inherit the notation $(X_i)_{i\in[n]}, \rho, K, (\Delta_\ell)_{\ell\in[r']}$ from Lemma 5.15 with one minor change: here for each integer $a \geq 3$, we define

$$p_{\rho,a,\ell} = \max_{c\in\mathbb{Z}} \mathbf{Pr}\left[\Delta_\ell \equiv c \pmod{a} \mid \rho\right].$$

By a similar analysis as in Claim 5.16, we obtain the following claim.

**Claim 5.23.** $\mathbb{E}_\rho[(p_{\rho,a,\ell})^2] \leq 1 - 2^{-14t+11}$ holds for any $\ell \in [r']$ and $a \geq 3$.

*Proof.* We only highlight the difference from the proof of Claim 5.16. We apply Lemma 4.4 with $\gamma = 1/2$ and modulus $\overline{q} = \min\left\{a, t+1\right\}$ here. Then $2 \leq \overline{q} \leq t+1$ and

$$\frac{\gamma}{4\overline{q}} \cdot 2^{-50\gamma(t-1)/\overline{q}^2} \geq \frac{1}{8(t+1)} \cdot 2^{-13(t-1)} \geq 2^{-14t+10} \geq \varepsilon,$$

which implies $\mathbb{E}_\rho[(p_{\rho,a,\ell})^2] \leq 1 - 2^{-14t+11}$ by Lemma 4.4. □

Then we have $\mathbb{E}_\rho[1 - p_{\rho,a,\ell}] \geq 2^{-14t+10}$. As before, since $p_{\rho,a,\ell}$'s are independent for fixed $a$, by Fact 3.3 we have

$$\mathbf{Pr}\left[\sum_{\ell\in[r']} (1 - p_{\rho,a,\ell}) \leq 2^{-14t+9} \cdot r'\right] \leq \exp\left\{-2^{-14t+6} \cdot r'\right\}.$$

Now we say $\rho$ is bad if for some $a \geq 3$ dividing $q$ the above event happens, and good otherwise. Since there are at most $q$ possible such $a$'s, by the union bound we have

$$\mathbf{Pr}\left[\rho \text{ is bad}\right] \leq q \cdot \exp\left\{-2^{-14t+6} \cdot r'\right\}. \tag{15}$$

Thus,

$$\begin{aligned}
&\mathcal{P}_g(\mathsf{supp}\left(\mathcal{D}_{q,\Lambda}\right) \\
\leq &\mathbf{Pr}\left[\rho \text{ is bad}\right] + \mathbf{Pr}\left[K + \sum_{\ell \in [r']} \Delta_\ell \bmod q \in \Lambda \,\middle|\, \rho \text{ is good}\right] \\
\leq &q \cdot \exp\left\{-2^{-14t+6} r'\right\} + q \cdot \exp\left\{-\frac{2^{-14t+11} r'}{q^2}\right\} + \begin{cases} |\Lambda|/q & q \text{ odd}, \\ 2 \cdot \max\left\{|\Lambda_{\mathsf{even}}|, |\Lambda_{\mathsf{odd}}|\right\}/q & q \text{ even}, \end{cases}
\end{aligned}$$

where for the last inequality we use (15) and Lemma 3.7 with $L = 2^{-14t+10} \cdot r'$ and the observation that $\Lambda$ shifted by $K$ still has the same maximum of even and odd numbers. □

Finally we choose $\varepsilon$ and conclude the proof of Proposition 5.19.

*Proof of Proposition 5.19.* Firstly the bound is trivial if $r < 1$. For $r \geq 1$, we set $\varepsilon = 2^{-14t+10}$. By Definition 5.11, there are $r$ non-connected neighborhoods of size at most $t$. If $\lceil r/2 \rceil$ of them are $\mathsf{Type}$-$1$, then we apply Lemma 5.21 with $r' = \lceil r/2 \rceil$ and obtain

$$\|\mathcal{P}_g - \mathcal{D}_{q,\Lambda}\|_{\mathsf{TV}} \geq 1 - \frac{2}{\eta} \cdot \exp\left\{-2^{-28t+19} \cdot r\right\}.$$

Otherwise there are $\lceil r/2 \rceil$ of $\mathsf{Type}$-$2$, and we apply Lemma 5.22 with $r' = \lceil r/2 \rceil$ to obtain

$$\mathcal{P}_g(\mathsf{supp}\left(\mathcal{D}_{q,\Lambda}\right)) \leq 2q \cdot \exp\left\{-\frac{r \cdot 2^{-14t+10}}{q^2}\right\} + \begin{cases} |\Lambda|/q & q \text{ odd}, \\ 2 \cdot \max\left\{|\Lambda_{\mathsf{even}}|, |\Lambda_{\mathsf{odd}}|\right\}/q & q \text{ even}, \end{cases}$$ □

## 6 Upper Bounds

In this section, we provide upper bounds on the locality of functions sampling specific distributions. Subsection 6.1 contains two incomparable bounds for $\mathcal{D}_k$.

**Theorem 6.1.** *For all $k \leq n$, there exists a $d$-local Boolean function $f : \{0,1\}^m \to \{0,1\}^n$ such that $f(\mathcal{U}^m)$ is $\varepsilon$-close to $\mathcal{D}_k$, where*

$$d = O\left(\min\left\{\log(n) \cdot \log(n/\varepsilon), \log(n/k) + \log^2(k/\varepsilon)\right\}\right).$$

Subsection 6.2 uses the previous theorem to prove an upper bound on $\mathcal{D}_{q,\Lambda}$.

**Theorem 6.2.** *For all $q \in \mathbb{N}$ and non-empty $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$, there exists an $O(q^2 \cdot \log^2(n/\varepsilon))$-local Boolean function $f : \{0,1\}^m \to \{0,1\}^n$ such that $f(\mathcal{U}^m)$ is $\varepsilon$-close to $\mathcal{D}_{q,\Lambda}$.*

Throughout, we will need to sample from various distributions. The following standard lemma allows us to do so approximately with low locality.

**Lemma 6.3.** *Any discrete distribution of support size $m$ can be approximated to $\varepsilon$ error in total variation distance with $\lceil \log(m/\varepsilon) \rceil$ uniform random bits.*

*Proof.* Let $\mathcal{D}$ be a discrete distribution with support $\{x_1, \ldots, x_m\}$, and let $B := 2^{\lceil \log(m/\varepsilon) \rceil}$. Furthermore, define $\mathcal{T}$ to be the distribution by discretizing $\mathcal{D}$'s probability density function:

$$\mathcal{T}(x_i) = \begin{cases} \frac{1}{B} \cdot \lfloor B \cdot \mathcal{D}(x_i) \rfloor & i \in [m-1], \\ 1 - \sum_{i=1}^{m-1} \mathcal{T}(x_i) & i = m. \end{cases}$$

Then,

$$\|\mathcal{T} - \mathcal{D}\|_{\mathsf{TV}} = \mathcal{T}(x_m) - \mathcal{D}(x_m) = 1 - \left( \sum_{i=1}^{m-1} \frac{1}{B} \cdot \lfloor B \cdot \mathcal{D}(x_i) \rfloor \right) - \mathcal{D}(x_m) \leq \varepsilon. \qquad \square$$

**Decision Forest Depth.** Our work focuses on quantifying a circuit's complexity by its locality. Another common measure is *decision forest depth*, which can be viewed as an adaptive variant of locality. A function $f : \{0,1\}^m \to \{0,1\}^n$ is computable by a depth-$d$ decision forest if every output bit of $f$ is the result of a depth-$d$ decision tree of the input bits. Observe that such a function is $2^d$-local. This notion was studied in earlier works, such as [Vio12b, FLRS23]. In fact, several of their aforementioned lower bounds also hold in this stronger model. Their upper bounds also only appear in terms of this quantity, rather than locality.

It is known that there exists a sampler for $\mathcal{D}_k$ of depth $O(\log n)$ [Vio12a] using a (randomized) switching network construction of [Czu15]. Our Theorem 6.1 provides comparable bounds in terms of the weaker locality parameter. No samplers seem to appear in the literature for $\mathcal{D}_{q,\Lambda}$.

## 6.1 A Single Hamming Slice

In this subsection, we prove Theorem 6.1 in two parts: Theorem 6.4 gives an $O(\log^2 n)$ bound, while Theorem 6.5 gives an $O(\log(n/k) + \log^2 k)$ bound.

**Theorem 6.4.** *For all $k \leq n$, there exists an $O(\log(n) \cdot \log(n/\varepsilon))$-local Boolean function $f : \{0,1\}^m \to \{0,1\}^n$ such that $f(\mathcal{U}^m)$ is $\varepsilon$-close to $\mathcal{D}_k$.*

*Proof.* We sample $\mathcal{D}_k$ by iteratively partitioning the interval $[n]$ into two (essentially) equally sized intervals, which contain $a$ and $k-a$ ones to place, respectively, for an $a$ sampled from the appropriate hypergeometric distribution. More precisely, $\mathcal{D}_k$ can be viewed as the resulting distribution of the following process.

1. Consider the interval $S := [n]$. There are $\ell := k$ ones to place inside, and the remaining entries are zeros.
2. Partition $S$ into two consecutive intervals $S_1$ and $S_2$ of sizes $\lfloor |S|/2 \rfloor$ and $\lceil |S|/2 \rceil$.
3. Pick an element $a \in \{0, \ldots, \ell\}$ according to the hypergeometric distribution $\mathcal{H}_{|S|,\ell,|S_1|}$:

$$\mathbf{Pr}[a = r] = \frac{\binom{\ell}{r} \cdot \binom{|S|-\ell}{|S_1|-r}}{\binom{|S|}{|S_1|}}.$$

4. Repeat the process with $S_1$ and $S_2$ given $a$ and $\ell - a$ ones, respectively, to place in their intervals.

We would like to approximate $\mathcal{D}_k$ with a distribution $\mathcal{L}_k := f(\mathcal{U}^m)$ produced by a function $f : \{0,1\}^m \to \{0,1\}^n$ of small locality. By Lemma 6.3, we can sample from $\mathcal{H}_{|S|,\ell,|S_1|}$ with at most

$\lceil \log(k/\varepsilon') \rceil$ bits of locality and error $\varepsilon'$. The key insight is that each output bit depends only on the interval containing it in each of the at most $\lceil \log n \rceil$ steps, so the total locality is $O(\log(n) \cdot \log(k/\varepsilon'))$.

Since each low locality approximation to the hypergeometric distribution sampled in the above process is $\varepsilon'$-close to the true distribution, $\|\mathcal{L}_k - \mathcal{D}_k\|_{\mathsf{TV}} \le \varepsilon' n$ by the union bound. Setting $\varepsilon' = \varepsilon/n$, we find that $f$ is an $O(\log(n) \cdot \log(n/\varepsilon))$-local function with $f(\mathcal{U}^m)$ $\varepsilon$-close to $\mathcal{D}_k$. □

We can use the above theorem to derive a tighter bound in the case of small $k$.

**Theorem 6.5.** *For all $k \le n$, there exists an $O(\log(n/k) + \log^2(k/\varepsilon))$-local Boolean function $f : \{0,1\}^m \to \{0,1\}^n$ such that $f(\mathcal{U}^m)$ is $\varepsilon$-close to $\mathcal{D}_k$.*

For clarity in the proof, we will use $\mathcal{D}_{s,t}$ to denote the uniform distribution over $x \in \{0,1\}^t$ of Hamming weight $s$. When only one parameter is provided, $\mathcal{D}_s$ denotes the uniform distribution over $x \in \{0,1\}^n$ of Hamming weight $s$, as in prior contexts.

*Proof.* Assume $k \le \sqrt{\varepsilon n}$, as otherwise the bound follows from Theorem 6.4. We first describe a randomized process to approximately generate $\mathcal{D}_k$: split $[n]$ into (essentially) equally sized parts, pick $k$ at random, and place a single 1 randomly in each of the chosen parts. The number of parts in the original split determines the accuracy of this approximation. More precisely, let $\mathcal{P}_k$ be the resulting distribution of the following process.

1. Divide $[n]$ into $t$ intervals, each of size $\lceil n/t \rceil$ or $\lfloor n/t \rfloor$ for a $t$ to be determined later.
2. Pick $k$ distinct intervals according to the distribution $\mathcal{D}_{k,t}$.
3. Put a 1 in each of the selected intervals according to the distribution $\mathcal{D}_{1,\lceil n/t \rceil}$ or $\mathcal{D}_{1,\lfloor n/t \rfloor}$, depending on the block size. (The remaining entries are zeros.)

By direct calculation, we find

$$\begin{aligned} \|\mathcal{P}_k - \mathcal{D}_k\|_{\mathsf{TV}} &= \mathop{\mathbf{Pr}}_{x \sim \mathcal{D}_k}[x \text{ has two ones in the same interval}] \\ &\le \mathop{\mathbb{E}}_{x \sim \mathcal{D}_k}[\text{number of pairs of ones in } x \text{ in the same interval}] \\ &\le k^2/t. \end{aligned}$$

Hence, as long as $t \ge \lceil k^2/\varepsilon \rceil$, the two distributions are $\varepsilon$-close.

We would like to approximate $\mathcal{P}_k$ with a distribution $\mathcal{L}_k \coloneqq f(\mathcal{U}^m)$ produced by a function $f : \{0,1\}^m \to \{0,1\}^n$ of small locality. Each output bit $b$ depends on the choice of intervals and the location of the 1 within the interval containing $b$. Using Theorem 6.4, we can sample from $\mathcal{D}_{k,t}$ with $O(\log(t) \cdot \log(t/\varepsilon_1))$ bits of locality and error $\varepsilon_1$. Similarly using Lemma 6.3, we can sample from $\mathcal{D}_{1,\lceil n/t \rceil}$ or $\mathcal{D}_{1,\lfloor n/t \rfloor}$ with $O(\log(n/(t\varepsilon_2)))$ bits of locality and error $\varepsilon_2$, so the total locality is $O(\log(t) \cdot \log(t/\varepsilon_1) + \log(n/t\varepsilon_2))$.

Union bounding over the errors implies $\|\mathcal{L}_k - \mathcal{P}_k\|_{\mathsf{TV}} \le \varepsilon_1 + k\varepsilon_2$. Setting $\varepsilon_1 = \varepsilon/4$ and $\varepsilon_2 = \varepsilon/4k$, we find that $\mathcal{L}_k$ is $(\varepsilon/2)$-close to $\mathcal{P}_k$. Finally setting $t = \lceil k^2/(\varepsilon/2) \rceil$ yields that $\mathcal{P}_k$ is $(\varepsilon/2)$-close to $\mathcal{D}_k$, so we conclude that $f$ is an $O(\log(n/k) + \log^2(k/\varepsilon))$-local function with $f(\mathcal{U}^m)$ $\varepsilon$-close to $\mathcal{D}_k$. □

## 6.2 Periodic Hamming Slices

In this subsection, we prove Theorem 6.2. Recall that $\mathcal{D}_{q,\Lambda}$ for $\Lambda \in \mathbb{Z}/q\mathbb{Z}$ is the uniform distribution over $n$-bit strings of Hamming weight modulo $q$ in $\Lambda$.

**Theorem** (Theorem 6.2 Restated)**.** *For all $q \in \mathbb{N}$ and non-empty $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$, there exists an $O(q^2 \cdot \log^2(n/\varepsilon))$-local Boolean function $f : \{0,1\}^m \to \{0,1\}^n$ such that $f(\mathcal{U}^m)$ is $\varepsilon$-close to $\mathcal{D}_{q,\Lambda}$.*

We start with the case of $|\Lambda| = 1$, where our approach is similar to the PARITY example in the introduction.

**Lemma 6.6.** *For all $q \in \mathbb{N}$ and $r \in \mathbb{Z}/q\mathbb{Z}$, there exists an $O(q^2 \cdot \log^2(n/\varepsilon))$-local Boolean function $f : \{0,1\}^m \to \{0,1\}^n$ such that $f(\mathcal{U}^m)$ is $\varepsilon$-close to $\mathcal{D}_{q,\{r\}}$.*

We again will use $\mathcal{D}_{s,t}$ to denote the uniform distribution over $x \in \{0,1\}^t$ of Hamming weight $s$ in cases where $t$ may not equal $n$.

*Proof.* Assume $q = o(\sqrt{n/\log(n/\varepsilon)})$, as otherwise the bound follows from Lemma 6.3. We can view $\mathcal{U}^n$ as the resulting distribution of the following processing.

1. Divide $[n]$ into $\lfloor n/t \rfloor$ blocks, each of size either $t$ or $t+1$.

   (This is the reverse of what was done in the proof of Theorem 6.5, where there were $t$ blocks. It will be more convenient for later calculations.)
2. Sample $y_1, \ldots, y_{\lfloor n/t \rfloor} \in \mathbb{Z}/q\mathbb{Z}$ according to the binomial distribution $\mathcal{B}_{q,1/2}$:

$$\mathbf{Pr}[y_i = j] = \frac{1}{2^q} \cdot \binom{q}{j}.$$

3. Suppose the $i$-th block has size $h$. Sample $w_i$ according to the binomial distribution $\mathcal{B}_{h,1/2}$ conditioned on the result being $y_i \pmod q$.
4. For each $i$, put $w_i$ ones in block $i$ according to the distribution $\mathcal{D}_{w_i,h}$.

We claim that the $y_i$'s are approximately uniform over $\mathbb{Z}/q\mathbb{Z}$. Thus, if we perform a variant of the above process where each $y_i$ is chosen uniformly at random from $\mathbb{Z}/q\mathbb{Z}$, we can get an $(\varepsilon/2)$-approximation to the uniform distribution over $\{0,1\}^n$.

**Claim 6.7.** Setting $t = \Omega(q^2 \cdot \log(n/\varepsilon))$ implies the uniform distribution over $\{0,1\}^n$ restricted to a block has weight mod $q$ $(\varepsilon/2n)$-close to the uniform distribution over $\mathbb{Z}/q\mathbb{Z}$.

*Proof.* It suffices to show that $\left|\mathbf{Pr}_{X\sim\{0,1\}^t}\left[\sum_{j\in[t]} X_j \bmod q = v\right] - 1/q\right| \le e^{-t/q^2}$ for any $v \in \mathbb{Z}/q\mathbb{Z}$. By Fourier expansion,

$$\mathbf{Pr}\left[\sum_{j\in[t]} X_j \bmod q = v\right] = \mathbb{E}\left[\frac{1}{q}\sum_{a\in\mathbb{Z}/q\mathbb{Z}} \omega_q^{a\cdot(\sum_j X_j - v)}\right] = \frac{1}{q} + \frac{1}{q}\sum_{a\in\mathbb{Z}/q\mathbb{Z},a\neq 0}\left(\frac{1+\omega_q^a}{2}\right)^t \cdot \omega_q^{-a\cdot v},$$

where $\omega_q$ is the $q$-th unit root.

By Claim B.3, we can bound the error by

$$\left|\frac{1}{q}\sum_{a\in\mathbb{Z}/q\mathbb{Z},a\neq 0}\left(\frac{1+\omega_q^a}{2}\right)^t \cdot \omega_q^{-a\cdot v}\right| \le \max_{a\in\mathbb{Z}/q\mathbb{Z},a\neq 0}\left|\frac{1+\omega_q^a}{2}\right|^t \le \left(1-\frac{1}{q^2}\right)^t \le e^{-t/q^2}. \qquad \square$$

Observe that $\mathcal{D}_{q,\{r\}}$ is the uniform distribution over $\{0,1\}^n$ conditioned on the $y_i$'s summing to $r \pmod q$. We can sample $y_i$'s with this property by choosing $x_1, \ldots, x_{\lfloor n/t \rfloor} \in \mathbb{Z}/q\mathbb{Z}$ uniformly at random and setting

$$y_1 \coloneqq x_1 - x_2, \quad y_2 \coloneqq x_2 - x_3, \quad \ldots, \quad y_{\lfloor n/t \rfloor - 1} \coloneqq x_{\lfloor n/t \rfloor - 1} - x_{\lfloor n/t \rfloor}, \quad y_{\lfloor n/t \rfloor} \coloneqq x_{\lfloor n/t \rfloor} - x_1 + r.$$

Call the resulting distribution of the above process with this modification $\mathcal{P}$. Then we have $\left\|\mathcal{P} - \mathcal{D}_{q,\{r\}}\right\|_{\mathsf{TV}} \le \varepsilon/2$.

We would like to approximate $\mathcal{P}$ with a distribution $\mathcal{L} \coloneqq f(\mathcal{U}^m)$ produced by a function $f : \{0,1\}^m \to \{0,1\}^n$ of small locality. By Lemma 6.3, we can sample from the uniform distribution over $\mathbb{Z}/q\mathbb{Z}$ with $\lceil \log(q/\varepsilon') \rceil$ bits of locality and error $\varepsilon'$. Similarly, we can sample from the uniform distribution over $\{0,1\}^{t+1}$ (or $\{0,1\}^t$) conditioned on the $y_i$'s summing to $0 \pmod q$ with $\lceil (t+1)\log(2/\varepsilon') \rceil$ bits of locality and error $\varepsilon'$. Finally, we can sample from $\mathcal{D}_{w_i,h}$ with $O(\log(t) \cdot \log(t/\varepsilon'))$ bits of locality and error $\varepsilon'$ using Theorem 6.1. Output bits in block $i$ depend on the two $x_j$'s that affect $y_i$, $w_i$, and the placement of the $w_i$ ones, so the total locality is $O(\log(q/\varepsilon') + t\log(2/\varepsilon') + \log(t) \cdot \log(t/\varepsilon'))$.

Union bounding over the errors implies $\|\mathcal{L} - \mathcal{P}\|_{\mathsf{TV}} \le \frac{n}{t} \cdot 3\varepsilon'$. Setting $\varepsilon' = t\varepsilon/6n$ implies $\mathcal{L}$ depends on $O(t\log(n/t\varepsilon) + \log(t) \cdot \log(n/\varepsilon))$ bits of locality and is $\varepsilon/2$-close to $\mathcal{P}$. Finally, setting $t = \Theta(q^2 \cdot \log(n/\varepsilon))$, we find that $f$ is an $O(q^2 \cdot \log^2(n/\varepsilon))$-local function with $f(\mathcal{U}^m)$ $\varepsilon$-close to $\mathcal{D}_{q,\{r\}}$. □

By randomly choosing a Hamming weight from $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$, we get Theorem 6.2.

*Proof of Theorem 6.2.* Randomly choose an element $r \in \Lambda$ with probability $\frac{|\mathsf{supp}(\mathcal{D}_{q,\{r\}})|}{|\mathsf{supp}(\mathcal{D}_{q,\Lambda})|}$, and then apply Lemma 6.6 with error $\varepsilon/2$. Lemma 6.3 implies we can sample $r$ with locality $\lceil \log(2q/\varepsilon) \rceil$ and error $\varepsilon/2$. Applying the union bound concludes the proof. □

## Acknowledgements

We thank Yuval Filmus for pointing this topic to us and answering questions regarding [FLRS23, WP23], and we thank Emanuele Viola for answering questions regarding [Vio23] and clarifying prior literature. KW wants to thank Daniel Grier, Jiaqing Jiang, and Jack Morris for answering questions in quantum computing, and thank Mingmou Liu for questions in data structure lower bounds. AO wants to thank Max Hopkins for helpful conversations on the upper bounds. We are also indebted to anonymous reviewers for their useful feedback, especially their discussion on prior techniques and results.

Part of this work was done while AO and KW were visiting the Simons Institute for the Theory of Computing.

## References

[Bab87] Lászió Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Information Processing Letters*, 26(1):51–53, 1987. 3

[BGK18] Sergey Bravyi, David Gosset, and Robert König. Quantum advantage with shallow circuits. *Science*, 362(6412):308–311, 2018. 6

[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. 3, 5

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

[CGZ22] Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The space complexity of sampling. In *13th Innovations in Theoretical Computer Science Conference,(ITCS 2022)*, 2022. 3, 10

[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. 3

[CT06] Thomas M Cover and Joy A Thomas. Elements of information theory, 2006. 15

[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. 3

[Czu15] Artur Czumaj. Random permutations using switching networks. In *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*, pages 703–712, 2015. 38

[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. 3

[Erd45] Paul Erdős. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51(11):898–902, 1945. 10

[FLRS23] Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. Sampling and certifying symmetric functions. In *Approximation, Randomization, and Combinatorial Optimization. (APPROX/RANDOM)*, 2023. 3, 5, 6, 7, 8, 38, 41

[GM07] Anna Gál and Peter Bro Miltersen. The cell probe complexity of succinct data structures. *Theoretical computer science*, 379(3):405–417, 2007. 5

[GS20] Daniel Grier and Luke Schaeffer. Interactive shallow clifford circuits: Quantum advantage against $\mathrm{NC}^1$ and beyond. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 875–888, 2020. 7

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

[Hås86] Johan Håstad. *Computational limitations for small depth circuits.* PhD thesis, Massachusetts Institute of Technology, 1986. 3

[IN96] Russell Impagliazzo and Moni Naor. Efficient cryptographic schemes provably as secure as subset sum. *Journal of cryptology*, 9(4):199–216, 1996. 3

[KOWar] Daniel Kane, Anthony Ostuni, and Kewen Wu. Locally sampleable (uniform) symmetric distributions, (To appear). 7

[LLYZ23] Tianxiao Li, Jingxun Liang, Huacheng Yu, and Renfei Zhou. Tight cell-probe lower bounds for dynamic succinct dictionaries. *arXiv preprint arXiv:2306.02253*, 2023. 6

[LO43] John E Littlewood and Albert C Offord. On the number of real roots of a random algebraic equation (iii). *Rec. Math.[Mat. Sbornik] NS*, 12(54):3, 1943. 10

[LPPZ23] Kasper Green Larsen, Rasmus Pagh, Toniann Pitassi, and Or Zamir. Optimal non-adaptive cell probe dictionaries and hashing. *arXiv preprint arXiv:2308.16042*, 2023. 5

[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. 3, 5

[PY20] Giuseppe Persiano and Kevin Yeo. Tight static lower bounds for non-adaptive data structures. *arXiv preprint arXiv:2001.05053*, 2020. 5, 6

[SW22] Zbigniew Szewczak and Michel Weber. Classical and almost sure local limit theorems. *arXiv preprint arXiv:2208.02700*, 2022. 16

[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. 15

[Vio12a] Emanuele Viola. Bit-probe lower bounds for succinct data structures. *SIAM Journal on Computing*, 41(6):1593, 2012. 3, 5, 6, 7, 25, 38

[Vio12b] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012. 3, 5, 6, 8, 10, 38

[Vio12c] Emanuele Viola. Extractors for Turing-machine sources. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 663–671. Springer, 2012. 3

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014. 3

[Vio16] Emanuele Viola. Quadratic maps are hard to sample. *ACM Transactions on Computation Theory (TOCT)*, 8(4):1–4, 2016. 3

[Vio20] Emanuele Viola. Sampling lower bounds: boolean average-case and permutations. *SIAM Journal on Computing*, 49(1):119–137, 2020. 3, 5, 8, 17

[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. 3, 4, 7, 25, 41

[Wik23] Wikipedia. Binary entropy function — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Binary%20entropy%20function&oldid=1071507954, 2023. [Online; accessed 04-December-2023]. 15

[WP23] Adam Bene Watts and Natalie Parham. Unconditional quantum advantage for sampling with shallow circuits. *arXiv preprint arXiv:2301.00995*, 2023. 3, 6, 7, 41

# A Tightness of the Graph Elimination Results

In this section, we show the tightness of our graph elimination results, which are the main bottlenecks preventing better locality lower bounds. Recall that these graph elimination reductions aim to remove a few right vertices in a $d$-left-bounded bipartite graph to obtain non-connected left vertices or neighborhoods, which corresponds to conditioning on a few input bits in a $d$-local function to obtain a $(d, r)$-local or $(d, r, t)$-local function.

## A.1 Non-Connected Vertices

We start with the tightness of the graph elimination result for non-connected vertices, i.e., Property 4.6 and Corollary 4.8. This is used in the lower bounds for sampling biased distributions (Theorem 5.3) and single Hamming slices of non-dyadic weight (Theorem 5.7).

**Property** (Property 4.6 Restated)**.** *There exists $S \subseteq [m]$ such that deleting those right vertices (and their incident edges) produces a bipartite graph with $r$ non-connected left vertices satisfying*

$$|S| \le \frac{r}{\beta} \quad \textit{and} \quad r \ge \frac{n}{\lambda}.$$

In light of Property 4.6, we prove the following statement.

**Lemma A.1.** *Let $\beta \ge 2$ be an integer parameter (not necessarily constant). If $n = (\beta - 1)^d$, then there exists a $d$-left-bounded bipartite graph $G = ([n], [m], E)$ such that for any $S \subseteq [m]$, deleting those right vertices (and their incident edges) gives at most $\max\{1, (\beta - 1)|S|\}$ non-connected left vertices.*

The instance $G$ above does not satisfy Property 4.6 if $\lambda < n = (\beta-1)^d$. Recall that Corollary 4.8 shows that Property 4.6 holds as long as $\lambda \ge (d\beta)^{\Omega(d)}$. Hence Lemma A.1 provides a sharp example for Corollary 4.8 when $\beta \ge d^{\Omega(1)}$, which is the typical setting for us.

In the whole analysis, it implies a barrier for improving the $\delta^{O(d)}$ factor in Theorem 5.3 and Theorem 5.7 to $\delta^{o(d)}$. Put concretely, the $2^{O(d^2)}$ factor in Theorem 1.3 and Theorem 1.10 is inevitable in our analysis framework.

Now we proceed to the proof of Lemma A.1.

*Proof of Lemma A.1.* The right vertices of $G$ will form a complete $(\beta - 1)$-ary tree on top of the left vertices. To be more precise, let $\mathcal{T}_d$ be a complete $(\beta - 1)$-ary tree of depth $d$, where the root has depth zero. We identify the $(\beta-1)^d$ leaves of $\mathcal{T}_d$ as the left vertices $[n]$, and identify the internal nodes of $\mathcal{T}_d$ as the right vertices $[m]$. From now on, we will use internal nodes for right vertices and leaves for left vertices.

In the bipartite graph $G$, each internal node is connected with all the leaves below it in $\mathcal{T}_d$. It is clear that $G$ is $d$-left-bounded as $\mathcal{T}_d$ has depth $d$. Suppose we removed internal nodes $S$ and obtained leaves $T$ that are not connected to each other in $G$. If $|S| = 0$, then clearly $|T| = 1$. Hence now we assume $|S| \ge 1$. Observe that for any distinct leaves in $T$, their common ancestors in $\mathcal{T}_d$ must be removed in $S$ to disconnect them. Therefore, for any $v \in S$ and its child $v' \in \mathcal{T}_d$, if $v' \notin S$ then at most one leaf in the sub-tree rooted from $v'$ can be contained in $T$. This means that the size of $T$ is upper bounded by the number of leaves in a $(\beta - 1)$-ary tree with at most $|S|$ internal nodes, where the latter is at most $(\beta - 1)|S|$ when $|S| \ge 1$. □

## A.2 Non-Connected Neighborhoods

Now we turn to the tightness of the graph elimination result for non-connected neighborhoods, i.e., Property 4.9 and Corollary 4.11. This accounts for the gigantic dependence on $d$ in the lower bounds for sampling general (Theorem 5.10) and periodic (Theorem 5.17) Hamming slices.

**Property** (Property 4.9 Restated)**.** *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*

$$|S| \le \frac{r}{F(t)} \quad \textit{and} \quad r \ge \frac{n}{\lambda} \quad \textit{and} \quad t \le \kappa.$$

In light of Property 4.9, we present the following construction.

**Lemma A.2.** *Assume $d \ge 2$ is an even number and $n = \mathrm{tow}_2(d/2)$. There exists a $d$-left-bounded bipartite graph $G = ([n], [m], E)$ such that for any $S \subseteq [m]$, deleting those right vertices (and their incident edges) gives at most $\max\{1, |S|\}$ non-connected left neighborhoods.*

The instance $G$ above serves as a counterexample to Property 4.9 if $F(t) > 1$ and $\lambda < n = \mathrm{tow}_2(d/2)$. In addition, by "open-boxing" the construction, there is a right vertex $v^*$ in $G$ incident to all left vertices. Hence $G$ is also a counterexample to Property 4.9 if $F(t) > 1$ and $\kappa < n = \mathrm{tow}_2(d/2)$, since $t \le \kappa < n$ enforces that $v^*$ must be removed. On the other hand, setting $F(t) = O(1)$ in Corollary 4.11, we know that Property 4.9 holds if $\lambda$ and $\kappa$ are indeed a tower of $d$. Hence Lemma A.2 shows that Corollary 4.11 is surprisingly tight.

Recall that in our actual applications, Theorem 5.10 needs to set $F(t) = \Omega(t)$ and Theorem 5.17 needs to set $F(t) = 2^{\Omega(t)}$. Hence the assumption that $F(t) > 1$ is extremely weak. Yet we still cannot hope for a bound without a tower-type blowup on $d$.

Now we prove Lemma A.2. Its construction is similar to the one in Lemma A.1.

*Proof of Lemma A.2.* Define $k = d/2$. Let $\mathcal{L}_k$ be a rooted tree of depth $k$, where the root has depth zero. Each internal node in $\mathcal{L}_k$ of depth $i \le k-1$ has arity $a_i = \frac{\log^{(i)}(n)}{\log^{(i+1)}(n)} = \frac{\mathrm{tow}_2(k-i)}{\mathrm{tow}_2(k-i-1)}$.[14] By construction, there are $b_i = \mathrm{tow}_2(k-i)$ leaves in a sub-tree rooted from a node of $\mathcal{L}_k$ at depth $i$.

Now similar to the proof of Lemma A.1, we identify the $\mathrm{tow}_2(k)$ leaves of $\mathcal{L}_k$ as the left vertices $[n]$, and put the internal nodes of $\mathcal{L}_k$ as right vertices (see Figure 1). Then in the bipartite graph $G$, each internal node is connected with all the leaves below it in $\mathcal{L}_k$. It is clear that, at this point, $G$ is $k$-left-bounded as $\mathcal{L}_k$ has depth $k$.

Suppose that we want to ensure that leaves $T \subseteq [n]$ have non-connected left *neighborhoods*. Then at least, we need $T$ to be a set of non-connected left *vertices*. Hence, analogous to the proof of Lemma A.1, this implies the following claim.

**Claim A.3.** For any distinct leaves in $T$, their common ancestors in $\mathcal{L}_k$ must be removed.

Unfortunately there are not many such common ancestors, since the arities of the internal nodes of $\mathcal{L}_k$ are extremely large. To strengthen the connectivity among left neighborhoods, we now introduce more right vertices.

Fix an arbitrary internal node $v \in \mathcal{L}_k$. Let $0 \le i \le k-1$ be its depth, and let $v_1, v_2, \ldots, v_{a_i} \in \mathcal{L}_k$ be its children, where recall that $a_i = \frac{\mathrm{tow}_2(k-i)}{\mathrm{tow}_2(k-i-1)}$ is its arity. In the current construction, we have only put $v_1, \ldots, v_{a_i}$ as right vertices on top of their *respective* leaves. In other words, $v_1, \ldots, v_{a_i}$ support on disjoint sets of leaves. To enhance the connectivity, we will now create more right vertices

[14] $\log^{(i)}(\cdot)$ is the iterated logarithm of order $i$, where $\log^{(0)}(x) = x$ and $\log^{(i)}(x) = \log(\log^{(i-1)}(x))$ for $i \ge 1$.

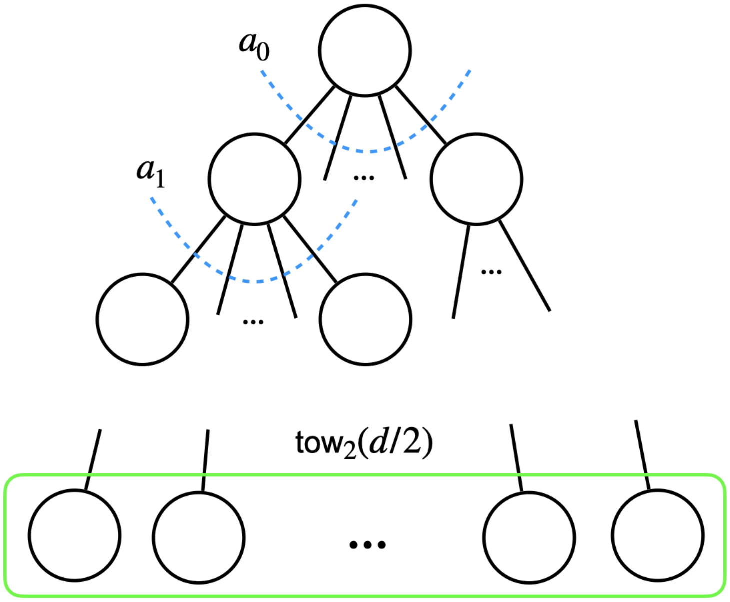

Figure 1: The tree $\mathcal{L}_k$. The vertices in the green box are identified as the left vertices of the bipartite graph $G$, while those outside the box correspond to the right vertices.

to link between these disjoint set of leaves. For each $j \in [a_i]$, we list (in an arbitrary order) the leaves in the sub-tree rooted from $v_j$ as $u_{j,0}, \ldots, u_{j,b_{i+1}-1}$, where recall that $b_{i+1} = \mathrm{tow}_2(k-i-1)$. Then we will build a complete binary tree for $v_1, \ldots, v_{a_i}$, and use the $u_{j,\ell}$'s to spread out the degrees. More precisely, let $\mathcal{B}_v$ be a complete binary tree with $a_i$ leaves, where the $j$-th leaf is equipped with $u_{j,0}, \ldots, u_{j,b_{i+1}-1}$. Since $a_i = \frac{\mathrm{tow}_2(k-i)}{\mathrm{tow}_2(k-i-1)}$ and $0 \le i \le k-1$, $\mathcal{B}_v$ has depth $d_i = \log(a_i) \le b_{i+1}$ where $d_i \ge 1$ is an integer and the root of $\mathcal{B}_v$ root has depth zero. For each internal node $w_v \in \mathcal{B}_v$ of depth $0 \le \ell \le d_i - 1 \le b_{i+1} - 1$, we identify it as a new right vertex in $G$ and add an edge between $w_v$ and $u_{j,\ell}$ for each leaf $j$ below $w_v$ in $\mathcal{B}_v$ (see Figure 2). Observe that different $w_v$ in the same depth will use different index $j$ of the $u$'s, and different $w_v$ of different depths will use different index $\ell$ of the $u$'s. Hence these new right vertices are incident to disjoint sets of input vertices.

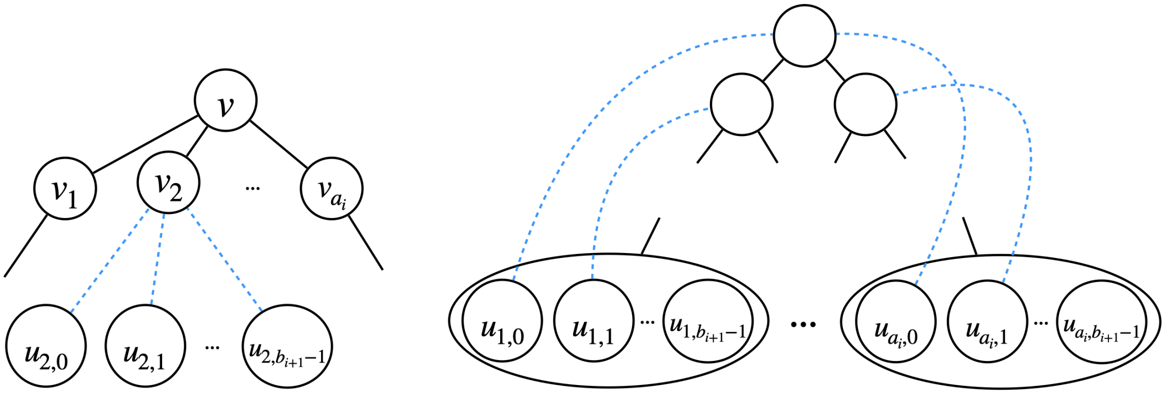

Figure 2: Left: The sub-tree rooted at a node $v \in \mathcal{L}_k$ of depth $i$. The blue dashed edges represent that the $u_{2,j}$'s are leaves in the sub-tree rooted at $v_2$. Right: The corresponding $\mathcal{B}_v$. The blue dashed edges correspond to edges in the bipartite graph $G$.

The construction of $G$ is completed by performing the above procedure for each internal node $v \in \mathcal{L}_k$. During the procedure for $v \in \mathcal{L}_k$, only the degree of leaves (left vertices of $G$) below $v$ gets increased by at most one. Therefore each left vertex of $G$ has total degree at most $2k = d$ as desired.

Recall that we aim to ensure that leaves $T \subseteq [n]$ have non-connected left neighborhoods. Assume that we removed right vertices in $S$ to get this. If $|T| \le 1$, then clearly one does not need to remove any right vertex, i.e., $S = \emptyset$ suffices. Now we assume $|T| \ge 2$. For each $u \in T$, we define $v_u \in \mathcal{L}_k$ to be the ancestor of $u$ that is not removed in $S$ and has the smallest depth. Note that if all ancestors of $u$ are removed, then $v_u = u$ itself. Let $p_u$ be the parent node of $v_u \in \mathcal{L}_k$. Since $|T| \ge 2$ and by Claim A.3, we at least need to remove the root of $\mathcal{L}_k$ and thus $p_u$ always exists.

Now for a fixed internal node $p \in \mathcal{L}_k$, we analyze all $u \in T$ with $p_u = p$. Let $T_p = \{u \in T\colon p_u = p\}$. Recall that we also constructed right vertices based on $\mathcal{B}_p$ to enhance the connectivity of the first-step construction based on $\mathcal{L}_k$. Define $S_p = \{v \in S\colon v = p \vee v \in \mathcal{B}_p\}$. Then it suffices to show $|T_p| \le |S_p|$, since

$$|S| = \sum_{\text{internal node } p \in \mathcal{L}_k} |S_p| \ge \sum_{\text{internal node } p \in \mathcal{L}_k} |T_p| = |T|.$$

Without loss of generality, assume $|T_p| \ge 1$. Firstly, $p \in S_p$ by definition. Now we look at the right vertices in $\mathcal{B}_p$. By Claim A.3, different $u$ have different $v_u$. Therefore, $v_u, v_{u'}$ for distinct $u, u' \in T_p$ correspond to distinct leaves in $\mathcal{B}_p$. Let $w \in \mathcal{B}_p$ be a common ancestor of $v_u, v_{u'}$. Since both $v_u, v_{u'}$ are not removed, $w$ must be removed; otherwise the left neighborhoods of $u$ and $u'$ are connected.

Indeed, the sub-trees (in $\mathcal{L}_k$) rooted at $v_u$ and $v_{u'}$ must have leaves $c_u$ and $c_{u'}$ with edges $(c_u, w)$ and $(c_{u'}, w)$, respectively. Then $u$ connects with $u'$ via $v_u, c_u, w, c_{u'}, v_{u'}$. This means that any common ancestors of $v_u, v_{u'}$ in $\mathcal{B}_p$ for pairs of distinct $u, u' \in T_p$ must be contained in $S_p$. Since $|T_p| \ge 1$ and $\mathcal{B}_p$ is a binary tree, we have at least $|T_p| - 1$ such ancestors. In summary, $|S_p| \ge 1 + (|T_p| - 1) = |T_p|$ as claimed. □

## B Local Limit Theorems on the Additive Group Modulo $q$

In this section, we prove Lemma 3.7, which is a local limit result for $\mathbb{Z}/q\mathbb{Z}$. One can also view it as the variant of Fact 3.4 in the cyclic groups with sharp estimates. The basic strategy is to use Fourier analysis to bound the sum of random variables, where the non-constancy assumption is used to show that certain coefficients are small.

We start by proving a more general statement.

**Theorem B.1.** *Let $q \ge 2$ be an integer, and let $X_1, \ldots, X_n$ be independent random variables in $\mathbb{Z}$. For each $j \in [n]$ and $r \ge 1$, define $p_{r,j} = \max_{x \in \mathbb{Z}} \mathbf{Pr}\left[X_j \equiv x \pmod{r}\right]$. Then for any $\Lambda \subseteq \mathbb{Z}/q\mathbb{Z}$, we have*

$$\mathbf{Pr}\left[\sum_{j \in [n]} X_j \mod q \in \Lambda\right] \le \frac{1}{q} \sum_{s \in [q]} \left(\sum_{a \in T_s} \left|\sum_{c \in \Lambda} \omega_q^{-a \cdot c}\right|\right) \cdot \exp\left\{-\frac{2 \cdot \sum_{j \in [n]}(1 - p_{q/s,j})}{q^2}\right\},$$

*where $\omega_q = e^{2\pi i/q}$ is the primitive $q$-th root of unit, $\gcd(\cdot,\cdot)$ is the greatest common divider function, and $T_s = \{a \in \mathbb{Z}/q\mathbb{Z}\colon \gcd(a, q) = s\}$.*

As a consequence, we finish the proof of Lemma 3.7.

**Lemma** (Lemma 3.7 Restated)**.** *Under the same conditions as Theorem B.1, assume further* $\sum_{j\in[n]}(1-p_{r,j})\ge L$ *holds for all* $r\ge 3$ *dividing* $q$*. Then for* $\Lambda\subseteq\mathbb{Z}/q\mathbb{Z}$*, we have*

$$\mathbf{Pr}\left[\sum_{j\in[n]}X_j \mod q\in\Lambda\right]\le q\cdot e^{-2L/q^2}+\begin{cases}|\Lambda|/q & q\text{ is odd,}\\ 2\cdot\max\left\{|\Lambda_{\mathsf{even}}|,|\Lambda_{\mathsf{odd}}|\right\}/q & q\text{ is even,}\end{cases}$$

*where* $\Lambda_{\mathsf{even}}=\{\text{even numbers in }\Lambda\}$ *and* $\Lambda_{\mathsf{odd}}=\{\text{odd numbers in }\Lambda\}$*.*

*Proof.* We follow the notation convention in Theorem B.1 and define for each $s\in[q]$

$$A_s=\left(\sum_{a\in T_s}\left|\sum_{c\in\Lambda}\omega_q^{-a\cdot c}\right|\right)\cdot\exp\left\{-\frac{2\cdot\sum_{j\in[n]}(1-p_{q/s,j})}{q^2}\right\}.$$

For $s=q$, we clearly have $A_s=|\Lambda|$. If $q$ is even and $s=q/2$, we have $T_s=\{q/2\}$ and

$$A_s\le\left|\sum_{c\in\Lambda}\omega_q^{-c\cdot q/2}\right|=\left||\Lambda_{\mathsf{even}}|-|\Lambda_{\mathsf{odd}}|\right|. \qquad \text{(since } \omega_q^{q/2}=-1\text{)}$$

For any $s\in[q]$ dividing $q$ and not equal to $q$ or $q/2$, we have $q/s\ge 3$ and thus $A_s\le|T_s|\cdot|\Lambda|\cdot e^{-2L/q^2}$. Then the desired bound follows from Theorem B.1 by observing that $|\Lambda|+\left||\Lambda_{\mathsf{even}}|-|\Lambda_{\mathsf{odd}}|\right|=2\cdot\max\left\{|\Lambda_{\mathsf{even}}|,|\Lambda_{\mathsf{odd}}|\right\}$ and $\sum_{s\in[q]}|T_s|\cdot|\Lambda|=q\cdot|\Lambda|\le q^2$. □

**Remark B.2.** We remark that Lemma 3.7 is roughly tight:

- If $q$ is odd, we let each $X_j$ be an unbiased coin. Then $\sum_j X_j \bmod q$ converges to the uniform distribution over $\mathbb{Z}/q\mathbb{Z}$ as $n\to+\infty$, which means the LHS converges to $|\Lambda|/q$.
- If $q$ is even, we let each $X_j$ be uniform in $\{0,2\}$. Then $\sum_j X_j \bmod q$ converges to the uniform distribution over even numbers in $\mathbb{Z}/q\mathbb{Z}$, which means the LHS converges to $2\cdot|\Lambda_{\mathsf{even}}|/q$. By changing $X_1$ to be uniform in $\{1,3\}$, we get the other bound $2\cdot|\Lambda_{\mathsf{odd}}|/q$.

Now we prove Theorem B.1.

*Proof of Theorem B.1.* We first note the following Fourier estimate, which will be proved later.

**Claim B.3.** For each $a\in\mathbb{Z}/q\mathbb{Z}$, let $s=\gcd(a,q)$. We have

$$\left|\mathop{\mathbb{E}}_{X_j}\left[\omega_q^{a\cdot X_j}\right]\right|\le 1-\frac{2\cdot\left(1-p_{q/s,j}\right)}{q^2}\quad\text{for each } j\in[n].$$

Assuming Claim B.3, we now complete the proof of Theorem B.1. Observe that

$$\begin{aligned}\mathbf{Pr}\left[\sum_{j\in[n]}X_j \mod q\in\Lambda\right]&=\mathbb{E}\left[\sum_{c\in\Lambda}\frac{1}{q}\sum_{a\in\mathbb{Z}/q\mathbb{Z}}\omega_q^{a\cdot(\sum_j X_j-c)}\right]\\&=\frac{1}{q}\sum_{a\in\mathbb{Z}/q\mathbb{Z}}\left(\prod_{j\in[n]}\mathop{\mathbb{E}}_{X_j}\left[\omega_q^{a\cdot X_j}\right]\right)\left(\sum_{c\in\Lambda}\omega_q^{-a\cdot c}\right)\\&\le\frac{1}{q}\sum_{a\in\mathbb{Z}/q\mathbb{Z}}\left|\prod_{j\in[n]}\mathop{\mathbb{E}}_{X_j}\left[\omega_q^{a\cdot X_j}\right]\right|\cdot\left|\sum_{c\in\Lambda}\omega_q^{-a\cdot c}\right|.\end{aligned}$$

Now for each $a \in \mathbb{Z}/q\mathbb{Z}$, by Claim B.3 we have

$$\left|\prod_{j\in[n]} \mathop{\mathbb{E}}_{X_j}\left[\omega_q^{a\cdot X_j}\right]\right| \le \prod_{j\in[n]}\left(1-\frac{2\cdot\left(1-p_{q/\gcd(a,q),j}\right)}{q^2}\right) \le \exp\left\{-\frac{2\cdot\sum_{j\in[n]}(1-p_{q/\gcd(a,q),j})}{q^2}\right\},$$

which implies

$$\mathbf{Pr}\left[\sum_{j\in[n]} X_j \mod q \in \Lambda\right] \le \frac{1}{q}\sum_{s\in[q]}\left(\sum_{a\in T_s}\left|\sum_{c\in\Lambda}\omega_q^{-a\cdot c}\right|\right)\cdot\exp\left\{-\frac{2\cdot\sum_{j\in[n]}(1-p_{q/s,j})}{q^2}\right\}. \qquad \square$$

Finally we prove Claim B.3.

*Proof of Claim B.3.* Let $z=\omega_q^\theta$ where $\theta\in\mathbb{R}$ is an arbitrary number. To prove Claim B.3, it suffices to show

$$\Re\left(z\cdot\mathbb{E}\left[\omega_q^{a\cdot X_j}\right]\right)\le 1-\frac{2\cdot\left(1-p_{q/s,j}\right)}{q^2}\quad\text{for all } z, \tag{16}$$

where $\Re(\cdot)$ is the real part of a complex number.

Observe that

$$\Re\left(z\cdot\mathbb{E}\left[\omega_q^{a\cdot X_j}\right]\right)=\mathbb{E}\left[\cos\left(\frac{2\pi(a\cdot X_j+\theta)}{q}\right)\right]=1-2\cdot\mathbb{E}\left[\sin^2\left(\frac{\pi(a\cdot X_j+\theta)}{q}\right)\right]. \tag{17}$$

Now it suffices to lower bound $\mathbb{E}\left[\sin^2\left(\frac{\pi(a\cdot X_j+\theta)}{q}\right)\right]$. Let $\mathcal{E}$ be the event that $(a\cdot X_j+\theta) \bmod q\in(-1/2,1/2]$, where the mod here is over the reals. If $\mathcal{E}$ does not happen, we have

$$\sin^2\left(\frac{\pi\left(a\cdot X_j+\theta\right)}{q}\right)\ge\sin^2\left(\frac{\pi}{2q}\right)\ge\frac{1}{q^2}, \tag{18}$$

where we used the fact that $\sin(\pi x)\ge 2x$ for $0\le x\le 1/2$. Hence

$$\mathbb{E}\left[\sin^2\left(\frac{\pi(a\cdot X_j+\theta)}{q}\right)\right]\ge\frac{1}{q^2}\cdot\mathbf{Pr}\left[\neg\mathcal{E}\right].$$

On the other hand, since both $a\cdot X_j$ and $q$ are integers, there is a unique value $b\in\mathbb{Z}/q\mathbb{Z}$ such that $\mathcal{E}$ holds if and only if $a\cdot X_j\equiv b \pmod q$. Now, if $s=\gcd(a,q)$ does not divide $b$, this can never happen and hence $\mathbf{Pr}[\mathcal{E}]=0$. Otherwise, setting $a'=a/s$ and $b'=b/s$, we have

$$\mathbf{Pr}\left[\mathcal{E}\right]=\mathbf{Pr}\left[a'\cdot X_j\equiv b' \pmod{q/s}\right]=\mathbf{Pr}\left[X_j\equiv(a')^{-1}\cdot b' \pmod{q/s}\right]\le p_{q/s,j},$$

where $(a')^{-1}$ is the inverse of $a'$ modulo $q/s$ and the last inequality uses the assumption in Theorem B.1. Therefore (16) follows by combining the above bound with (17) and (18). $\square$

# C Missing Proofs in Section 4

Here, we put omitted proofs from Section 4.

## Proof of Fact 4.1

*Proof of Fact 4.1.* Fix an event $\mathcal{E}'$ that attains $\mathcal{Q}'(\mathcal{E}') - \mathcal{P}'(\mathcal{E}') = \|\mathcal{P}' - \mathcal{Q}'\|_{\mathsf{TV}}$. Then we have

$$\begin{aligned}
\mathcal{Q}'(\mathcal{E}') - \mathcal{P}'(\mathcal{E}') &= \frac{\mathcal{Q}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{Q}(\mathcal{E})} - \frac{\mathcal{P}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{P}(\mathcal{E})} = \frac{\mathcal{Q}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{Q}(\mathcal{E})} - \frac{\mathcal{P}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{Q}(\mathcal{E})} + \frac{\mathcal{P}(\mathcal{E} \wedge \mathcal{E}') \cdot (\mathcal{P}(\mathcal{E}) - \mathcal{Q}(\mathcal{E}))}{\mathcal{Q}(\mathcal{E}) \cdot \mathcal{P}(\mathcal{E})} \\
&\le \frac{\mathcal{Q}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{Q}(\mathcal{E})} - \frac{\mathcal{P}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{Q}(\mathcal{E})} + \frac{\varepsilon \cdot \mathcal{P}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{Q}(\mathcal{E}) \cdot \mathcal{P}(\mathcal{E})} \quad (\text{since } \mathcal{P}(\mathcal{E}) - \mathcal{Q}(\mathcal{E}) \le \|\mathcal{P} - \mathcal{Q}\|_{\mathsf{TV}} \le \varepsilon) \\
&\le \frac{\mathcal{Q}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{Q}(\mathcal{E})} - \frac{\mathcal{P}(\mathcal{E} \wedge \mathcal{E}')}{\mathcal{Q}(\mathcal{E})} + \frac{\varepsilon}{\mathcal{Q}(\mathcal{E})} \quad (\text{since } \mathcal{P}(\mathcal{E} \wedge \mathcal{E}') \le \mathcal{P}(\mathcal{E})) \\
&\le \frac{\varepsilon}{\mathcal{Q}(\mathcal{E})} + \frac{\varepsilon}{\mathcal{Q}(\mathcal{E})} \quad (\text{since } \mathcal{Q}(\mathcal{E} \wedge \mathcal{E}') - \mathcal{P}(\mathcal{E} \wedge \mathcal{E}') \le \|\mathcal{P} - \mathcal{Q}\|_{\mathsf{TV}} \le \varepsilon) \\
&= \frac{2\varepsilon}{\mathcal{Q}(\mathcal{E})}.
\end{aligned}$$

Applying the data processing inequality concludes the proof. □

## Proof of Claim 4.5

*Proof of Claim 4.5.* Let $\omega_q = e^{2\pi i/q}$ be the primitive $q$-th root of unit. We consider the following quantity

$$Q = \left| \mathop{\mathbb{E}}_{X \sim \mathcal{D}_0} \left[\omega_q^X\right] - \mathop{\mathbb{E}}_{X \sim \mathcal{D}_1} \left[\omega_q^X\right] \right|.$$

On the one hand, we have

$$Q \le \sum_{c \in \mathbb{Z}/q\mathbb{Z}} \left|\omega_q^c \cdot (\mathcal{D}_0(c) - \mathcal{D}_1(c))\right| = \sum_{c \in \mathbb{Z}/q\mathbb{Z}} |\mathcal{D}_0(c) - \mathcal{D}_1(c)| = 2 \cdot \|\mathcal{D}_0 - \mathcal{D}_1\|_{\mathsf{TV}}. \tag{19}$$

On the other hand, we have

$$\begin{aligned}
Q &= \left|(1-\gamma+\gamma\cdot\omega_q)^{t-1} - \omega_q \cdot (1-\gamma+\gamma\cdot\omega_q)^{t-1}\right| \qquad (\text{by the definition of } \mathcal{D}_0 \text{ and } \mathcal{D}_1) \\
&= |1-\omega_q| \cdot |1-\gamma+\gamma\cdot\omega_q|^{t-1}. \qquad (20)
\end{aligned}$$

Let $r = \sin^2\left(\frac{\pi}{q}\right)$. Then

$$|1-\omega_q| = \sqrt{\left(1-\cos\left(\frac{2\pi}{q}\right)\right)^2 + \sin^2\left(\frac{2\pi}{q}\right)} = 2 \cdot \left|\sin\left(\frac{\pi}{q}\right)\right| = 2\sqrt{r}$$

and

$$\begin{aligned}
|1-\gamma+\gamma\cdot\omega_q| &= \sqrt{\left(1-\gamma+\gamma\cdot\cos\left(\frac{2\pi}{q}\right)\right)^2 + \gamma^2\cdot\sin^2\left(\frac{2\pi}{q}\right)} \\
&= \sqrt{1-4\gamma(1-\gamma)\cdot\sin^2\left(\frac{\pi}{q}\right)} = \sqrt{1-4\gamma(1-\gamma)r}.
\end{aligned}$$

Combining these with (19) and (20), we have

$$\|\mathcal{D}_0 - \mathcal{D}_1\|_{\mathsf{TV}} \ge \sqrt{r\cdot(1-4\gamma(1-\gamma)r)^{t-1}} =: \sqrt{F}. \tag{21}$$

If $q = 3$, then $r = \sin^2(\pi/q) = 3/4$. Hence

$$F = \frac{3}{4} \cdot (1 - 3\gamma(1-\gamma))^{t-1} \geq \frac{3}{4} \cdot 2^{-8\gamma(1-\gamma)(t-1)} \geq \frac{3}{4} \cdot 2^{-8\gamma(t-1)} \geq \frac{4}{9} \cdot 2^{-100\gamma(t-1)/9}$$

where we used the fact that $3\gamma(1-\gamma) \leq 3/4$ and $(1-x)^{1/x} \geq 2^{-8/3}$ for $x \leq 3/4$. Otherwise $q \geq 4$. We use a different inequality that $2x \leq \sin(\pi x) \leq \pi x$ for $0 \leq x \leq 1/2$, and obtain $4/q^2 \leq r \leq 10/q^2$. Hence

$$F \geq \frac{4}{q^2} \cdot \left(1 - \frac{40\gamma(1-\gamma)}{q^2}\right)^{t-1} \geq \frac{4}{q^2} \cdot 2^{-100\gamma(1-\gamma)(t-1)/q^2} \geq \frac{4}{q^2} \cdot 2^{-100\gamma(t-1)/q^2},$$

where we used the fact that $40\gamma(1-\gamma)/q^2 \leq 10/q^2 \leq 5/8$ and $(1-x)^{1/x} \geq 2^{-10/4}$ for $x \leq 5/8$. Putting these back to (21) gives

$$\|\mathcal{D}_0 - \mathcal{D}_1\|_{\mathsf{TV}} \geq \frac{2}{q} \cdot 2^{-50\gamma(t-1)/q^2}$$

as desired. □

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il