Electronic Colloquium on Computational Complexity, Report No. 75 (2026)

On the Advantage of Adaptivity for Sampling with Cell Probes

Farzan Byramji* Daniel M. Kane† Jackson Morris‡ Anthony Ostuni§

Abstract

We construct an explicit distribution $\mathbf{D}$ over $\{0,1\}^N$ that exhibits an essentially optimal separation between adaptive and non-adaptive cell-probe sampling. The distribution can be sampled exactly when each output bit is allowed two adaptive probes to an arbitrarily long sequence of independent uniform symbols from $[N]$. In contrast, any non-adaptive sampler requires $\widetilde{\Omega}(N)$ non-adaptive cell probes to generate a distribution with total variation distance less than $1 - o(1)$ from $\mathbf{D}$. This provides a 2-vs-$\widetilde{\Omega}(N)$ separation for sampling with adaptive versus non-adaptive cell probes, improving upon the 2-vs-$\widetilde{\Omega}(\log N)$ separation of Yu and Zhan (ITCS '24) and the $(\log N)^{O(1)}$-vs-$N^{\Omega(1)}$ separation of Alekseev, Göös, Myasnikov, Riazanov, and Sokolov (STOC '26).

1 Introduction

In recent years, the family of computational tasks involving *sampling* has received increased attention, particularly from a complexity-theoretic viewpoint. Broadly, the goal is to take uniform randomness as input and produce samples from some target distribution. Sampling has connections and applications to many other areas of complexity including data structure lower bounds [Vio12, LV11, BIL12, Vio20, CGZ22, Vio23, YZ24, KOW24, AGM+26], quantum-classical separations [BWP26, KOW24, GKM+26], randomness extractors [DW12, Vio14, CGZ22, BKMO26], and more.

One may study the complexity of sampling tasks in various computational models (e.g., circuits, branching programs, bounded-space Turing machines) and with a variety of randomness sources, but in this work we focus on the *cell-probe* model with uniformly random inputs. In this setting the sampler has query access to arbitrarily many independent samples drawn uniformly from $[N] := \{1, 2, \ldots, N\}$, and outputs are determined by arbitrary functions depending on some subset of inputs queried, i.e., our sampler corresponds to some function $f\colon [N]^m \to \{0,1\}^N$, and we consider the distribution on $\{0,1\}^N$ which results from evaluating f on uniformly random input. We write $f(\mathbf{U}^m_{[N]})$ to denote this distribution where $\mathbf{U}^m_{[N]}$ is the uniform distribution on $[N]^m$. We say that f is d-local if every output bit is a function of at most d input symbols; that is, it can be computed with at most d non-adaptive cell probes.

We also consider an adaptive version of this model wherein the sampler can perform intermediate computations with later queries depending on the results of earlier ones. Indeed, the focus of this work is on understanding the relative power of adaptivity for sampling tasks in this model:

*UC San Diego. Email: fbyramji@ucsd.edu. Supported by Simons Investigator Award #929894, and NSF Awards CCF-2425349 and AF: Medium 2212136.

†UC San Diego. Email: dakane@ucsd.edu. Supported by NSF Medium Award CCF-2107547.

‡UC San Diego. Email: jrm035@ucsd.edu.

§UC San Diego. Email: aostuni@ucsd.edu.

ISSN 1433-8092

How much can adaptive queries help to sample in the cell-probe model?

The challenge of exhibiting a distribution which requires asymptotically fewer adaptive queries than non-adaptive queries was raised by Viola in [Vio20] and reiterated in [Vio23]. He suggested the following candidate distribution $\mathbf{D}_V$ on $[N]^N$. Fix some integer $k = N^{1-\Omega(1)}$, and sample a string $r \in [N]^k$ uniformly at random. Each output symbol x_i of $\mathbf{D}_V$ is determined by independently and uniformly sampling an index $j \in [k]$ and setting $x_i = r_j$. It is easy to see that two adaptive probes are sufficient to sample from this distribution (assuming $k|N$), but a separation was not known until Yu and Zhan [YZ24] proved that $\Omega(\log N/\log\log N)$ non-adaptive probes are necessary to sample from $\mathbf{D}_V$ (even approximately).

More recently, Alekseev, Göös, Myasnikov, Riazanov, and Sokolov [AGM+26] showed that sampling a uniform random permutation in $[N]^N$ also gives a separation between adaptive and non-adaptive cell-probe samplers. They proved that $N^{\Omega(1)}$ non-adaptive probes are required, while Czumaj [Czu15] showed that $O(\log^2 N)$ adaptive probes suffice to get polynomially small distance. Alekseev et al. [AGM+26] also showed that $(\log N)^{\Omega(1)}$ adaptive probes are necessary, and highlighted obtaining an $O(1)$-vs-$N^{\Omega(1)}$ separation as an interesting open question. Weaker lower bounds on the number of non-adaptive probes required to sample permutations had been shown earlier in [Vio20, YZ24].

In this work, we give an essentially optimal $O(1)$-vs-$\widetilde{\Omega}(N)$ separation.

Theorem 1.1. *For any $\varepsilon > 0$ and infinitely many N, there exists an explicit distribution $\mathbf{D}_1$ over $\{0,1\}^N$ which can be sampled exactly by two adaptive probes to independent uniform symbols from $[N]$, yet for any $(N^{1-\varepsilon})$-local function $f\colon [N]^m \to \{0,1\}^N$, we have*

$$\left\| f(\mathbf{U}_{[N]}^m) - \mathbf{D}_1 \right\|_{\mathsf{TV}} \geq 1 - \exp(-N^{\Omega_\varepsilon(1)}). \tag{1}$$

Similarly, there exists a constant $c > 0$ such that for infinitely many N, there exists an explicit distribution $\mathbf{D}_2$ over $\{0,1\}^N$ which can be sampled exactly by two adaptive probes to independent uniform symbols from $[N]$, yet for any $(N/\log^c N)$-local function $g\colon [N]^m \to \{0,1\}^N$, we have

$$\left\| g(\mathbf{U}_{[N]}^m) - \mathbf{D}_2 \right\|_{\mathsf{TV}} \geq 1 - \exp(-\Omega(\log^2 N)). \tag{2}$$

Remark 1.2. These results are essentially the best possible in several ways.

- Any adaptive sampler which only makes one cell probe (per output bit) is trivially equivalent to a non-adaptive sampler, so at least two probes are necessary for a separation.
- The error in (1) cannot be appreciably improved, since any distribution over $\{0,1\}^N$ can be approximated to distance $1 - e^{-\Omega(N)}$ by a point mass (which needs no locality to generate).
- The locality in (2) cannot be appreciably improved, since a $\Theta(N/\log N)$-local sampler can use the probed cells as a shared seed with $2^{\Theta(N)}$ possible values to approximate any distribution over $\{0,1\}^N$ with exponentially small error (see, e.g., [Vio12, Lemma 5.2]).
- The alphabet size s of the random symbols must be large for such a separation to hold. If s were smaller than any polynomial in N, then any sampler making $d = O(1)$ adaptive probes could be simulated by one making $s^{O(d)} \ll N^{\Omega(1)}$ non-adaptive probes by considering all possible queried inputs.

The distributions in Theorem 1.1 are instantiations of the following distribution with different choices of parameters.

Definition 1.3 (Distribution $\mathbf{D}_{n,\ell}$)**.** Let $n \geq 2$ be a power of two, and let $\ell \geq 1$ be an integer. We define $\mathbf{D}_{n,\ell}$ to be the distribution over $(\mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(\ell)}, \mathbf{x}, \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(\ell)}) \in \{0,1\}^{\ell \log(n) + (\ell+1)n}$, where each $\mathbf{i}^{(j)}$ is independently uniform over $\{0,1\}^{\log(n)}$ (viewed as an element of $[n]$), $\mathbf{x}$ is uniform over $\{0,1\}^n$, and $\mathbf{y}^{(j)}_k = \mathbf{x}_{k+\mathbf{i}^{(j)} \bmod n}$ for all $j \in [\ell]$ and $k \in [n]$.

Note that $\mathbf{D}_{n,\ell}$ is similar in spirit to the aforementioned distribution $\mathbf{D}_V$, but with a more explicit indexing structure. In $\mathbf{D}_V$, each output samples a private index $j \in [k]$ and outputs r_j. This produces clusters of coordinates which must take the same value, although the output itself does not reveal which coordinates belong to which cluster. Our method instead explicitly produces many random groups of $\ell + 1$ outputs that must correlate and encodes which clusters those must be in the index portion of the output.

As with $\mathbf{D}_V$, the distribution $\mathbf{D}_{n,\ell}$ can be sampled with two adaptive cell probes.

Claim 1.4. Let n, ℓ be as in Definition 1.3, and let N be a (positive) integer multiple of $2n$. Then $\mathbf{D}_{n,\ell}$ can be sampled exactly by two adaptive probes to independent uniform symbols from $[N]$.

Proof. Let the first $\ell + n$ random input symbols be $S_1, \ldots, S_\ell, R_1, \ldots, R_n \in [N]$. For each $j \in [\ell]$, use S_j to generate the shift $\mathbf{i}^{(j)}$ using some balanced mapping from $[N] \to [n]$. Similarly for each $k \in [n]$, use R_k to generate the bit $\mathbf{x}_k$ using some balanced mapping from $[N] \to \{0,1\}$. (Since $(2n) | N$, both mappings exist.) To produce $\mathbf{y}^{(j)}_k$, first probe S_j to learn the shift $\mathbf{i}^{(j)}$, then probe $R_{k+\mathbf{i}^{(j)} \bmod n}$ and output the corresponding bit. □

Our main technical result is the following lower bound, which shows that $\mathbf{D}_{n,\ell}$ is hard to sample non-adaptively.

Theorem 1.5. *Let $m \geq d \geq 1$, $\ell \geq 1$, and $n, N \geq 2$ be integers where n is a power of two. Additionally, let $f \colon [N]^m \to \{0,1\}^{\ell \log(n) + (\ell+1)n}$ be a d-local function, and let $\mathbf{D}_{n,\ell}$ be the distribution in Definition 1.3. Then $f(\mathbf{U}^m_{[N]})$ and $\mathbf{D}_{n,\ell}$ have total variation distance at least*

$$1 - \exp\left(-\Omega\left(\frac{n}{\ell^2 d \log N}\right)\right) - d\left(\frac{\Theta(d \log N)}{n}\right)^{\ell} - \exp\left(d\ell \log(n) \log(N) - \Omega(n)\right).$$

Taking Theorem 1.5 as a black-box, we can immediately derive Theorem 1.1.

Proof of Theorem 1.1 (assuming Theorem 1.5). Both $\mathbf{D}_1$ and $\mathbf{D}_2$ will be chosen to be padded versions of $\mathbf{D}_{n,\ell}$ (for different values of ℓ), where $n \geq 2$ is a power of 2, $\ell \geq 1$ is an integer, and $N = 4\ell n$. Note that under these conditions, Claim 1.4 guarantees we can exactly sample $\mathbf{D}_{n,\ell}$ with two adaptive cell probes, so the same must also be true when the distribution is padded with zeros.

In the case of $\mathbf{D}_1$, we set $\delta = 3\varepsilon/4$, $\ell = \lfloor n^{\delta/3} \rfloor$, and $d = \lceil n^{1-\delta} \rceil$. Then Theorem 1.5 implies that for any d-local $f \colon [N]^m \to \{0,1\}^{\ell \log n + (\ell+1)n}$, we have

$$\left\| f(\mathbf{U}^m_{[N]}) - \mathbf{D}_{n,\ell} \right\|_{\mathsf{TV}} \geq 1 - \exp(-\widetilde{\Omega}(n^{\delta/3})) \geq 1 - \exp(-\Omega(N^{\varepsilon/6}))$$

for large enough n. Since $d \geq N^{1-\varepsilon}$ and $\mathbf{D}_{n,\ell}$ is a projection of $\mathbf{D}_1$, the desired lower bound for $\mathbf{D}_1$ follows. Similarly in the case of $\mathbf{D}_2$, we set $\ell = \lfloor \log^2 n \rfloor$ and $d = \lceil n / \log^7 n \rceil$. Then for any d-local g, we have

$$\left\| g(\mathbf{U}^m_{[N]}) - \mathbf{D}_{n,\ell} \right\|_{\mathsf{TV}} \geq 1 - \exp(-\Omega(\log^2 n)) \geq 1 - \exp(-\Omega(\log^2 N))$$

for large enough n. Using $d \geq N / \log^{10} N$, we obtain the desired lower bound for $\mathbf{D}_2$. □

We will prove Theorem 1.5 in Section 3 after setting up some preliminary material in the subsequent section.

2 Preliminaries

In this section, we review some basic notation, terminology, and standard results. For a positive integer n, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. All logarithms given in the paper are base 2. The indicator function is denoted by $\mathbb{1}(\cdot)$.

Asymptotics. We use the standard $\Omega(\cdot), O(\cdot), \Theta(\cdot)$ asymptotic notation to hide universal positive constants. Occasionally, we will use subscripts to indicate an unspecified dependence on a particular parameter (e.g., $\Omega_\varepsilon(1)$). Additionally, we write $\widetilde{\Omega}(\cdot)$ to suppress polylogarithmic factors.

Cell-Probe Samplers. We study sampling in the cell-probe model. A sampler is given access to an arbitrarily long sequence of independent random cells, each uniformly distributed over $[N]$, and produces an output in $\{0,1\}^n$ according to probes made to the random cells. The complexity of the sampler is the maximum number of cells queried in order to produce any single output bit.

A central distinction is whether these probes are adaptive. In an adaptive sampler, the location of a later probe may depend on the values observed in earlier probes. In a non-adaptive sampler, all probe locations for a given output bit are fixed in advance. In the non-adaptive setting, this cost is often called the *locality* of the sampler. Thus, a function $f\colon [N]^m \to \{0,1\}^n$ is said to be d-local if each output bit of f depends on at most d input cells.

Probability. We use bold letters to denote probability distributions and random variables. We reserve $\mathbf{U}_{[n]}$ for the uniform distribution over $[n]$. For an event $\mathcal{E}$, we define $\mathbf{X}(\mathcal{E})$ to be the probability mass assigned to $\mathcal{E}$ by $\mathbf{X}$. For a function f, we use $f(\mathbf{X})$ to denote the output distribution of $f(\mathbf{x})$ on randomly drawn $\mathbf{x} \sim \mathbf{X}$.

Given a distribution $\mathbf{X}$ and positive integer t, we use $\mathbf{X}^t$ to denote the t-fold product distribution $\mathbf{X} \times \cdots \times \mathbf{X}$. If s is a finite set, we write $\mathbf{X}^s$ to emphasize that the coordinates of $\mathbf{X}^{|s|}$ are indexed by s. We refer to $\mathbf{X}$ as a mixture if it can be written as a convex combination of other distributions. That is, there exist $c_1, \ldots, c_k \in [0,1]$ with $\sum_i c_i = 1$ and distributions $\mathbf{X}_1, \ldots, \mathbf{X}_k$ such that $\mathbf{X}(\mathcal{E}) = \sum_{i=1}^k c_i \cdot \mathbf{X}_i(\mathcal{E})$ for every event $\mathcal{E}$. Occasionally, we write this more concisely as $\mathbf{X} = \sum_{i=1}^k c_i \mathbf{X}_i$. We measure the similarity of two (discrete) distributions $\mathbf{P}$ and $\mathbf{Q}$ by the *total variation (TV) distance*

$$\|\mathbf{P} - \mathbf{Q}\|_{\mathsf{TV}} = \max_{\text{event } \mathcal{E}} \mathbf{P}(\mathcal{E}) - \mathbf{Q}(\mathcal{E}) = \frac{1}{2} \sum_x |\mathbf{P}(x) - \mathbf{Q}(x)| .$$

We say $\mathbf{P}$ is ε-*close* to $\mathbf{Q}$ if $\|\mathbf{P} - \mathbf{Q}\|_{\mathsf{TV}} \le \varepsilon$, and ε-*far* otherwise.

Concentration Inequalities. We will need the following standard concentration inequality.

Lemma 2.1 (Chernoff)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent indicator random variables with $\mathbf{Pr}[\mathbf{X}_i = 1] \ge p$ for all i. Then for any $\delta \in (0,1)$, we have*

$$\mathbf{Pr}\left[\sum_i \mathbf{X}_i \le (1-\delta)pn\right] \le \exp\left(-\Omega(\delta^2 pn)\right) .$$

We will also require a variant of Lemma 2.1 in which the random variables are only partially independent. There are a plethora of standard results giving concentration under these weaker conditions (e.g., McDiarmid's or Azuma's inequalities). For our purposes, a lesser-known inequality

derived from Shearer's lemma (from information theory) by Gavinsky, Lovett, Saks, and Srinivasan [GLSS15] appears to provide the best bounds. A collection of random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is said to be a *read-k family* if they are functions of independent random variables $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$, where each $\mathbf{Y}_j$ influences at most k of the $\mathbf{X}_i$'s. The following is essentially [GLSS15, Theorem 1.1].

Lemma 2.2 (Read-k Chernoff)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a read-k family of indicator random variables with $\mathbf{Pr}[\mathbf{X}_i = 1] \geq p$ for all i. Then for any $\delta \in (0, 1)$, we have*

$$\mathbf{Pr}\left[\sum_i \mathbf{X}_i \leq (1-\delta)pn\right] \leq \exp\left(-\Omega\left(\frac{\delta^2 pn}{k}\right)\right).$$

3 Proof of Theorem 1.5

In this section, we prove Theorem 1.5, restated below for the reader's convenience.

Theorem 1.5. *Let $m \geq d \geq 1$, $\ell \geq 1$, and $n, N \geq 2$ be integers where n is a power of two. Additionally, let $f\colon [N]^m \to \{0,1\}^{\ell \log(n) + (\ell+1)n}$ be a d-local function, and let $\mathbf{D}_{n,\ell}$ be the distribution in Definition 1.3. Then $f(\mathbf{U}^m_{[N]})$ and $\mathbf{D}_{n,\ell}$ have total variation distance at least*

$$1 - \exp\left(-\Omega\left(\frac{n}{\ell^2 d \log N}\right)\right) - d\left(\frac{\Theta(d \log N)}{n}\right)^{\ell} - \exp\big(d\ell \log(n) \log(N) - \Omega(n)\big).$$

At a high level, the proof reduces to a setting in which many groups of output bits that are forced to be equal under $\mathbf{D}_{n,\ell}$ do not share a common input cell. For each such group, this absence of a shared input creates a dichotomy: either the sampler produces unequal bits with noticeable probability, immediately violating the defining constraints of $\mathbf{D}_{n,\ell}$, or the entire group is biased toward one fixed constant value. If the latter occurs for many groups, then the sampler places too much mass on outputs where many coordinates simultaneously take prescribed values, an event that is exponentially unlikely under $\mathbf{D}_{n,\ell}$. Thus in either case the sampler is far from the target distribution. We now proceed to the details.

For clarity, we will write $\mathbf{U}$ for $\mathbf{U}_{[N]}$ and $\mathbf{D}$ for $\mathbf{D}_{n,\ell}$. It will also be convenient to discuss the indices in the output of $f(\mathbf{U}^m)$ in a similar fashion to the definition of $\mathbf{D}$ (in Definition 1.3). That is, we will abuse notation and often use (say) $\mathbf{i}^{(1)}$ for the first $\log n$ bits of $f(\mathbf{U}^m)$. It will be evident from context which distribution (i.e., $f(\mathbf{U}^m)$ or $\mathbf{D}$) the random variables $\mathbf{i}^{(j)}, \mathbf{x}, \mathbf{y}^{(j)}$ are being taken from.

Let $t > 0$ be a threshold parameter to be defined later, and let $S \subseteq [m]$ be the set of input symbols to the sampler f that either affect more than t output bits or affect one of the first $\ell \log n$ outputs (corresponding to $\mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(\ell)}$). Observe that $s := |S| \leq d\ell \log(n) + d(\ell+1)n/t$. By conditioning on the values that the symbols in S take, we can express $f(\mathbf{U}^m)$ as the mixture

$$f(\mathbf{U}^m) = \sum_{\rho \in [N]^S} \frac{1}{N^s} \cdot f(\mathbf{U}^{[m]\setminus S}, \rho).$$

Below, we will write the distribution $f(\mathbf{U}^{[m]\setminus S}, \rho)$ more concisely as $\mathbf{F}_\rho$. Observe that any such conditioning ρ fixes the value of the shift vector $\mathbf{i}(\rho) := (\mathbf{i}^{(1)}(\rho), \ldots, \mathbf{i}^{(\ell)}(\rho))$, where $\mathbf{i}^{(j)}(\rho)$ is the value of $\mathbf{i}^{(j)}$ after conditioning on ρ, and it guarantees that every unfixed input cell affects at most t output bits. We will use these properties to show that for certain restrictions ρ, the distribution $\mathbf{F}_\rho$ is far from $\mathbf{D}$ conditioned on their shift vectors being equal. Then, we amalgamate the behavior of these conditional distributions to obtain our final distance bound.

In order for $f(\mathbf{U}^m)$ to approximate $\mathbf{D}$, there must be clusters of $k := \ell+1$ output bits which all take the same value. More precisely, for each shift vector $i \in (\{0,1\}^{\log n})^\ell$ and $u \in [n]$, define the equality block $B_u(i) = (\mathbf{x}_u, \mathbf{y}^{(1)}_{u-i^{(1)}}, \ldots, \mathbf{y}^{(\ell)}_{u-i^{(\ell)}})$. Call such an i *bad* if at least $n/4$ of the equality blocks have some unfixed input cell $c \notin S$ which affects every bit in the block (and *good* otherwise); if a block has such an input symbol, we say the block is *covered*.

Claim 3.1. The fraction of bad $i \in (\{0,1\}^{\log n})^\ell$ is at most $4d\left(\frac{t}{\ell n}\right)^\ell$.

Proof. For each unfixed input cell $c \notin S$, define $A_{x,c} \subseteq [n]$ to be the set of $\mathbf{x}$ coordinates affected by c. Similarly, for all $j \in [\ell]$ define $A_{j,c} \subseteq [n]$ to be the set of $\mathbf{y}^{(j)}$ coordinates affected by c. Observe that $\sum_j |A_{j,c}| \le t$, and that a block $B_u(i)$ is covered by a cell c only if $u \in A_{x,c}$ and $u - i^{(j)} \in A_{j,c}$ for all j. In a uniformly sampled $\mathbf{i}$, each $\mathbf{i}^{(j)}$ is chosen independently and uniformly at random from $\{0,1\}^{\log n}$, so the probability that $u - \mathbf{i}^{(j)} \in A_{j,c}$ is $|A_{j,c}|/n$. Thus,

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{i}\in(\{0,1\}^{\log n})^\ell}[\#\text{ blocks covered}] &\le \sum_{c\notin S} \mathop{\mathbb{E}}_{\mathbf{i}\in(\{0,1\}^{\log n})^\ell}[\#\text{ blocks covered by } c] \\
&= \sum_{c\notin S}\sum_{u\in A_{x,c}} \prod_{j=1}^{\ell} \frac{|A_{j,c}|}{n} && \text{(since } \mathbf{i}^{(j)}\text{'s independent)} \\
&\le \sum_{c\notin S} |A_{x,c}| \left(\frac{\sum_j |A_{j,c}|}{n\ell}\right)^\ell && \text{(by AM-GM inequality)} \\
&\le \left(\sum_{c\notin S} |A_{x,c}|\right) \cdot \left(\frac{t}{n\ell}\right)^\ell && \text{(by def'n of } S\text{)} \\
&\le dn\left(\frac{t}{n\ell}\right)^\ell. && \text{(since } \mathbf{x} \text{ affected by } \le dn \text{ inputs)}
\end{aligned}$$

Applying Markov's inequality concludes the proof. □

By Claim 3.1, we can primarily focus our attention on the case of good i's, where we have many uncovered blocks. We will show that the lack of correlation within the bits of such a block causes behavior which does not occur in the target distribution. Below, we view $\mathbf{Z}$ as the marginal distribution of $\mathbf{F}_\rho$ onto an uncovered block.

Claim 3.2. Let $\mathbf{Z}$ be the marginal distribution over some $k \ge 2$ output bits. If no input cell affects every bit in $\mathbf{Z}$, then either

1. $\mathbf{Pr}[\mathbf{Z} \notin \{0^k, 1^k\}] \ge 1/(100k)$, or
2. $\mathbf{Pr}[\mathbf{Z} = z] \ge 2/3$ for some fixed $z \in \{0^k, 1^k\}$.

Proof. Arbitrarily partition the input cells into sets $I_1, \ldots, I_k$ where the cells in I_j do not affect the j-th output bit of $\mathbf{Z}$. We proceed by a hybrid argument. Consider two independent random inputs $\mathbf{A}, \mathbf{B}$. We define the sequence of inputs

$$\mathbf{A} = \mathbf{C}^{(0)}, \mathbf{C}^{(1)}, \ldots, \mathbf{C}^{(k)} = \mathbf{B},$$

where $\mathbf{C}^{(j)}$ has the input symbols in $I_1 \cup \cdots \cup I_j$ taken from $\mathbf{B}$ and the rest from $\mathbf{A}$. We also define $\mathbf{Z}^{(j)}$ to be the value of $\mathbf{Z}$ on input $\mathbf{C}^{(j)}$, and observe that each $\mathbf{Z}^{(j)}$ has the same marginal distribution as $\mathbf{Z}$. Moreover, let $\delta = \mathbf{Pr}[\mathbf{Z} \notin \{0^k, 1^k\}]$, $p_0 = \mathbf{Pr}[\mathbf{Z} = 0^k]$, and $p_1 = \mathbf{Pr}[\mathbf{Z} = 1^k]$.

Whenever $\mathbf{Z}^{(0)} = 0^k$ and $\mathbf{Z}^{(k)} = 1^k$ (or vice-versa), there must exist some $\mathbf{Z}^{(j)} \notin \{0^k, 1^k\}$. Indeed, any two consecutive outputs $\mathbf{Z}^{(j-1)}, \mathbf{Z}^{(j)}$ agree on their j-th bit, since the cells in I_j do not affect it. Thus, we cannot transition directly from 0^k to 1^k, so if the sequence contains both strings, then it must also contain some other output. This implies $2p_0p_1 \leq (k+1)\delta$. Assume by contradiction that $\delta < 1/(100k)$ and $p_0, p_1 < 2/3$. Then $\min(p_0, p_1) > 1 - \frac{1}{100k} - \frac{2}{3} \geq \frac{1}{4}$, so

$$\frac{1}{8} \leq 2\left(\frac{1}{4}\right)^2 < \frac{k+1}{100k} \leq \frac{1}{100} + \frac{1}{100k} \leq \frac{1}{8} - \frac{11}{100},$$

a contradiction. □

Fix a conditioning ρ such that $\mathbf{i}(\rho)$ is good. That is, at least $3n/4$ blocks $B_u(\mathbf{i}(\rho))$ are uncovered. We break into cases depending on which conclusion of Claim 3.2 is satisfied by most of these blocks. Below, let $n' = 3n/8$.

Case 1: Most blocks satisfy (1). Let $B_1, \ldots, B_{n'}$ be blocks satisfying (1). While they are not fully independent, one would expect their behavior to resemble that of independent random variables, since each unfixed input cell only affects a small number of blocks.[1] More formally, we have that the indicator random variables $\{\mathbb{1}(B_j \notin \{0^k, 1^k\})\}_j$ form a read-t family, so Lemma 2.2 implies

$$\Pr_{\mathbf{F}_\rho}\left[\sum_{j=1}^{n'} \mathbb{1}(B_j \notin \{0^k, 1^k\}) \geq \frac{n'}{200k}\right] \geq 1 - \exp\left(-\Omega\left(\frac{n}{kt}\right)\right).$$

For comparison, this event never occurs under $\mathbf{D}$ conditioned on its shift vector being $\mathbf{i}(\rho)$.

Case 2: Most blocks satisfy (2). Let $B_1, \ldots, B_{n'}$ be blocks satisfying (2), where their most common values are $b_1, \ldots, b_{n'}$, respectively. Again applying Lemma 2.2, we have

$$\Pr_{\mathbf{F}_\rho}\left[\sum_{j=1}^{n'} \mathbb{1}(B_j = b_j) \geq \frac{7n'}{12}\right] \geq 1 - \exp\left(-\Omega\left(\frac{n}{t}\right)\right).$$

For comparison, $\sum_{j=1}^{n'} \mathbb{1}(B_j = b_j)$ is distributed like the binomial $\mathsf{Bin}(n', 1/2)$ under $\mathbf{D}$ conditioned on its shift vector being $\mathbf{i}(\rho)$, so this event occurs with probability at most $e^{-\Omega(n)}$ by Lemma 2.1.

In either case, there exists an event $\mathcal{E}_\rho$ which occurs with probability at least $1 - e^{-\Omega\left(\frac{n}{kt}\right)}$ under $\mathbf{F}_\rho$, but probability at most $e^{-\Omega(n)}$ under $\mathbf{D}$ conditioned on its shift vector. We now define the global event $\mathcal{E}$ witnessing the TV distance between $f(\mathbf{U}^m)$ and $\mathbf{D}$ to be

$$\mathcal{E} = \{\text{bad } \mathbf{i}\} \cup \bigcup_{\rho\,:\,\mathbf{i}(\rho) \text{ is good}} (\mathbf{i} = \mathbf{i}(\rho) \text{ and } \mathcal{E}_\rho).$$

Under $f(\mathbf{U}^m)$, each conditioning $\rho \in [N]^S$ produces a shift vector $\mathbf{i}(\rho)$. If $\mathbf{i}(\rho)$ is bad, then the output automatically lies in $\mathcal{E}$, and if $\mathbf{i}(\rho)$ is good, then we have already shown it lies in $\mathcal{E}$ with

[1] A greedy construction allows us to find $\Omega(n/dt)$ actually independent blocks, but pursuing this line of analysis gives worse bounds.

probability at least $1 - e^{-\Omega(\frac{n}{kt})}$. To analyze the probability under $\mathbf{D}$, we apply a union bound to find that

$$\begin{aligned}\Pr_{\mathbf{D}}[\mathcal{E}] &\le \Pr_{\mathbf{D}}[\mathbf{i} \text{ is bad}] + \sum_{\rho\,:\,\mathbf{i}(\rho) \text{ is good}} \Pr_{\mathbf{D}}\left[\mathcal{E}_\rho \,|\, \mathbf{i} = \mathbf{i}(\rho)\right] \\ &\le 4d\left(\frac{t}{\ell n}\right)^{\ell} + N^s \exp(-\Omega(n)). \qquad \text{(by Claim 3.1)}\end{aligned}$$

Hence,

$$\|f(\mathbf{U}^m) - \mathbf{D}\|_{\mathsf{TV}} \ge 1 - \exp\left(-\Omega\left(\frac{n}{(\ell+1)t}\right)\right) - 4d\left(\frac{t}{\ell n}\right)^{\ell} - N^s \exp(-\Omega(n)).$$

Setting $t = \Theta(\ell d \log N)$ with a sufficiently large implicit constant gives $s \le d\ell \log(n) + O(n/\log N)$, where $O(n/\log N)$ has a sufficiently small implicit constant, so we find the distance between $f(\mathbf{U}^m)$ and $\mathbf{D}$ is at least

$$1 - \exp\left(-\Omega\left(\frac{n}{\ell^2 d \log N}\right)\right) - d\left(\frac{\Theta(d \log N)}{n}\right)^{\ell} - \exp\left(d\ell \log(n) \log(N) - \Omega(n)\right).$$

This concludes the proof of Theorem 1.5.

References

[AGM+26] Yaroslav Alekseev, Mika Göös, Konstantin Myasnikov, Artur Riazanov, and Dmitry Sokolov. Sampling permutations with cell probes is hard. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing (to appear)*, 2026. 1, 2

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

[BKMO26] Farzan Byramji, Daniel M Kane, Jackson Morris, and Anthony Ostuni. Hard-to-sample distributions from robust extractors. *arXiv preprint arXiv:2604.26179*, 2026. 1

[BWP26] Adam Bene Watts and Natalie Parham. Unconditional quantum advantage for sampling with shallow circuits. In *17th Innovations in Theoretical Computer Science Conference*, volume 362 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 17, 12. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2026. 1

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

[Czu15] Artur Czumaj. Random permutations using switching networks. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 703–712. ACM, New York, 2015. 2

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

[GKM+26] Daniel Grier, Daniel M. Kane, Jackson Morris, Anthony Ostuni, and Kewen Wu. Quantum advantage from sampling shallow circuits: beyond hardness of marginals. In *17th Innovations in Theoretical Computer Science Conference*, volume 362 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 73, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2026. 1

[GLSS15] Dmitry Gavinsky, Shachar Lovett, Michael Saks, and Srikanth Srinivasan. A tail bound for read-k families of functions. *Random Structures Algorithms*, 47(1):99–108, 2015. 5

[KOW24] Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locality bounds for sampling Hamming slices. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1279–1286, 2024. 1

[LV11] Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 243–251. IEEE, 2011. 1

[Vio12] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012. 1, 2

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

[Vio20] Emanuele Viola. Sampling lower bounds: boolean average-case and permutations. *SIAM Journal on Computing*, 49(1):119–137, 2020. 1, 2

[Vio23] Emanuele Viola. New sampling lower bounds via the separator. In *38th Computational Complexity Conference (CCC 2023)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2023. 1, 2

[YZ24] Huacheng Yu and Wei Zhan. Sampling, flowers and communication. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, pages 100–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024. 1, 2

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il