Electronic Colloquium on Computational Complexity, Report No. 164 (2025)

Constant-time source decoding

Jordan Horacsek
jordan_horacsek@sfu.ca
Simon Fraser University

Chin Ho Lee
chinho.lee@ncsu.edu
North Carolina State University

Igor Shinkar
ishinkar@sfu.ca
Simon Fraser University

Emanuele Viola*
mathematicsoftheimpossible@gmail.com
Northeastern University

Renfei Zhou
renfeiz@andrew.cmu.edu
CMU

November 2, 2025

Abstract

We give versions of Shannon's coding theorem where the decoder runs in constant time:

1. Let $D = (D_1, D_2, \ldots, D_n)$ be a product distribution where the D_i have constant support and have dyadic probability masses (i.e., of the form $a/2^b$ where a, b are integers). Then D can be sampled in constant time in the bit-probe model (equivalently, in NC^0) and randomness complexity $(h(D) + \epsilon)n$, up to an exponentially small statistical error. The dyadic requirement is necessary.
2. Every p-biased distribution can be sampled in constant time in the cell-probe model with randomness complexity $h(p)n + \sqrt{n} \cdot \mathrm{polylog}(n)$, up to a polynomially small statistical distance.
3. We determine the tradeoffs between locality and statistical distance for sampling the $1/4$-biased distribution using non-trivial randomness complexity (e.g., $1.99n$). For 2 bit probes, essentially no non-trivial approximation is possible; for 3 bit probes, we give a sampler with $1/\mathrm{poly}(n)$ statistical distance and show that this is best possible; finally, 4 bit probes suffice for exponentially small distance.

Our constructions use various tools from low-density parity-check codes, and recent results on succinct and retrieval data structures [HLYZZ, STOC 2025].

*Supported by NSF grant CCF-2430026.

ISSN 1433-8092

1 Introduction

Shannon's source coding theorem, see for example [CT06, Theorem 3.2.1], says that n i.i.d. samples from a source D can be compressed into about $nH(D)$ bits from which the samples can be decoded with high probability. In Shannon's result the decoder is not explicit. A vast literature in information and computer science theory has developed codes with various guarantees. However, approaching $nH(D)$ bits with decoders that run in constant time per output symbol has proved elusive. Surprisingly to us, in this work we achieve that in various settings. For example, in Theorem 1 below we achieve it in the bit-probe (a.k.a. NC^0) model, for any *dyadic* source, a requirement which is necessary by previous work. Our results give the stronger guarantee that the output of the decoder (over uniform input) is close to the distribution of the samples from the source D. (This is known as *source simulation* or *resolvability* in the information-theory literature.) Indeed, a main motivation for this paper comes from the study of the *complexity of distributions*. We now discuss this perspective in greater detail and then present our results.

Recent years have witnessed substantial work and progress on the complexity of sampling distributions. Let us make the setting precise. Given a distribution $\mathcal{D}$, say supported on $\{0,1\}^n$, the goal is to design a mapping $f\colon \{0,1\}^m \to \{0,1\}^n$ such that for a uniformly random $x \in \{0,1\}^m$ the distribution $f(x)$ is equal (or close) to $\mathcal{D}$. Note that as opposed to the standard setting of computing a function, in this setting we do not require that f outputs any specific value on a particular input x, instead only considering the distribution $f(x)$ for a uniformly random $x \in \{0,1\}^n$. This is motivated by earlier works which showed that *sampling* can be easier than *computing*.

For a concrete example, consider the parity function. While classical results in the 80s [Has86, Smo87] showed that $\mathbf{AC}^0$ circuits have small correlation with the parity function, the works [Bab87, BL87] showed that the uniform distribution on n-bit strings with the same parity can be sampled exactly by the 2-local function

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

Other surprising examples include sampling the inner-product mod 2 function [IN96] and random permutations [MV91, Hag91, Vio12]. For background and more discussion we refer the readers to [Vio12].

The work [Vio12] initiated a study of the complexity of sampling, with a focus on restricted computational models and lower bounds. Since then, a large body of works have established many exciting unconditional results on sampling distributions in several restricted models, including local functions [Vio12, Vio23, FLRS23, KOW24b, KOW24a], small-depth circuits [LV12, BIL12], one-way space-bounded computation [CGZ22], and communication protocols [GW20, YZ24]. By now, this line of research has found a wide range of applications in various areas such as randomness extractors [Vio14a, CZ19, CS16], data structures [Vio12, Vio23, YZ24], low-distortion embeddings [BCS16, BS23], quantum and classical separation [WP23], and coding theory [SS24]. In fact, jumping ahead, this work will also further develop some of these connections (in particular, to data structures). We refer the readers to the blog post [Vio24] for more details on these connections.

In this work, we study the complexity of sampling *product distributions*. The special case of *p-biased distributions* on n bits, denoted $\mathrm{Ber}(p)^n$ and where the bits are i.i.d. with probability of 1

equal to p is already omnipresent in computer science. For example, the complexity of sampling p-biased distributions has been studied in a series of recent works including [Vio23, FLRS23, KOW24b, KOW24a]. Some of the motivation for this line of research comes from a connection with data structures from [Vio12], discussed more below. Such distributions also arise as *noise* or *random restrictions* in various areas ranging from distributed computing, to Boolean function analysis, coding theory, randomized algorithms, and learning theory. For example, in cryptography the Learning Parity with Noise (LPN) problem [BKW03] or its cousin the Learning With Errors (LWE) problem [Reg05] are considered standard hardness assumptions. Instantiation of cryptographic primitives based on these assumptions typically requires perturbing a binary vector with p-biased noise. Hence very efficient (or parallel) implementations typically require correspondingly efficient ways to sample such noise. In pseudorandomness, recent approaches to constructing generators involve summing bounded-independence generators with p-biased distributions, see the monograph [HH23] and the works [DILV24a, DILV24b]. Again, efficient implementations of such generators require efficient samplers for p-biased distributions.

Our main interest in this work is to understand the tradeoffs between *locality*, *input length* which we also call *seed length* or *randomness complexity*, and *statistical distance* for sampling product distributions. To illustrate, let us consider the task of sampling the $1/4$-biased distribution on n bits, denoted $\mathrm{Ber}(1/4)^n$. On the one hand, the distribution $\mathrm{Ber}(1/4)^n$ can be sampled with randomness complexity $2n$ and locality 2. This trivial construction partitions the $2n$ input bits into n pairs, and for each pair computes AND of the two bits. On the other hand, by Shannon's source coding theorem, any p-biased distributions on n bits can be sample with randomness complexity $h(p)n$, where $h(p) := p\log_2(1/p) + (1-p)\log_2(1/(1-p))$ is the binary entropy function, which is best possible. However, this coding theorem does not take into account resources such as locality required in the sampling procedure.

It is natural to ask if one can simultaneously achieve small locality and randomness complexity. Indeed, this question was explicitly posed in a blog post [Vio14b] about 10 years ago, yet to the best of our knowledge, the question of understanding these tradeoffs has remained largely open. In particular, the following basic question has remained open:

> Can you sample $\mathrm{Ber}(1/4)^n$ with constant locality and randomness complexity $(h(1/4)+\epsilon)n$?
> Can you even get randomness complexity $1.99n$ with constant locality?

1.1 Our results

We resolve the aforementioned basic question in the affirmative. Surprisingly to us, we show that with constant locality we can sample *any* product distribution (in particular, $\mathrm{Ber}(1/4)^n$) with nearly optimal randomness complexity. For this result we need the distribution to be *dyadic*, i.e., the probability masses have to be of the form $a/2^b$ where a, b are integers. This dyadic requirement is necessary, since $\mathrm{Ber}(1/3)^n$ cannot be sampled locally, not even close. This follows by techniques in [Vio23], though the result there is stated for the Hamming slice; alternatively see [KOW24b, Theorem 1.10]. Thus, we illustrate a stark contrast between sampling $\mathrm{Ber}(p)^n$ for dyadic and non-dyadic p. We denote statistical distance by dist.

Theorem 1 (Special case of Theorem 7)**.** *Let $D = (D_1, D_2, \ldots, D_n)$ be a product distribution where each D_i is dyadic and supported on $\leq s$ points. For every $\epsilon > 0$, there is a $O_{s,\epsilon}(1)$-local $f\colon \{0,1\}^{H(D)+\epsilon n}$ such that $\operatorname{dist}(f(U), D) \leq e^{-\Omega_{s,\epsilon}(n)}$.*

Increasing the locality to $O(\log n)$ we can approximate any distribution by a dyadic one and sample any *arbitrary* product distribution to within statistical distance $1/\operatorname{poly}(n)$ (see Corollary 8).

The above result is in the bit-probe model. The next result is in the *cell*-probe model: The input randomness is organized in words of $\log n$ bits, and one probe reads an entire word. We show how to sample $\mathrm{Ber}(p)^n$ with randomness complexity $h(p)n + \tilde{O}(\sqrt{n})$ to within distance $1/\operatorname{poly}(n)$, in constant time.

Theorem 2. *The distribution $\mathrm{Ber}(p)^n$ can be sampled using $h(p)n + \sqrt{n} \cdot \operatorname{polylog}(n)$ uniform bits within statistical distance $1/\operatorname{poly}(n)$ with $O(1)$ word-probes.*

Returning to the bit-probe model, recall the trivial sampler of $\mathrm{Ber}(1/4)^n$ that is 2-local and uses randomness complexity $2n$. We ask ourselves what can be achieved using constant locality and non-trivial randomness complexity $(2-\epsilon)n$. We determine the tradeoff between locality and statistical distance: For 2 bit probes, no non-trivial approximation is possible; for 3 bit probes, we give a sampler with $1/\operatorname{poly}(n)$ error and show that this is best possible; finally, 4 bit probes suffice for exponentially small distance. We state these results in two theorems, the first focusing on negative results, the other on positive.

Theorem 3 (Theorem 21 and Theorem 29)**.** *For $\epsilon > 0$ and $f\colon \{0,1\}^{(2-\epsilon)n} \to \{0,1\}^n$ be any d-local function. We have*

$$\operatorname{dist}(f(U), \mathrm{Ber}(1/4)^n) \geq \begin{cases} 1 - e^{-\Omega(n)} & \text{if } d = 2 \\ n^{-O(1)} & \text{if } d = 3. \end{cases}$$

Theorem 4 (Theorem 25 and Theorem 32)**.** *For $d \in \{3,4\}$, there is an $\epsilon > 0$ and a d-local sampler $f\colon \{0,1\}^{(2-\epsilon)n} \to \{0,1\}^n$ such that*

$$\operatorname{dist}(f(U), \mathrm{Ber}(1/4)^n) \leq \begin{cases} n^{-\Omega(1)} & \text{if } d = 3 \\ e^{-\Omega(n)} & \text{if } d = 4. \end{cases}$$

Our constructions are explicit in the following sense. The claimed samplers (viewed, for example, as circuits) can be constructed by an efficient randomized algorithm, with a small error probability. Jumping ahead, the error probability arises from the need of constructing certain matrices (cf. Lemma 10) for which we do not know of a deterministic construction. However, at least in the cell-probe model we also obtain a *deterministic* construction of the sampler (Appendix D).

While we have focused on product distributions, we mention that a body of works has established strong negative results for sampling distributions in NC^0 or even AC^0 *regardless of the input length of the sampler*. For example, [LV12] has shown the existence of *linear maps* that cannot be sampled in AC^0. Still, there remains some interesting open questions. For example, it would be interesting to sample random walks on graphs (equivalently, Markov chains), a problem studied in [VWY20].

1.2 Proof overview

We now give an overview of the proofs. We focus on sampling $\mathrm{Ber}(1/4)^n$, which captures all the key ideas in our arguments.

Overview of the proof of Theorem 1. A building block for the proof is the construction of a (possibly inefficient) sampler of $\mathrm{Ber}(1/4)^b$ with *expected* randomness complexity close to the optimal $h(1/4)b$. We call this the *block-sampler*. To illustrate the basic idea, consider sampling one bit, i.e., $\mathrm{Ber}(1/4)^1$. We can do so as follows. First, read an input bit. If it's 0, output 0; otherwise, read another input bit and output it. This samples perfectly $\mathrm{Ber}(1/4)^1$. While in the worst case we use a trivial randomness complexity 2, the expected number of input bits read is only 1.25, which is much better. This idea can be suitably generalized, and can approach the optimum when taking the block-size b large enough.

Given such a block sampler, we divide the n output bits in blocks of length b, and consider sampling each block with an independent copy of the block sampler. By concentration bounds, we know that with high probability over the randomness of the input bits, the actually number of random bits used to sample a typical output is close to optimal.

Now in some sense we *derandomize* this construction. We in turn sample the inputs to the block samplers pseudorandomly, via a local linear transformation. Specifically, we take a nearly optimal number of bits, and we multiply them by a sparse matrix that expands these to the larger number of bits which can be needed by the block samplers in the worst case. The key property we need from the matrix is that *most* small subsets of the rows of the matrix (that correspond to the coordinates read by the block samplers) are linearly independent. We remark that this is a *weaker* notion than bounded uniformity, which demands that *every* small subset of rows of the matrix be linearly independent. In fact, the Plotkin bound shows that no matrix satisfying the latter requirement would achieve optimal seed length. So utilizing the weaker condition is crucial in our construction. The matrix property that we need seems relatively basic, yet we cannot find a result in the literature that we can use directly, so we give a self-contained analysis that a suitable random construction works.

We obtain Theorem 1 by dividing the n bits into blocks of constant size.

Overview of the proof of Theorem 2. Theorem 2 is obtained via a new connection between sampling and succinct data structure. While a link between these two areas was already observed in [Vio12] (see Claim 6) and used in a number of following works, our connection is different. [Vio12] pointed out that a succinct data structure is immediately a non-trivial sampler, but the statistical distance can be quite large and close to 1. This connection can be used to establish data-structure lower bounds from sampling lower bounds that rule out even such large statistical distance, but it is not clear how one can use it to obtain useful samplers, even with statistical distance $1/2$. Indeed, we are not aware of any construction of samplers that is based on data structure. Moreover, as our target distribution is not uniform on a set, it is not clear we can use any existing data structure directly in a blackbox way. Instead, we leverage and adapt the *techniques* used in recent exciting progress on the *set membership* (and *dictionary*) data structure problems [HLY+24], in particular the use of *retrieval data structures*.

To explain we begin with a key concept, originating in [Păt08] (cf [DPT10]).

Definition 5 (Spillover representation)**.** *Given an injective map from a set S to $\{0,1\}^M \times [K]$, the* spillover representation *of an element in S is its corresponding element $(m,k) \in \{0,1\}^M \times [K]$, where k is called the* spill*.*

The work [Vio12] observed the following connection between sampling the uniform distribution over a set and membership data structure.

Claim 6. *Suppose a set of n keys in a universe U can be represented by a spillover representation $(m,k) \in \{0,1\}^M \times [K]$ with $M + \log_2 K \le \log_2 \binom{U}{n} + \epsilon$. Then a uniform key can be sampled from $\{0,1\}^M \times [K]$ with error ϵ.*

Proof. The error is at most the probability that a uniform element from $\{0,1\}^M \times [K]$ is not a spillover representation of any keys. Using $1 - 1/x \le \log_2 x$ for $x > 0$, this probability is

$$1 - \frac{\binom{U}{n}}{2^M \cdot K} \le \log\left(\frac{2^M \cdot K}{\binom{U}{n}}\right) \le \epsilon. \qquad \square$$

We again divide the n bits into blocks of $B = \mathrm{polylog}(n)$ bits. To sample a block with constant word-probes, we will now use a succinct membership data structure by Yu [Yu22]. It shows that one can represent B-bit strings *of Hamming weight s* by spillover representations in $\{0,1\}^M \times [K]$ so that each string can be retrieved using $O(1)$ word-probes to the representation. Moreover, the redundancy $M + \log_2 K - \log \binom{B}{s}$ is polynomially small.

A critical point here is that s is not fixed, but a random variable. Consequently, even M and K are random variables.

To sample $\mathrm{Ber}(1/4)^B$, as in [HLY+24] we encode the weight distribution $\mathrm{Bin}(B, 1/4)$ into the first $O(1)$ words in each representation with a $1/\mathrm{poly}(n)$ increase in redundancy. This gives us a block sampler for $\mathrm{Ber}(1/4)^B$: we first sample (the first) $O(1)$ words to determine the Hamming weight $\boldsymbol{s}$ distributed according to $\mathrm{Bin}(B, 1/4)$, and then sample a uniform spill representation in $\{0,1\}^{M^{(\boldsymbol{s})}} \times [K^{(\boldsymbol{s})}]$. One can show that $M^{(\boldsymbol{s})} + \log_2 K^{(\boldsymbol{s})} \le h(1/4)B + 1/\mathrm{poly}(n)$ *in expectation*.

Now we can apply Claim 6 to obtain a $O(1)$-word-probe sampler for $\mathrm{Ber}(1/4)^B$.

Our plan is to use $L := n/B$ independent copies of the sampler to sample the L blocks. However, as the sizes of the representations depend on $\boldsymbol{s}$, sampling the L representations $(\boldsymbol{m}_i, \boldsymbol{k}_i) \sim \{0,1\}^{M^{(\boldsymbol{s}_i)}} \times [K^{(\boldsymbol{s}_i)}]$ together with small redundancy becomes a challenge. The issue here is what we alluded to before. The $\boldsymbol{m}_i$ and $\boldsymbol{k}_i$ are random variables, so we need to put together data structures of varying length which is not obvious: where are the relevant input bits for a specific output bit?

The work [HLY+24] addressed this challenge using *augmented retrieval* data structure. We will not define it here, but the key observation behind their construction is that the random variable $M^{\boldsymbol{s}}$ typically is *at least* $M_{\mathrm{fixed}} = \Omega(\log \binom{B}{pB - B^{2/3}})$, which is much larger than its deviation $O(B^{2/3} \log B)$, and $\log_2 K^{(\boldsymbol{s})} = O(\log n)$.

Based on this observation, [HLY+24] constructs random sparse matrices to concatenate the L representations with $\mathrm{polylog}(n)$ redundancy. Here, we use the same random sparse matrices to sample the L spillover representations. However, our construction does not achieve $\mathrm{polylog}(n)$

redundancy as in [HLY+24], because unlike in the data structure setting, we cannot write down the sizes of each retrieval data structures and point to their starting positions in the memory. So we instead put an upper bound on the sizes of all retrieval data structures except the last one.

Also, the reconstruction in [HLY+24] requires switching between spillover representations over symbols with different alphabet sizes with small redundancy. In the sampling setting, we also have to ensure these transformations also maintain closeness to the uniform distribution (see Lemma 14).

Overview of Theorem 3 and Theorem 4. Our 2-local lower bound is based on a win-win argument. Given a sampler $f\colon \{0,1\}^m \to \{0,1\}^n$ where $m = (2-\epsilon)n$. We consider the bipartite graph representing the input-output dependency of f.

Suppose there is a subset of m' inputs which connects to $n' := 100m'$ neighbors, then for every fixing of these input bits, f restricted to the n' bits is a 1-local, which can be shown to be exponentially far from $\mathrm{Ber}(1/4)^{n'}$, and this remains so after summing over all $2^{m'}$ fixings of the inputs.

Therefore, if $m' = \Omega(n)$, then the result follows. Otherwise, by removing these input vertices and their neighbors, we are left with a 2-local map from $(2-\Omega(\epsilon))n''$ bits to $n'' = \Omega(n)$ bits where *every input has bounded degree*. So we can decompose the outputs into $\Omega(n)$ groups so that each group depends on disjoint inputs. We show that each group has some constant distance away from the $1/4$-biased distribution. So the overall distance is at least $1 - e^{-\Omega(n)}$.

Our 3-local lower bound (Theorem 29) is shown by finding a set of output coordinates of size $k = O(\log(n))$ which depend on at most $2k-1$ inputs. Indeed, by granularity it follows that we see all zeros on these k coordinates with probability either 0 or at least $2^{-(2k-1)} = \frac{2}{4^k}$, while $\mathrm{Ber}(1/4)^k$ outputs all zeros with probability $\frac{1}{4^k}$. Therefore, the statistical distance of our sampler to $\mathrm{Ber}(1/4)^n$ is at least $\frac{1}{4^k}$. In order to find such a set, we consider the bipartite graph representing the input-output dependency of the sampler. Noting that the degree of each output vertex is at most 3, the problem essentially reduces to finding a cycle of length $O(\log n)$ in any graphs whose average degree is bounded above by 2.

The construction of our 3-local and 4-local samplers (Theorem 4) is inspired by the recent iterative framework in constructing pseudorandom generators [HH23]. Recall that the output of the trivial 2-local sampler is the bitwise AND $x \wedge y$ for two independent uniform n-bit strings x and y. We can think of x as picking a uniformly random subset of the n positions. Then to get close to $\mathrm{Ber}(1/4)^n$, we only require y to be uniform on the subset of positions chosen by x *with high probability*.

To generate such y, we assign each y_i to two input bits z_i, z_i' according to a 3-regular expander graph G, where y_i corresponds to the edge (z_i, z_i'). Then we let y_i to be $z_i \oplus z_i'$. To analyze the construction, we show that a random subgraph of G has no cycle with probability $1 - 1/\operatorname{poly}(n)$. That means the y_i's are uniform when restricted to most subsets chosen by x, and the result follows.

Our 4-local sampler construction follows the same idea. Again, we use n random bits to select a random subset of $[n]$. Then to sample $y \in \{0,1\}^n$ we use a 3-local LPDC code instead of an expander. By analyzing the weight distribution of the code, we show that a random subset of rows in the corresponding parity-check matrix is full rank with probability $1 - e^{-\Omega(n)}$.

2 Local sampler for product distributions

In this section, we prove Theorem 1.

Theorem 7. *Let q be an integer, and $D_1, \ldots, D_n$ be n distributions on $\{0,1\}^w$, where $D_i(s)$ is an integer multiple of 2^{-q} for every $i \in [n]$ and $s \in \{0,1\}^w$. Let $D = D_1 \times \cdots \times D_n$ the product distribution of the D_i's.*

For every $\epsilon > 0$, and let $m = H(D) + \epsilon n$. There exists a sampler $f\colon \{0,1\}^m \to (\{0,1\}^w)^n$ with locality $O(\frac{q}{\epsilon}\log(\frac{1}{\epsilon}))$ such that $\operatorname{dist}(f(U_m), D) < 2^{-\Omega\left(\frac{\epsilon^3 n}{q^2}\right)}$. The sampler f is adaptive, in the sense that for each output query, f makes $O(\frac{q}{\epsilon}\log(\frac{1}{\epsilon}))$ sequential queries to the inputs, where each query may depend on the previous queries.

The following is an almost immediate corollary from Theorem 7.

Corollary 8. *Let $D_1, \ldots, D_n$ be n distributions on $\{0,1\}^w$ for $w = O(\log n)$. For every $\epsilon \in (0, \frac{\log n}{n^{1/3}})$, the product distribution $D = D_1 \times \cdots \times D_n$ can be sampled using $H(D) + \epsilon n$ bits with locality $O(\frac{\log(1/\epsilon)}{\epsilon} \cdot \log(n))$ and error $1/\operatorname{poly}(n)$.*

Proof. We can approximate each D_i with a distribution D_i' whose probability masses are integer multiples of $2^{-(w+\lceil \log_2(n/\gamma)\rceil)}$ such that $|D_i(s) - D_i'(s)| \le \frac{2\gamma}{n\cdot 2^w}$ for all $s \in \{0,1\}^w$, and in particular $\operatorname{dist}(D_i, D_i') \le \gamma/n$ for all $i \in [n]$ (cf. [Vio12, Lemma 5.2]). Setting $\gamma = 1/n^C$ for a sufficiently large constant C, the two distributions D and $D' := D_1' \times \cdots \times D_n'$ are $1/\operatorname{poly}(n)$-close in total variation distance. Note that for each $i \in [n]$ we have

$$H(D_i') - H(D_i) \le \sum_{s\in\{0,1\}^w} \left| D_i'(s) \cdot \log_2(\frac{1}{D_i'(s)}) - D_i(s) \cdot \log_2(\frac{1}{D_i(s)}) \right|.$$

Since $|D_i(s) - D_i'(s)| \le \frac{2\gamma}{n\cdot 2^w}$, each term in the sum is at most $\frac{2\gamma}{n\cdot 2^w}\log_2\left(\frac{n\cdot 2^w}{2\gamma}\right)$, and hence

$$H(D_i') \le H(D_i) + 2^w \times \frac{2\gamma}{n\cdot 2^w}\log_2\left(\frac{n\cdot 2^w}{2\gamma}\right) = \frac{2\gamma}{n}\cdot\log_2\left(\frac{n\cdot 2^w}{2\gamma}\right).$$

Therefore, $H(D') \le H(D) + 1/\operatorname{poly}(n)$, and the corollary follows by applying Theorem 7 on D'. □

We now turn to the proof of Theorem 7.

Lemma 9. *Let D be any distribution on $\{0,1\}^w$, where there is some $q \in \mathbb{N}$ such that $D(s)$ is an integer multiple of 2^{-q} for all $s \in \{0,1\}^w$. Then D can be sampled by a decision tree with q input bits, where the expected depth of a leaf is at most $H(D) + 3$.*

Proof. We start with a complete binary tree of height q and 2^q leaves. For each $s \in \{0,1\}^w$, we label $D(s)\cdot 2^q$ consecutive leaves in the tree by s, each leaf having weight 1. To obtain our decision tree, if two siblings have the same label s, we remove them, label their parent by s, and assign its weight to be the sum of the weight of the two removed vertices. We repeat this process until no 2

siblings share the same label. To sample D, we use the q input bits one by one to traverse the tree until we reach a leaf, in which case we output its label.

For a node v in the tree, denote by $d(v)$ is the depth of v, i.e., the distance of v from the root. For a label s, denote by $d(s)$ the minimal depth of a leaf in the tree, whose label is s. Note that the probability of sampling a leaf v is $2^{-d(v)}$, and the probability of sampling a node labeled s is equal to $\sum_{v \in V(s)} 2^{-d(v)}$, $V(s)$ denote the set of all leaves v labeled with s.

We now bound from above the expected depth of a leaf. Note that at each level of the tree there can be at most two nodes that share the same label s, for otherwise there must be two siblings who share the same label. Hence, in order to maximize the depth, we put two nodes labeled s on each level between $d(s)$ and q. Hence, we have

$$D(s) \leq \sum_{j=0}^{q-d(s)} 2 \cdot \frac{1}{2^{q-j}} \leq 2^{-d(s)+2}.$$

Hence, the minimal depth of a nodes labeled s satisfies $d(s) \leq \log_2(1/D(s)) + 2$. Thus, the expected depth of a leaf in the decision tree is at most

$$\begin{aligned}
\sum_{s \in \{0,1\}^w} \sum_{v \in V(s)} 2^{-d(v)} \cdot d(v) &\leq \sum_{s \in \{0,1\}^w} \sum_{j=0}^{q-d(s)} 2 \cdot \frac{2^j}{2^q} \cdot (q-j) \\
&= 2^{-q+1} \sum_{s \in \{0,1\}^w} \sum_{j=0}^{q-d(s)} 2^j \cdot q - 2^{-q+1} \sum_{s \in \{0,1\}^w} \sum_{j=0}^{q-d(s)} 2^j \cdot j \\
&= 2^{-q+1} \cdot q \sum_{s \in \{0,1\}^w} (2^{q-d(s)+1} - 1) - 2^{-q+1} \sum_{s \in \{0,1\}^w} (q - d(s) - 1) \cdot 2^{q-d(s)+1} + 2 \\
&< q \sum_{s \in \{0,1\}^w} 2^{-d(s)+2} - \sum_{s \in \{0,1\}^w} (q - d(s) - 1) \cdot 2^{-d(s)+2} \\
&\leq \sum_s 2^{-(d(s)} \cdot (d(s)+1) \\
&\leq \sum_s D(s) \cdot \big(\log_2(1/D(s)) + 3\big) \\
&= H(D) + 3.
\end{aligned}$$

So the expected depth of a leaf in the decision tree is at most $H(D) + 3$. □

For the proof of Theorem 7 we also need a sparse matrix with the following properties.

Lemma 10. *Fix $k \in \mathbb{N}$ and a sufficiently small $\alpha > 0$. Let $m \geq \max\{(1 + 2h(\alpha))k, \alpha n\}$, and $D = \lceil \ln(1/\alpha)/\alpha \rceil$. Let $\mathcal{S}$ be any distribution supported on subsets $S \subseteq [n]$ of size $|S| = k$. There exists a matrix $M \in \mathbb{F}_2^{n \times m}$ with at most D ones in each row, such that if we sample a subset $\boldsymbol{S}$ according to $\mathcal{S}$, then the corresponding submatrix $M_{\boldsymbol{S}} \in \mathbb{F}_2^{k \times m}$ is full rank with probability at least*

$$\Pr_{\boldsymbol{S} \sim \mathcal{S}} [M_{\boldsymbol{S}} \text{ is full rank}] \geq 1 - 2^{-\Omega_\alpha(n)}.$$

Remark 11. In Lemma 10 we are looking for a matrix M such that *most subsets* of k rows are linearly independent, where most is with respect to the distribution $\boldsymbol{S}$. We note that with the required parameters we cannot possibly hope for a matrix where *any* k rows are linearly independent, as such matrix would correspond to a party check matrix of a linear error correcting codes with block length n, distance k, and dimension at least $n-m$, which is impossible over small alphabet (e.g., by Plotkin bound, stating that a linear code of length m with distance k has dimension at most $m-2k+o(1)$).

The proof of Lemma 10 is below in Section 2.1.

Proof of Theorem 7. Divide the D_i's into n/t blocks each of size $t=\lceil 12/\epsilon\rceil$. Let D^j be the product of the D_i's in the j-th block. We first use Lemma 9 to sample each D^j independently using qt bits. Let $f\colon \{0,1\}^{qn}\to\{0,1\}^{wn}$ be our sampler (which is simply a concatenation of the samplers from Lemma 9). Then, we sample these qn bits pseudorandomly by applying the sparse matrix from Lemma 10 to a seed of length $m=H(D)+\epsilon n$.

For each $z\in\{0,1\}^{qn}$, let $S_z\subseteq[qn]$ denote the subset of positions read (adaptively) by f to evaluate $f(z)$. We emphasize that S_z are the only positions read by f to evaluate $f(z)$. Let $\boldsymbol{S}$ be the distribution of S_z induced by choosing a uniformly random $z\in\{0,1\}^m$. Note that S_z can be written as $S_z=S^1\cup\ldots S^{n/t}$ with $S^j\subseteq\{(j-1)t+1,(j-1)t+2,\ldots,jt\}$, where S^j's are distributed independently, and $\mathbb{E}[|S_z^j|]\le H(D^j)+3$. Therefore,

$$\mathbb{E}_z[|S_z|]\le\sum_{i=1}^{n/t}H(D^j)+3\cdot(t/n)\le H(D)+(\epsilon/4)n.$$

Let $k=H(D)+(\epsilon/3)n$. Note that $\mathbb{E}_z[|S_z|]+\epsilon n/12\le k\le(1-\epsilon/2)m$ for all $\epsilon<1/3$.

Let M be the $qn\times m$ matrix obtained by applying Lemma 10 with the distribution $\boldsymbol{S}$. Our sampler will take the input $x\in\{0,1\}^{m=H(D)+\epsilon n}$ and output $f(Mx)$.

Clearly the input length of f is $m=H(D)+\epsilon n$. The locality of the sampler is at most $qt\cdot D=O(\frac{q}{\epsilon}\log\left(\frac{1}{\epsilon}\right))$, where D is the bound on the sparsity of M from Lemma 10.

Below we show that

$$\operatorname{dist}(f(U_m),\mathsf{Ber}(1/4)^n)\le\Pr[|S_z|\ge k]+2^{-\Omega(\epsilon n)}.$$

By Hoeffding's inequality, we have

$$\Pr[|S_z|\le k]\ge\Pr\left[|S_z|\le\mathbb{E}_z[|S_z|+\epsilon n/12]\right]\ge 1-e^{-\Omega\left(\frac{(\epsilon n)^2}{(n/t)\cdot(qt)^2}\right)}\ge 1-e^{-\Omega(\frac{\epsilon^3}{q^2}n)}.$$

For a uniformly random $z\in\{0,1\}^m$

$$\Pr_z[M_{S_z}\text{ is full rank}\,||S_z|\le k]\ge 1-2^{-\Omega(\epsilon n)}.$$

Let us condition on the event that $|S_z|\le k$ and M_{S_z} is full rank. Then for a uniformly random input $x\in\{0,1\}^m$ to our sampler, the $\le k$ bits in $(Mx)_{S_z}$ are uniformly random, and thus in each of the (n/t) blocks the output is distributed according to D^j. Therefore,

$$\operatorname{dist}(f(U_m),\mathsf{Ber}(1/4)^n)\le\Pr[|S_z|\ge k]+2^{-\Omega(\epsilon n)}\le e^{-\Omega(\frac{\epsilon^3}{q^2}n)},$$

as required. □

2.1 Proof of Lemma 10

We prove the lemma by considering a random $n \times m$ matrix $\boldsymbol{M}$, where each row of $\boldsymbol{M}$ is sampled independently according to the following distribution: Select D indices $\boldsymbol{i}_1, \boldsymbol{i}_2, \ldots, \boldsymbol{i}_D \in [m]$ uniformly and independently, and define the row of $\boldsymbol{M}$ to be $e_{\boldsymbol{i}_1} + e_{\boldsymbol{i}_2} + \cdots + e_{\boldsymbol{i}_D}$. Clearly, each row of $\boldsymbol{M}$ has at most D ones.

For an integer $1 \le \ell \le k$, let $p_\ell := \Pr[z\boldsymbol{M} = 0]$, where $z \in \mathbb{F}_2^n$ is a vector of Hamming weight ℓ. We start with the following expression for the probability that z is a null vector of $\boldsymbol{M}$.

Claim 12. *For all $1 \le \ell \le m$ we have*

$$p_\ell = \frac{1}{2^m} \sum_{S \subseteq [m]} \left(1 - \frac{2|S|}{m}\right)^{D\ell} = \frac{1}{2^m} \sum_{i=0}^{m} \binom{m}{i} \left(1 - \frac{2i}{m}\right)^{D\ell}.$$

In particular,

$$p_\ell \le \begin{cases} 2\left(\frac{2D\ell}{m}\log\left(\frac{m}{D\ell}\right)\right)^{\frac{D\ell}{2}} & \text{if } 1 \le \ell \le \frac{m}{4D} \\ 2 \cdot 2^{-\frac{m}{4}} & \text{if } \frac{m}{4D} \le \ell \le \alpha m \\ m \cdot 2^{-(1-h(\alpha))m} & \text{if } \alpha m \le \ell. \end{cases}$$

Let us see how the claim above proves Lemma 10.

Proof of Lemma 10. Consider the random matrix $\boldsymbol{M}$ with at most D ones in each row as described above. Denote by L the event that any αm rows of $\boldsymbol{M}$ are linearly independent. Then, using the assumption about α being sufficiently small, Claim 12 implies that

$$\begin{aligned} \Pr[L] &\le \sum_{\ell=1}^{\alpha m} \binom{n}{\ell} p_\ell \\ &\le \sum_{\ell=1}^{\frac{m}{4D}} \binom{n}{\ell} 2\left(\frac{2D\ell}{m}\log\left(\frac{m}{D\ell}\right)\right)^{\frac{D\ell}{2}} + \sum_{\ell=\frac{m}{4D}+1}^{\alpha m} \binom{n}{\ell} 2 \cdot 2^{-\frac{m}{4}} \\ &\le 2\sum_{\ell=1}^{\frac{m}{4D}} \left(\frac{en}{\ell} \cdot \left(\frac{2D\ell}{m}\log\left(\frac{m}{D\ell}\right)\right)^{\frac{D}{2}}\right)^{\ell} + 2 \cdot 2^{h(\frac{\alpha m}{n})n} \cdot 2^{-\frac{m}{4}} \\ &\le n^{-\Omega(D)}. \end{aligned} \tag{1}$$

Fix a subset of the rows $S \subseteq [n]$ of size k. Then by Claim 12 and using $(1-\epsilon)(1+2\epsilon) \ge 1 + \epsilon/2$ for $\epsilon \in [0, 1/4]$.

$$\begin{aligned} &\Pr_{\boldsymbol{M}}\big[\text{there exists a subset of rows } T \subseteq S \text{ with } T \ge \alpha m \text{ whose sum is } 0\big] \\ &\quad \le \sum_{\ell=\alpha m}^{k} \binom{k}{\ell} p_\ell \le 2^k \cdot m \cdot 2^{-(1-h(\alpha))m} \le O(k) \cdot 2^k \cdot 2^{-(1+\frac{h(\alpha)}{2})k} \le 2^{-\Omega(h(\alpha)k)}. \end{aligned}$$

Note that if we consider the random matrix $\boldsymbol{M}$ *conditioned on* L, then for any $S \subseteq [n]$ of size k we have

$$\Pr_{\boldsymbol{M}}\big[\text{there exists a subset of rows } T \subseteq S \text{ with } T \geq \alpha m \text{ whose sum is } 0 \mid L\big] \leq \frac{2^{-\Omega(\alpha n)}}{\Pr[L]} \leq 2^{-\Omega(\alpha n)}.$$

Therefore, by the averaging argument, there exists a matrix M such that

$$\Pr_{S \sim \boldsymbol{S}}[M_S \text{ is full rank}] \geq 1 - 2^{-\Omega(\alpha n)}.$$

This completes the proof of Lemma 10. □

We now return to the proof of Claim 12.

Proof of Claim 12. Let $f\colon \{0,1\}^m \to \{0,1\}$ be the indicator function of the all zeros vector. We can write f in its Fourier expansion

$$f(x) = \prod_{i=1}^{m} \frac{1 + (-1)^{x_i}}{2} = \frac{1}{2^m} \sum_{S \subseteq [m]} (-1)^{\sum_{i \in S} x_i}.$$

Observe that for a uniform random index $\boldsymbol{i} \sim [m]$, we have $\mathbb{E}[(-1)^{\sum_{j \in S} (e_{\boldsymbol{i}})_j}] = 1 - \frac{2|S|}{m}$. As the D indices $\boldsymbol{i}_j$'s in each row of $\boldsymbol{M}$ are sampled independently, for a vector $z \in \mathbb{F}_2^n$ of Hamming weight ℓ, we have

$$\begin{aligned} p_\ell = \Pr[f(z\boldsymbol{M}) = 1] = 2^{-m} \cdot \sum_{S \subseteq [m]} \mathbb{E}\Big[(-1)^{\sum_{i \in S} (z\boldsymbol{M})_i}\Big] &= 2^{-m} \cdot \sum_{S \subseteq [m]} \left(1 - \frac{2|S|}{m}\right)^{D\ell} \\ &= 2^{-m} \cdot \sum_{i=0}^{m} \binom{m}{i} \left(1 - \frac{2i}{m}\right)^{D\ell}. \end{aligned} \tag{2}$$

Next, we prove the "in particular" part of the claim. We will consider 3 cases depending on the values of $1 \leq \ell \leq m$; in each case, we will decompose the sum in Eq. (2) into two parts according to some threshold t that depends on ℓ, and bound each part separately.

The case of $1 \leq \ell \leq \frac{m}{4D}$: Let $t = \frac{m}{2}(1 - \sqrt{h(D\ell/m)})$. Note that we have $2^{-m} \sum_{i=t+1}^{m} \binom{m}{i} (1 - \frac{2i}{m})^{D\ell} \leq (1 - \frac{2t}{m})^{D\ell}$, and so

$$\begin{aligned} p_\ell &\leq 2^{-m} \sum_{i=0}^{t} \binom{m}{i} \left(1 - \frac{2i}{m}\right) + \left(1 - \frac{2t}{m}\right)^{D\ell} \\ &\leq 2^{-(1-h(\frac{t}{m}))m} + h\left(\frac{D\ell}{m}\right)^{\frac{D\ell}{2}}. \end{aligned}$$

The first term can be upper bounded as follows. Using the fact that $h(1/2 - \sqrt{x}) < 1 - 2x$ with $x = h(D\ell/m)$, we have $h(t/m) = h(1/2 - \sqrt{h(D\ell/m)}/2) \leq 1 - \frac{h(D\ell/m)}{2}$. So,

$$2^{-(1-h(\frac{t}{m}))m} \leq 2^{-\frac{1}{2}h(\frac{D\ell}{m})m} \leq \frac{1}{\binom{m}{D\ell}^{1/2}} \leq \left(\frac{D\ell}{m}\right)^{\frac{D\ell}{2}}.$$

For the second term we use the fact that $h(x) \le 2x\log_2(1/x)$ for $x \in [0,1/2]$, which gives us

$$h\Big(\frac{D\ell}{m}\Big)^{\frac{D\ell}{2}} \le \Big(\frac{2D\ell}{m}\log\Big(\frac{m}{D\ell}\Big)\Big)^{\frac{D\ell}{2}}.$$

Therefore,

$$p_\ell \le \left(\frac{D\ell}{m}\right)^{\frac{D\ell}{2}} + \left(\frac{2D\ell}{m}\log\Big(\frac{m}{D\ell}\Big)\right)^{\frac{D\ell}{2}} \le 2\left(\frac{2D\ell}{m}\log\Big(\frac{m}{D\ell}\Big)\right)^{\frac{D\ell}{2}}.$$

The case of $\frac{m}{4D} \le \ell \le \alpha m$: Let $t = m/4$. Then

$$p_\ell \le 2^{-m}\sum_{i=0}^{t}\binom{m}{i}\left(1-\frac{2i}{m}\right) + \left(1-\frac{2t}{m}\right)^{D\ell} \le 2^{-(1-h(1/4))m} + 2^{-D\ell} < 2\cdot 2^{-\frac{m}{4}},$$

where the last inequality follows because $1 - h(1/4) \ge 1/4$ and $D\ell \ge m/4$ by our assumption.

The case of $\ell \ge \alpha m$: We first show that for every $0 \le i \le m$, it holds that

$$\binom{m}{i}\left(1-\frac{2i}{m}\right)^{D\ell} \le 2^{h(\alpha)m}.$$

When $0 \le i \le \alpha m$, this simply follows from $\binom{m}{i} \le 2^{h(\frac{i}{m})m} \le 2^{h(\alpha)m}$. Now, suppose $i \in [\alpha m, m/2]$. Using $\ell \ge \alpha m$ and our choice of $D \ge \ln(1/\alpha)/\alpha$, together with the fact that $h(x) \le 2x\log_2(1/x)$ for $x \in [0,1/2]$, we have

$$\left(1-\frac{2i}{m}\right)^{D\ell} \le e^{-\frac{2i}{m}D\ell} \le e^{-2i\ln(\frac{1}{\alpha})} \le e^{-2i\ln(\frac{m}{i})} \le 2^{-h(\frac{i}{m})m} \le \frac{1}{\binom{m}{i}}.$$

Thus, $\binom{m}{i}(1-\frac{2i}{m})^{D\ell} \le 1$. Finally, for $i \ge m/2$, note that $\binom{m}{i}(1-\frac{2i}{m})^{D\ell} \le \binom{m}{m-i}(1-\frac{2(m-i)}{m})^{D\ell}$ and so we can apply the previous bounds.

Therefore,

$$p_\ell = 2^{-m}\sum_{i=0}^{m}\binom{m}{j}\left(1-\frac{2i}{m}\right)^{D\ell} \le m\cdot 2^{-(1-h(\alpha))m}.$$

This completes the proof of Claim 12. □

3 Sampling p-biased distributions from static dictionary

In this section we prove Theorem 2.

Theorem 2. *The distribution* $\mathrm{Ber}(p)^n$ *can be sampled using* $h(p)n + \sqrt{n}\cdot\mathrm{polylog}(n)$ *uniform bits within statistical distance* $1/\mathrm{poly}(n)$ *with* $O(1)$ *word-probes.*

3.1 Changing bases

In this subsection, we prove local transformations between uniform distributions on sequences over different domains with little overheads and errors.

Claim 13. *The uniform distribution over $[K]$ can be sampled by m elements in $[q]$ with error K/q^m.*

Proof. We think of $[p]^m$ as $\{0, \dots, p^m - 1\}$. Given a uniform $\boldsymbol{u} \sim \{0, \dots, p^m - 1\}$, we output $\lfloor \frac{\boldsymbol{u}}{K} \rfloor$. The statistical distance is at most the probability that $\boldsymbol{u}$ lies in the last $p^m \bmod K$ elements, which is at most K/p^m. □

Lemma 14. *Given $p, q \le \mathrm{poly}(n)$, there is a function $f\colon [q]^m \to [p]^n$ such that*

- $m \le n \log_q p + O(\log_q n)$;
- *each output coordinate depends on $O(\log_q n)$ many input coordinates;*
- *for every subset $S \subseteq [n]$, if the coordinates $f(U)_S$ depends on are ϵ-close to uniform, then $f(U)_S$ is $(\epsilon + 1/\mathrm{poly}(n))$-close to uniform over $[p]^S$.*

Proof. We modify the proof in [DPT10, Section 4] as follows. They showed that one can represent $x_p \in [p]^n$ by a spillover representation $(x_q, y) \in [q]^{m'} \times [K]$ where $K = \mathrm{poly}(n)$ and

$$m' \log_2 q + \log_2 K \le n \log_2 p + \frac{1}{\mathrm{poly}(n)}.$$

Moreover, each element of $[p]^n$ only depends on $O(\log_q n)$ coordinates of $[q]^{m'} \times [K]$. It follow from Claim 6 that the uniform distribution on $[p]^n$ can be sampled from the uniform distribution on $[q]^{m'} \times [K]$ with error $1/\mathrm{poly}(n)$, with each output coordinate depending on at most $O(\log_q n)$ of the input coordinates. Finally, we use Claim 13 to sample the uniform distribution over $[K]$ using $O(\log_q n)$ elements of $[q]$ with error $1/\mathrm{poly}(n)$. □

3.2 Sampling p-biased distributions on polylog bits

In this subsection, we show how to sample $\mathrm{polylog}(n)$ many p-biased bits with $O(1)$-word probes.

Theorem 15. *Let $B = \mathrm{polylog}(n)$ and $C > 0$ be any constant. The distribution $\mathrm{Ber}(p)^B$ can be sampled adaptively from $\{0,1\}^{\boldsymbol{M}} \times [\boldsymbol{K}]$ with error $1/\mathrm{poly}(n)$, where $\boldsymbol{K} \le \mathrm{poly}(n)$, with the following properties:*

- *A random variable $\boldsymbol{s} \sim [pB - B^{2/3}, pB + B^{2/3}]$ that is $n^{-\Omega(C)}$-close to $\mathrm{Bin}(B, p)$ can be sampled using the first $t := C \log_2 n$ bits of $\{0,1\}^{\boldsymbol{M}}$.*
- *Given $\boldsymbol{s} = s$, the lengths $\boldsymbol{M} = M^{(\boldsymbol{s})}$ and $\boldsymbol{K} = K^{(\boldsymbol{s})}$ are fixed and*

$$\mathbb{E}_{\boldsymbol{s}}\left[M^{(\boldsymbol{s})} + \log_2 K^{(\boldsymbol{s})}\right] \le h(p)B + \frac{1}{n^C}.$$

- *Given both $\boldsymbol{s}$ and $K^{(\boldsymbol{s})}$, each output coordinate of a sample can be computed from $O(1)$ many words of $\boldsymbol{m} \sim \{0,1\}^{\boldsymbol{M}}$.*

The proof of Theorem 15 follows [HLY+24], where we encode information of $\boldsymbol{s}$ into the succinct data structure in [Yu22] with a small increase in redundancy.

Lemma 16 (Lemma 28 in [Yu22])**.** *Let $B = \mathrm{polylog}(n)$, and $C > 0$ be any constant. A size-s subset $S \subseteq [B]$ can be represented by a spillover representation $(m', k') \in \{0,1\}^{M'} \times [K']$ such that*

- $K' = \mathrm{poly}(n)$,
- $M' + \log K' \le \log \binom{B}{s} + O(1/n^C)$,
- *each query can be answered with $O(1)$ word probes to m' and k'.*

Proof Sketch of Theorem 15. As in [HLY+24, Lemma 4.2], we first instantiate Lemma 16 to represent a size-s subset of $[B]$ with a spillover representation $(m', k') \in \{0,1\}^{M'} \times [K']$. Then we encode s and $D(s)$ into (m', k'), for some distribution D that is $n^{-\Omega(C)}$-close to $\mathrm{Bin}(B,p)$ into the first t bits of m'.

We think of a t-bit string as the set $T := \{0, \dots, 2^t - 1\}$. Let D' be $\mathrm{Bin}(B,p)$ conditioned on its value lies in $[pB - B^{2/3}, pB + B^{2/3}]$. Note that D' is $n^{-\omega(1)}$ close to $\mathrm{Bin}(B,p)$. For each s in the support of D', we assign an interval $T_s \subseteq T$ of $\lfloor D'(s) \cdot 2^t \rfloor$ elements. For any point $x \in T$ that is not in any T_s, we assign it to an arbitrary T_s. Defining the distribution $D(s) = \frac{|T_s|}{|T|}$, one can verify that D is $n^{-\Omega(C)}$-close to $\mathrm{Bin}(B,p)$.

To encode $\boldsymbol{s}$, we take the first $2t$ bits m_0 of m', and view m_0 as a number in $\{0, \dots, 2^{2t} - 1\}$. We can write m_0 as

$$m_0 = \left\lfloor \frac{m_0}{|T_s|} \right\rfloor \cdot |T_s| + m_0 \bmod |T_s|.$$

We replace the first t bits of m' with the $(m_0 \bmod |T_s|)$-th value in the interval T_s in binary. Then we remove the second block of t bits of m', and encode $\left\lfloor \frac{m_0}{|T_s|} \right\rfloor$ along with k' as the spill, which increases K' to $K' \cdot \lceil \frac{2^{2t}}{|T_s|} \rceil$. A similar calculation as in [HLY+24] shows that the redundancy increases by $1/\mathrm{poly}(n)$.

To sample $(\boldsymbol{m}, \boldsymbol{k})$, our sampler will sample the first t bits of $\boldsymbol{m}$ to sample $\boldsymbol{s} \sim D$. This lets us determine $\boldsymbol{s}$, $\boldsymbol{m}_0 \bmod |T_{\boldsymbol{s}}|$, and $(M^{(s)}, K^{(s)})$. Given the sizes, we can sample the rest of $(\boldsymbol{m}, \boldsymbol{k})$. From $\boldsymbol{k}$, we can recover $\boldsymbol{m}_0$. This lets us recover $(\boldsymbol{m}', \boldsymbol{k}')$.

We have obtained a spillover representation $(\boldsymbol{m}, \boldsymbol{k}) \in \{0,1\}^{\boldsymbol{M}} \times [\boldsymbol{K}]$ of a set $S \subseteq [B]$ of size $\boldsymbol{s}$, with $\boldsymbol{s}$ encoded in the first t bits of $\boldsymbol{m}$ with redundancy $1/\mathrm{poly}(n)$. Let X be the resulting distribution. By Lemma 16, we have

$$\mathbb{E}_{\boldsymbol{s}\sim D}[\boldsymbol{M} + \log_2 \boldsymbol{K}'] \le \mathbb{E}_{\boldsymbol{s}\sim D}\left[\log \frac{1}{D(\boldsymbol{s})} + \log \binom{B}{\boldsymbol{s}}\right] + O\left(\frac{1}{n^C}\right) = H(X) + O\left(\frac{1}{n^C}\right).$$

Since D is $n^{-\Omega(C)}$-close to $\mathrm{Bin}(B,p)$, X is $1/n^{-\Omega(C)}$-close to $\mathrm{Ber}(p)^B$, and therefore $H(X) \le h(p)B + 1/n^{\Omega(C)}$. The theorem then follows from Claim 6. □

3.3 Concatenation

We now show to sample the L copies of the sampler in Theorem 15. Specifically, we will sample L i.i.d. copies of $(\boldsymbol{m}_i, \boldsymbol{k}_i) \in \{0,1\}^{M^{(\boldsymbol{s}_i)}} \times [K^{(\boldsymbol{s}_i)}]$.

Recall that D is a distribution supported on $[pB - B^{2/3}, pB + B^{2/3}]$, and each $(m, k) \in \{0,1\}^{M^{(s)}} \times [K^{(s)}]$ is a spill representation of a size-s subset in $[B]$. Thus, for every s in its support of D, we have $M^{(s)} \in [M_{\min}, M_{\max}]$ for some $M_{\min} \geq \log \binom{B}{pB-B^{2/3}}/2$ and $M_{\max} := \log \binom{B}{pB+B^{2/3}}$. Moreover, as $s \in [pB - B^{3/2}, pB + B^{3/2}]$, there are $S \leq 2B^{3/2} + 1$ many possible different values $K_1, \ldots, K_S$ for $K^{(s)}$. Let $p_j := \Pr[K^{(\boldsymbol{s})} = K_j] = \sum_{s:K^{(s)}=K_j} D(s)$. The number of $K^{(\boldsymbol{s}_i)}$'s that are equal to K_j is distributed according to $\mathrm{Bin}(L, p_j)$, and thus is at most $p_j L + O(\sqrt{L \log n})$ with probability $1/\operatorname{poly}(n)$.

For each $i \in [L]$, we partition $\boldsymbol{m}_i$ into $(\boldsymbol{m}_{i,1}, \boldsymbol{m}_{i,2}, \boldsymbol{m}'_i)$, where

- $\boldsymbol{m}_{i,1}$ has length $t := C \log_2 n$,
- $\boldsymbol{m}_{i,2}$ has length $M_{\min} - t$,
- $\boldsymbol{m}'_i$ is the remaining of $\boldsymbol{m}_i$.

(Recall that $\boldsymbol{M}^{(\boldsymbol{s}_i)} \geq M_{\min}$.) We will assume the lengths $|\boldsymbol{m}_{i,2}|$ and $|\boldsymbol{m}'_i|$ are integer multiples of the word size w, and $K^{(\boldsymbol{s}_i)} \geq n^C$ for a large enough c. These can be achieved by moving some bits in $\boldsymbol{m}_{i,2}$ and $\boldsymbol{m}'_i$ to the spill if necessary, which only change the spill size $K^{(\boldsymbol{s}_i)}$ by $\operatorname{poly}(n)$.

Let $\boldsymbol{m}_{\text{fixed}}$ denote the concatenation of the $\boldsymbol{m}_{i,2} : i \in [L]$, which has length at least $L \cdot (M_{\min} - t) \geq LM_{\min}/2$. We partition $\boldsymbol{m}_{\text{fixed}}$ into $S + 1$ blocks

$$\left(\boldsymbol{m}_{\text{fixed},1}, \ldots, \boldsymbol{m}_{\text{fixed},S}, \boldsymbol{m}'_{\text{fixed}}\right),$$

where $|\boldsymbol{m}_{\text{fixed},j}| = \frac{LM_{\min}}{4S}$ for $j \in [S]$ and $\boldsymbol{m}'_{\text{fixed}}$ contains the remaining $\geq \frac{LM_{\min}}{4}$ bits.

First, for each $i \in [L]$, we sample $\boldsymbol{s}_i$ using each $\boldsymbol{m}_{i,1}$. Then, we will sample different portions of $\boldsymbol{m}_{\text{fixed}}$ together with $\boldsymbol{m}'_i$ and the spills $\boldsymbol{k}_i$'s. For each $j \in [S]$, we sample $\boldsymbol{k}_i : i \in R_j$ together with $\boldsymbol{m}_{\text{fixed},j}$ using Lemma 18. Finally, we sample $\boldsymbol{m}'_1, \ldots, \boldsymbol{m}'_L$ together with $\boldsymbol{m}_{\text{fixed},S+1}$ using Lemma 19.

The proofs of both lemmas use the following sparse matrix for the augmented retrieval data structure in [HLY+24].

Lemma 17. *Let $\mathbb{F}$ be a finite field of size at least n^C. Let $\boldsymbol{S}$ be a random subset of $[U]$ of size at most r. Suppose $M_{\text{fixed}} \geq U \log n$. Then there exists a $(U + M_{\text{fixed}}) \times (r + M_{\text{fixed}})$ matrix G over $\mathbb{F}$ with $O(1)$ nonzeros in every row such that*

$$\Pr\left[G_{\boldsymbol{S} \cup \{U+1, \ldots, U+M_{\text{fixed}}\}} \text{ is full rank}\right] \geq 1 - \frac{1}{n^{C/2}}.$$

We sketch its proof in Appendix C.

Sampling the spills. We first show how to sample the spills. We will use the following lemma whose proof is deferred to Section 3.3.1.

Lemma 18 (Concatenating Spills)**.** *Let $n^C \le K \le \mathrm{poly}(n)$, and $w = \Theta(\log n)$. Let $\boldsymbol{S} \subseteq [L]$ be a random subset of size at most r, and $M = \Omega(wL \log L)$. There is a function $f\colon \{0,1\}^m \to [K]^L \times \{0,1\}^M$ with*

$$m \le M + r \log_2 K + O(\log_2 n),$$

such that each element of $f(U)$ only depends on $O(1)$ words of size w of the input. Moreover, letting $(\boldsymbol{k}, \boldsymbol{m}) \in [K]^L \times \{0,1\}^M$ be the output distribution, and using $\boldsymbol{k}_S$ to denote $\boldsymbol{k}_i : i \in S$, we have

$$\Pr_{\boldsymbol{S}}\Big[(\boldsymbol{k}_{\boldsymbol{S}}, \boldsymbol{m}) \text{ is } 1/\mathrm{poly}(n)\text{-close to uniform}\Big] \ge 1 - 1/\mathrm{poly}(n).$$

Fix a $j \in [S]$. Our goal is to sample the $\boldsymbol{k}_i$'s where $K^{\boldsymbol{s}_i} = K_j$ together with $\boldsymbol{m}_{\mathrm{fixed},j}$. We apply Lemma 18 with

$$\begin{aligned} r_j &:= p_j L + O(\sqrt{L \log n}), \text{and} \\ M &:= |\boldsymbol{m}_{\mathrm{fixed},j}| = \Theta(n/B^{2/3}) \ge \Omega(wL \log L) \end{aligned}$$

to obtain a sampler that uses a seed of length

$$m_j = M + r \log_2 K + O(\log_2 n) \le |\boldsymbol{m}_{\mathrm{fixed},j}| + p_j L \log_2 K + O\big(\sqrt{L} \cdot \log_2^{3/2} n\big).$$

We now calculate the total number of bits used to sample $\boldsymbol{m}_{\mathrm{fixed},j}$ and all the $\boldsymbol{k}_i$'s. Note that

$$\begin{aligned} \sum_{j=1}^{S} (p_j L) \log_2 K_j &= L \sum_{j=1}^{S} \Big(\sum_{s : K^{\boldsymbol{s}_i} = K_j} D(s) \Big) \log_2 K^{(s)} \\ &= L \sum_{s} D(s) \log_2 K^{(s)} \\ &= L\, \mathbb{E}_{\boldsymbol{s}}\big[\log_2 K^{\boldsymbol{s}}\big]. \end{aligned}$$

Therefore, overall the sampler uses a seed of length

$$\sum_{j=1}^{S} |\boldsymbol{m}_{\mathrm{fixed},j}| + L\, \mathbb{E}[\log_2 K^{\boldsymbol{s}}] + O\big(\sqrt{L} \cdot \log_2^{3/2} n\big). \tag{3}$$

Sampling the variable length part. We will use the following lemma to concatenate the $\boldsymbol{m}_i'$'s.

Lemma 19 (Concatenating variable-length part)**.** *Let $\boldsymbol{S} \subseteq [Q]$ be a random subset of size at most r, and $M = \Omega(Q \log Q)$. There is a function $f\colon \{0,1\}^m \to (\{0,1\}^w)^{Q+M}$ with $m \le w(r+M)$ such that each word of $f(U)$ only depends on $O(1)$ words of the input. Moreover, letting $(\boldsymbol{x}, \boldsymbol{y}) \in (\{0,1\}^w)^r \times (\{0,1\}^w)^M$ be the output distribution, we have*

$$\Pr_{\boldsymbol{S}}\Big[(\boldsymbol{x}_{\boldsymbol{S}}, \boldsymbol{y}) \text{ is uniform}\Big] \ge 1 - 1/\mathrm{poly}(n).$$

Proof. Let G be the $(Q+M)\times(r+M)$ matrix over $\mathbb{F}=GF(2^w)$ given by Lemma 17. Let $(\boldsymbol{x},\boldsymbol{y})\in(\{0,1\}^w)^Q\times(\{0,1\}^w)^M$ be the output of G on an uniform input. We have

$$\Pr_{\boldsymbol{S}}\big[(\boldsymbol{x}_{\boldsymbol{S}},\boldsymbol{y}) \text{ is uniform}\big]\geq 1-1/\operatorname{poly}(n).$$

We can sample the uniform distribution over $(\{0,1\}^w)^{r+M}$ using $w(r+M)$ bits. □

Let

$$\Delta := M_{\max}-M_{\min}=O(B^{2/3}\log B).$$

Note that $\boldsymbol{m}'_1,\dots,\boldsymbol{m}'_L$ all together consist of $\boldsymbol{r}:=\sum_{i=1}^{L}\frac{(M^{(\boldsymbol{s}_i)}-M_{\min})}{w}\leq\frac{L\Delta}{w}$ words. Moreover, we have

$$O\Big(\frac{L\Delta}{w}\log(L\Delta)\Big)=O(LB^{2/3}\log n)=O\Big(\frac{n}{B^{1/3}}\log n\Big)\leq\frac{LM_{\min}}{4w}\leq\frac{|\boldsymbol{m}'_{\text{fixed}}|}{w}.$$

So we can apply Lemma 19 with $\boldsymbol{r}$ to sample $\boldsymbol{m}'_1,\dots,\boldsymbol{m}'_L$ together with $\boldsymbol{m}'_{\text{fixed}}$ using a seed of length

$$\sum_{i=1}^{L}|\boldsymbol{m}'_i|+|\boldsymbol{m}'_{\text{fixed}}|. \tag{4}$$

The error is $1/\operatorname{poly}(n)$.

Putting everything together. We first argue correctness. Observe that our choice of r_j is fixed when we sample the spills $\boldsymbol{k}_i$'s. Thus, to sample the spills $\boldsymbol{k}_i$'s uniformly, we need to ensure that the number of blocks i with $K^{\boldsymbol{s}_i}$ is at most $r_j=p_jL+O(\sqrt{L\log n})$ with high probability. Let $\boldsymbol{R}_j$ be the set of i where $K^{(\boldsymbol{s}_i)}=K_j$. Note that $\mathbb{E}[|\boldsymbol{R}_j|]=p_jL$, and thus by the Chernoff bound and a union bound, we have that for each $j\in[S]$,

$$\Pr\Big[|\boldsymbol{R}_j|\geq p_jL+10\sqrt{B\log n}\Big]\leq\frac{1}{n^{100}}.$$

So with probability $1-1/\operatorname{poly}(n)$, $|\boldsymbol{R}_j|\leq p_jL+10\sqrt{B\log n}$ for every $j\in[S]$, and the spills are sampled uniformly.

The overall seed length is the sum of the lengths of $\boldsymbol{m}_{i,1}:i\in[L]$, Eqs. (3) and (4). This is

$$\begin{aligned}&\sum_{i=1}^{L}\big(|\boldsymbol{m}_{i,1}|+|\boldsymbol{m}'_i|\big)+|\boldsymbol{m}'_{\text{fixed}}|+\sum_{j=1}^{S}|\boldsymbol{m}_{\text{fixed},j}|+L\,\mathbb{E}\big[\log_2 K^{\boldsymbol{s}}\big]+O\big(\sqrt{L}\log_2^{3/2}n\big)\\&\quad=\sum_{i=1}^{L}M^{(\boldsymbol{s}_i)}+L\,\mathbb{E}\big[\log_2 K^{\boldsymbol{s}}\big]+O(\sqrt{L}\log_2^{3/2}n).\end{aligned} \tag{5}$$

By Theorem 15, we have

$$\mathbb{E}\big[M^{(\boldsymbol{s}_i)}+\log_2 K^{(\boldsymbol{s}_i)}\big]\leq h(p)B+\frac{1}{\operatorname{poly}(n)}.$$

As $M^{(s)} \le B$ for every s, by Hoeffding's inequality, we have that with probability $1 - 1/\operatorname{poly}(n)$,

$$\sum_{i=1}^{L} M^{(\boldsymbol{s}_i)} \le L \cdot \mathbb{E}[M^{\boldsymbol{s}}] + O(B\sqrt{L \log n}),$$

and so

$$\begin{aligned} \text{Eq. (5)} &\le L\,\mathbb{E}\left[M^{(\boldsymbol{s}_i)} + \log_2 K^{(\boldsymbol{s}_i)}\right] + O(\sqrt{L}\log_2^{3/2} n) \\ &\le h(p)n + \sqrt{n} \cdot \operatorname{polylog}(n). \end{aligned}$$

3.3.1 Proof of Lemma 18

We now prove Lemma 18. The proofs are similar to the ones in [HLY+24], except we maintain the closeness to the uniform distribution when using Lemma 14 to change base.

We will need the following result from number theory [BHP01].

Lemma 20. *For every sufficiently large n, there is a prime between n and $n + n^{0.525}$.*

Proof of Lemma 18. We embed elements in $[K]$ into $[p]$ using Claim 13, where p is the smallest prime that is at least K, which, by Lemma 20, is at most $K + K^{0.525}$. Thus, the total variation distance between the uniform distributions over $[K]^L$ and $[p]^L$ is at most

$$L \cdot \frac{K^{0.525}}{K + K^{7/11}} \le 2 \cdot L \cdot K^{-0.475} \le 1/\operatorname{poly}(n). \tag{6}$$

We apply Lemma 14 to $(\{0,1\}^w)^M$ to obtain a sampler $f_1 \colon [p]^{M_p} \to \{0,1\}^M$ with $M_p \le \frac{M}{\log_2 p} + O(\log_p n)$ and error $1/\operatorname{poly}(n)$ Note that $M_p \ge \Omega(L \log L)$. We sample $[p]^{M_p + L}$ using Lemma 17. Specifically, let G be the $(L + M_p) \times (r + M_p)$ matrix over $\mathbb{F}_p$ given by Lemma 17 with respect to $\boldsymbol{S}$. This matrix has $O(1)$ many nonzeros in each row. Moreover, letting $(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_L, \boldsymbol{m})$ be the output of $G(U)$, we have

$$\Pr_{\boldsymbol{S}}\Big[(\boldsymbol{k}_{\boldsymbol{S}}, \boldsymbol{m}) \text{ is uniform}\Big] \ge 1 - 1/n^C. \tag{7}$$

Finally, we use Lemma 14 to sample the uniform distribution over $[p]^{r+M_p}$ from $\{0,1\}^m$ with error $1/\operatorname{poly}(n)$.

Therefore, we have

$$\begin{aligned} m &\le (r + M_p)\log_2 p + O(\log_2 n) \\ &\le \left(r + \frac{M}{\log_2 p}\right)\log_2 p + O(\log_2 n) \\ &= r \log_2 p + M + O(\log_2 n) \\ &= r \log_2 K + O(\log_2 n). \end{aligned}$$

Closeness to uniform follows from Lemma 14, and locality follows since each sampler is $O(1)$ word-local, and thus their composition is also $O(1)$ word-local. □

4 A lower bound for a 2-local construction with $m = (2-\epsilon)n$

In this section, we prove Theorem 21 to show that for any 2-local mapping with seed length is $(2-\epsilon)n$ its distance to $\mathrm{Ber}(1/4)^n$ approaches 1 as n increases.

Theorem 21. *Let $n \in \mathbb{N}$ be sufficiently large. Fix $\epsilon > 0$, and let $m = (2-\epsilon)n$. Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a 2-local mapping. Then*

$$\mathrm{dist}(f(U_m, \mathrm{Ber}(1/4)^n) \geq 1 - \exp(-c\epsilon n),$$

for some $c > 0$ is some absolute constant.

Before proving the theorem, we will prove several claims that will be needed later.

Proposition 22. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a 1-local mapping. Then*

$$\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) \geq 1 - 2 \cdot e^{-n/128}.$$

Note that since f is 1-local, we may assume without loss of generality that $m \leq n$.

Proof. For each $i \in [m]$ corresponding to the input bit x_i, let $N(i)$ be the output bits that depend on x_i. Note that we may assume without loss of generality that $|N(i)| \geq 1$ for all $i \in [m]$, as otherwise it we can remove the i'th coordinate. Since f is 1-local, the sets $N(i)$ and $N(i')$ are disjoint for $i \neq i'$, and the distributions $f(x_i)_{|N(i)}$ and $f(x_{i'})_{|N(i')}$ are independent. Next we consider the following two cases:

- If $m > n/2$, we may pick for each $i \in [m]$ one output coordinate $j_i \in N(i)$. Note that the corresponding output bit has distribution $\mathrm{Ber}(1/2)$, and the joint distribution $f(U_m)_{(j_i)_{i\in[m]}}$ is $\mathrm{Ber}(1/2)^m$. Thus

$$\begin{aligned}
\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) &\geq \mathrm{dist}(\mathrm{Ber}(1/2)^m, \mathrm{Ber}(1/4)^m)) \\
&\geq \Pr[\mathrm{Bin}(m,1/2) \geq 3m/8] - \Pr[\mathrm{Bin}(m,1/4) \geq 3m/8] \\
&\geq (1 - e^{-m/64}) - e^{-m/36} \geq 1 - 2 \cdot e^{-m/64} \geq 1 - 2 \cdot e^{-n/128}.
\end{aligned}$$

- If $m \leq n/2$, then $|\mathrm{supp}(f(U_m)| \leq 2^m \leq 2^{n/2}$. On the other hand for any subset $A \subseteq \{0,1\}^n$ of size at most 2^m it holds that $\Pr[\mathrm{Ber}(1/4)^n \in A] \leq \Pr[\mathrm{Bin}(n,1/4) \leq n/8]$, as $\mathrm{Ber}(1/4)^n$ assigns higher probability to the elements of lower weight and $\binom{n}{\leq n/8} \geq \frac{1}{\sqrt{n}} \cdot 2^{h(1/8)n} \geq 2^{n/2} \geq 2^m$. Therefore, by Claim 41 we have

$$\mathrm{dist}(\mathrm{Ber}(1/4)^n, f(U)) \geq 1 - \Pr[\mathrm{Bin}(n,1/4) \leq n/8] \geq 1 - e^{-n/64}.$$

This completes the proof of Proposition 22. □

Claim 23. *Let $f\colon \{0,1\}^m \to \{0,1\}^2$ be a 2-local mapping. Let $i \in [m]$ be a coordinates of the input, and let $N(i)$ be the output bits that are influenced by the i'th input bit. If $|N(i)| \geq 2$, then $\mathrm{dist}(f(U_m)_{|N(i)}, \mathrm{Ber}(1/4)^{|N(i)|}) \geq 1/8$.*

Proof. Take any two distinct coordinates $j, j' \in N(i)$. These two coordinates depend on at most three input bits, and hence all probabilities of $f(U_m)_{\{j,j'\}}$ are integer multiples of 1/8. On the other hand, the distribution $\mathrm{Ber}(1/4)^2$ has probabilities $(1/16, 3/16, 3/16, 9/16)$, and thus, each possible 2-bit string contributes at least $\frac{1}{8}$ to each term of the summation in the definition of the distance. Therefore $\mathrm{dist}(f(U_m)_{|N(i)}, \mathrm{Ber}(1/4)^{|N(i)|}) \geq \frac{1}{2} \cdot (4 \cdot \frac{1}{16}) = 1/8$. □

We are now ready to prove Theorem 21.

Proof of Theorem 21. Given a 2-local mapping $f\colon \{0,1\}^m \to \{0,1\}^n$ with $m = (2-\epsilon)n$, define a bipartite graph $G = (I \cup O, E)$, where the vertices in I correspond to the m coordinates of the input, O corresponds to the n coordinates of the output, and $(i, o) \in E$ if the o'th output bit depends on the i'th input coordinate. That is, $|I| = m$, $|O| = n$, and $|E| = 2n$ since f is 2-local.

We fix two large constants $C = 99$, and $D = 4(C+1)/\epsilon = 400/\epsilon$. Let $I^* \subseteq I$ be a maximal subset of I such that $|N(I^*)| \geq C|I^*|$, and consider the following two cases.

- $|N(I^*)| > n/D$: In this case, for any fixing of the inputs $(x_i)_{i \in I^*}$, the mapping $(f_j)_{j \in N(I^*)}$ is 1-local. Therefore, conditioning on $(x_i)_{i \in I^*}$ being fixed, by Proposition 22 the 1-local mapping satisfies

  $$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \geq \mathrm{dist}(f(U_{I^*}), \mathrm{Ber}(1/4)^{|N(I^*)|}) \geq 1 - 2 \cdot e^{-|N(I^*)|/128}.$$

  Applying Claim 38 with all $2^{|I^*|}$ assignments to the input bits in I^* we get

  $$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \geq 1 - 2 \sum_{s \in \{0,1\}^{|I^*|}} 2 \cdot e^{-|N(I^*)|/128} \geq 1 - 4 \cdot 2^{|I^*|} \cdot e^{-(|N(I^*)|/128}.$$

  Next, we use the assumption that $|I^*| \leq |N(I^*)|/C$ and $|N(I^*)| \geq n/D$ together with our choice of $C = 99$ and $D = 4(C+1)/\epsilon = 400/\epsilon$ to get

  $$\begin{aligned} \mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) &\geq 1 - 4 \cdot 2^{|N(I^*)|/C} \cdot e^{-|N(I^*)|/128} \\ &\geq 1 - 4 \cdot e^{-\frac{1/128 - \ln(2)/C}{D} n} \\ &\geq 1 - \exp(-\Omega(\epsilon n)). \end{aligned}$$

  This proves Theorem 21 in case of $|N(I^*)| \geq n/D$.

- $|N(I^*)| \leq n/D$: In this case our strategy is the following. We will remove $N(I^*)$ from the output coordinates. The remaining mapping $f'\colon \{0,1\}^m \to \{0,1\}^{n'}$ will satisfy the property that $m < (2-\epsilon/2)n'$ and each input coordinate influences at most C output nodes. This will allow us to find a collection $\mathcal{O}$ of $\Omega(\epsilon n)$ disjoint subsets of output coordinates $(O_i)_{i \in \mathcal{O}}$ such that

  1. $\mathrm{dist}(f(U_{m'})_{|O_i}, \mathrm{Ber}(1/4)^{|O_i|}) \geq 1/8$ for all $O_i \in \mathcal{O}$,
  2. $\left(f(U_{m'})_{|O_i}\right)_{O_i \in \mathcal{O}}$ are jointly independent.

Then, by applying Claim 40 we conclude that $\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \geq 1 - \exp(-\Omega(\epsilon n))$. We describe the details below.

Note that by maximality of I^* we have $|N(i) \setminus N(I^*)| \leq C$ for all $i \in I \setminus I^*$. Therefore, by removing $N(I^*)$ from the set of outputs, we get a graph $G' = (I' = I, O', E')$ such that the degree of each $i \in I'$ is at most C.

Since we removed at most $n/D = \epsilon n/400$ output vertices, the new graph has $m' = (2-\epsilon)n$ inputs and $n' \geq (1 - \epsilon/400)n$ outputs. Therefore, $m' \leq (2-\epsilon')n'$ for

$$\epsilon' = 2 - \frac{m'}{n'} = 2 - \frac{2-\epsilon}{1-\epsilon/400} > \epsilon/2.$$

Therefore, we now have a 2-local mapping $f' \colon \{0,1\}^{m=(2-\epsilon')n'} \to \{0,1\}^{n'}$ with $\epsilon' > \epsilon/2$ such that each input coordinate of f influences at most C output bits, and f' has the same distribution as f on the remaining output coordinates.

Let $J = \{i \in I' : \deg_{G'}(i) \geq 2\}$.

Claim 24. $|J| \geq \frac{\epsilon'}{2C} n'$.

Proof. The proof is a simple application of Markov's inequality. Since $\deg(v) \leq C$ for all $i \in I'$, we have

$$\frac{2n'}{(2-\epsilon')n'} = \mathbb{E}_{i \in I'}[\deg(i)] \leq \Pr[\deg(i) \leq 1] + C \cdot \Pr[\deg(i) > 1] \leq 1 + C \Pr[\deg(i) > 1].$$

Since $\deg(i)$ is an integer, we get $\Pr[\deg(i) \geq 2] = \Pr[\deg(i) > 1] \geq \frac{\epsilon'}{(2-\epsilon')C} > \frac{\epsilon'}{2C}$, as required. □

Now, since each input coordinate in J has degree at most C, we can find a subset $K \subseteq J$ of size $|K| \geq |K|/(C+1)$ such that $N(i)$ and $N(i')$ do not have common neighbours for all distinct $i, i' \in K$. Indeed, this is achieved by taking any $i \in J$, adding it to K and removing from J all neighbours of $N(i)$.

This gives us a collection of input coordinates $K \subseteq I'$ of size $|K| \geq |J|/(C+1) \geq \frac{\epsilon'}{2C(C+1)} n'$, such that each $i \in K$ has $\deg(i) \geq 2$ and $\left(f(U_m)_{|N(i)}\right)_{i \in K}$ are jointly independent.

By Claim 23 we have $\mathrm{dist}(f(U_m)_{|N(i)}, \mathrm{Ber}(1/4)^{|N(i)|}) \geq 1/8$ for all $i \in K$. Therefore, applying Claim 40 on $\left(f(U_m)_{|N(i)}\right)_{i \in K}$ we get

$$\begin{aligned}
\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) &\geq \mathrm{dist}(f'(U_m), \mathrm{Ber}(1/4)^{n'}) \\
&\geq 1 - 2e^{-\frac{(1/8)^2 |K|}{12}} \\
&\geq 1 - 2e^{-\frac{\epsilon' n'}{8^2 \cdot 12 \cdot 2C(C+1)}} \\
&\geq 1 - \exp(-\Omega(\epsilon n)).
\end{aligned}$$

This completes the proof of Theorem 21. □

5 A 3-local construction with $m = 1.99n$ that is $1/\operatorname{poly}(n)$-close to $\mathrm{Ber}(1/4)^n$

In this section, we show that in contrast to Theorem 21, if we allow the sampler to be 3-local, we can approximate the distribution $\mathrm{Ber}(1/4)^n$ within distance of $1/\operatorname{poly}(n)$, and this is optimal up to constant factor in the exponent.

Theorem 25. *Fix an integer $t \geq 3$ and let $\epsilon = 1/3t$. Let $n \in N$ be sufficiently large and let $m = (2-\epsilon)n$. Then, there is a 3-local mapping $f\colon \{0,1\}^m \to \{0,1\}^n$ such that*

$$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \leq (\frac{1}{2\epsilon n})^{\frac{2}{9\epsilon}-\frac{5}{3}}.$$

In particular, for $m = (2-1/9)n$ there is a 3-local mapping $f\colon \{0,1\}^m \to \{0,1\}^n$ such that

$$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \leq \frac{2}{n^{1/3}}.$$

Proof. Let $G' = (V', E')$ be a 3-regular graph with k vertices and $1.5k$ edges such that the girth of G' is $\geq \frac{2}{3}\log_2(|V'|) = \frac{2}{3}\log_2(k)$. Indeed, such graphs exists [Mor94, Theorem 5.13].

Claim 26. *Let $p = 2^{-t}$ for some $t \geq 3$, and let $G'_p = (V, E_p)$ be a random subgraph of G' obtained by keeping each edge in E with probability p independently. Then $\Pr[G'_p$ has a cycle$] < k^{-\frac{2t-5}{3}}$.*

Proof. By the assumption about G', it has no cycles of length $< \frac{2}{3}\log_2(k)$. For any $\ell \geq \frac{2}{3}\log_2(k)$ the number of cycles of length ℓ is at most $k \cdot 3 \cdot 2^{\ell-2}$. Therefore,

$$\Pr[G'_p \text{ has a cycle of length } \ell] \leq k \cdot 3 \cdot 2^{\ell-2} \cdot p^\ell = \frac{3k}{4}(2p)^\ell.$$

Taking the union bound over all lengths $\ell > \frac{2}{3}\log_2(k)$, we get

$$\Pr[G'_p \text{ has a cycle}] < \frac{3k}{4} \cdot \sum_{\ell=\frac{2}{3}\log_2(k)}^{\infty} (2p)^\ell = \frac{3k}{4} \cdot \frac{(2p)^{\frac{2}{3}\log_2(k)}}{1-2p} < \frac{1}{k^{\frac{2t-5}{3}}}, \tag{8}$$

as required. □

Given the graph G' above, we define a graph $G = (V, E)$ by subdividing each edge of G' into $t \geq 3$ edges. The number of vertices in G is $|V| = |V'| + (t-1)|E'| = k + 1.5(t-1)k = (1.5t - 0.5)k$, and the number of edges is $|E| = 1.5tk$.

Let $n = |E| = 1.5tk$ and $m = |V| + |E| = (3t - 0.5)k = (2 - 1/3t)n = (2-\epsilon)n$, and define the mapping $f\colon \{0,1\}^m \to \{0,1\}^n$ as follows. Treat the input to f as $x \circ y \in \{0,1\}^m$, where $x \in \{0,1\}^n$, and $y \in \{0,1\}^{m-n}$. It will be convenient to think of x_i as an assignment to the i'th edge of G, and y_j as an assignment to the j'th vertex of G. Define $g\colon \{0,1\}^{m-n} \to \{0,1\}^n$ by

letting $g_i(y) = y_{u_i} \oplus y_{v_i}$, where (u_i, v_i) are the endpoints of the i'th edge in G. Finally, we define f as the bitwise AND of x and $g(y)$, that is,

$$f(x \circ y) = x \wedge_n g(y).$$

We show below that

$$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \leq \Pr[G_{1/2} \text{ has a cycle}], \tag{9}$$

where $G_{1/2}$ is a random subgraph of G obtained by keeping each edge with probability 1/2. This proves Theorem 25 by combining Claim 26 with the observation that $\Pr[G_{1/2} \text{ has a cycle}] = \Pr[G'_{2^{-t}} \text{ has a cycle}]$.

For Eq. (9) recall that the input to our sampler f consists of $x \in \{0,1\}^n$ and $y \in \{0,1\}^{m-n}$, where x is the 0/1 assignment to the edges, and y is the 0/1 assignment to the vertices. For $x \in \{0,1\}^n$ let $E_x = \{i \in [n] : x_i = 1\}$, and observe that for all $i \in [n] \setminus E_x$ it holds that $f_i(x \circ y) = 0$ for any choice of y.

Denote by $F \subseteq \{0,1\}^n$ the event that the edges in G corresponding to the set E_x induce a forest, i.e., the subgraph (V, E_x) does not contain a cycle. Observe that if $x \in F$, then $(y_{u_i} \oplus y_{v_i})_{i \in E_x}$ is distributed as $\mathrm{Ber}(1/2)^{E_x}$.

Next, we fix any $A \subseteq \{0,1\}^n$, and show that $|\Pr[f(U_m) \in A] - \Pr[\mathrm{Ber}(1/4)^n \in A]| \leq \Pr[x \notin F]$. Using the natural correspondence between E_x and $G_{1/2}$, this clearly implies Eq. (9). Indeed,

$$\begin{aligned}
&\left|\Pr[f(U_m) \in A] - \Pr[\mathrm{Ber}(1/4)^n \in A]\right| \\
&\quad = \left| \Pr_{\substack{x \in \{0,1\}^n \\ y \in \{0,1\}^{m-n}}} [x \wedge_n g(y) \in A] - \Pr_{\substack{x \in \{0,1\}^n \\ z \in \{0,1\}^n}} [x \wedge_n z \in A] \right| \\
&\quad \leq \left| \Pr[(x \wedge_n g(y) \in A) \wedge x \in F] - \Pr[(x \wedge_n z \in A) \wedge x \in F] \right| \\
&\quad + \left| \Pr[(x \wedge_n g(y) \in A) \wedge x \notin F] - \Pr[(x \wedge_n z \in A) \wedge x \notin F] \right| \\
&\quad \leq 0 + \Pr[x \notin F] < \frac{1}{k^{\frac{2t-5}{3}}}.
\end{aligned}$$

This concludes the proof of Theorem 25. □

5.1 Theorem 25 is tight

Below we show that the analysis of the construction presented in the proof of Theorem 25 is tight up to the exponent of the polynomial. Specifically, we prove the following proposition.

Proposition 27. *Fix $\epsilon > 0$, and let $n \geq 1/\epsilon$ be an integer. Consider the sampler $f \colon \{0,1\}^{(2-\epsilon)n} \to \{0,1\}^n$ from Theorem 25. Then*

$$\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) \geq (2\epsilon n)^{-4\epsilon/3}.$$

The proposition relies on the following claim.

Claim 28. *Let G be the graph in the proof of Theorem 25. If G contains a cycle of length ℓ then,* $\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) \geq (1/4)^\ell$.

Proof. Let $\mathbf{b} = b_1, \ldots, b_\ell$ be the output bits of our cycle, further let $x_1, \ldots, x_\ell$ be the input bits associated with the edges of our cycle, and let $y_1, ..., y_\ell$ be the input bits associated with the nodes. That is for $i < \ell$, $b_i = x_i \wedge (y_i \oplus y_{i+1})$, and $b_\ell = x_\ell \wedge (y_\ell \oplus y_1)$.

It's clear that $\Pr[\mathrm{Ber}(1/4)^\ell = \mathbf{1}] = (1/4)^\ell$. Now, let's examine $\Pr[\mathbf{b} = \mathbf{1}]$. We consider two cases.

- ℓ is odd: In this case $\Pr[\mathbf{b} = \mathbf{1}] = 0$ because in order for this to occur all x_i must be 1, and the y_i's must alternate on the cycle, which is impossible of a cycle of odd length. Thus, $\Pr[\mathbf{b} = \mathbf{1}] = 0$ and $\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) = |\Pr[\mathbf{b} = \mathbf{1}] - \Pr[\mathrm{Ber}(1/4)^\ell = \mathbf{1}]| \geq |0 - (1/4)^\ell| = (1/4)^\ell$.
- ℓ is even: Here $\mathbf{b} = \mathbf{1}$ happens if and only if all $x_1, \ldots, x_\ell$ are 1, which happens with probability $(1/2)^\ell$, and $y_i's$ alternate, so $y_1 = 0, y_2 = 1, \ldots, y_\ell = 1$ or $y_1 = 1, y_2 = 0, \ldots, y_\ell = 0$, which happens with probability $2 \cdot (1/2)^\ell$. Thus $\Pr[\mathbf{b} = \mathbf{1}] = 0$ and $\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) = |\Pr[\mathbf{b} = \mathbf{1}] - \Pr[\mathrm{Ber}(1/4)^\ell = \mathbf{1}]| \geq |2(1/4)^\ell - (1/4)^\ell| = (1/4)^\ell$.

In both cases we have $\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) \geq (1/4)^\ell$. □

We now prove Proposition 27.

Proof of Proposition 27. We consider $G' = (V', E')$ as in the proof of Theorem 25. Recall that G' has k vertices, $1.5k$ edges and is 3-regular. We first show there exists a cycle in G' of length at most $\ell = O(\log_2(k))$

Let's run a Breadth First Search algorithm starting at any arbitrary node $s \in V$, and stop once the BFS tree reaches a cycle in G'. This cycle has length at most $2d + 1$, where d is the height of the tree.

Since G' is 3-regular, we add exactly 2 vertices (3 for the first node) into our visited queue on each iteration of BFS. This means $k = |V'| \geq 1 + 3\sum_{i=0}^{d-1} 2^i = 3 \cdot 2^d - 2$, and hence we have a cycle of length $2\log_2(\frac{k+2}{3}) + 1$ Recalling that $n = 1.5tk = k/2\epsilon$ and that each edge in G' corresponds to a path of length $t = 1/3\epsilon$ in G, we conclude that G has a cycle of length

$$\ell \leq \frac{2\log_2(\frac{2\epsilon n+2}{3}) + 1}{3\epsilon} \leq \frac{2\log_2(2\epsilon n)}{3\epsilon}.$$

Hence, by Claim 28, we get that $\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) \geq (1/4)^\ell \geq (2\epsilon n)^{-4\epsilon/3}$. □

6 A lower bound on 3-local constructions with $m = (2 - \epsilon)n$

In this section we prove a lower bound on the distance for all 3-local constructions with $m = 1.99n$.

Theorem 29. *Let $n \in \mathbb{N}$ be sufficiently large. Fix $\epsilon > 0$, and let $m = (2-\epsilon)n$. Then, for any 3-local mapping $f\colon \{0,1\}^m \to \{0,1\}^n$ it holds that*

$$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \geq n^{-O(1/\epsilon)}.$$

Proof. Let $G = (V = I \cup O, E)$ be a bipartite graph, where $|I| = m = (2-\epsilon)n$ represents the input bits of f, $|O| = n$ represents the output bits, and $(i,o) \in E$ if and only if the o'th output bit of f depends on the i'th input bit.

We prove Theorem 29 by finding a small set of outputs $S \subset O$ such that its neighbourhood $N(S)$ (i.e., the input bits of S) is small. Specifically, we will find a set $S \subseteq O$ of size $|S| = k = O(\log(n))$ such that $|N(S)| \leq 2k-1$. This indeed suffices, as for the distribution $\mathrm{Ber}(1/4)^k$ the probability of sampling all zeros in S is exactly 4^{-k}, while the granularity of the inputs to S implies that f outputs all zeros in S with probability either 0 or at least $2^{-|N(S)|} = 2 \cdot 4^{-k}$. Therefore,

$$|\Pr[f(U_m)_{|S} \equiv 0] - \Pr[(\mathrm{Ber}(1/4)^n)_{|S} \equiv 0]| \geq 4^{-k}.$$

In order to find such set S, note that the graph G has $|V| = (3-\epsilon)n$ vertices and $|E| = 3n$ edges. Therefore, $|E| = (1+\epsilon')|V|$ for $\epsilon' = \epsilon/(3-\epsilon)$.

We use the following lemma, saying that any sufficiently dense graph contains a set of vertices S that span at least $|S|+1$ edges, such that $|S| = O(\log(n))$.

Lemma 30 (Theorem 2 in [GGS23])**.** *Let $G = (V,E)$ be a multigraph with $|V| \geq 2$ vertices and $|E| = m \geq (1+\epsilon)|V|$ edges for some $\epsilon = \epsilon(|V|) \in (0,1]$. There exists a set of vertices $S \subseteq V$ of size $|S| \leq 8\log(|V|) \cdot \lceil 1/\epsilon \rceil$ spanning at least $|S|+1$ edges.*

Applying Lemma 30 to G, we get a subset of the vertices $C \subset V$ of size $|C| \leq 8\log(|V|) \cdot \lceil 1/\epsilon' \rceil$ that spans at least $|C|+1$ edges. By taking the minimal such subset C, we may assume[1] that all vertices $v \in C$ have at least two neighbours in C.

The key step of the proof is summarized in the following claim.

Claim 31. *Let $G' = (V' = I' \cup O', E')$ be the bipartite subgraph of G induced by C with $I' = I \cap C$ and $O' = O \cap C$, and let $k = |O'|$. Then $|N_G(O')| \leq 2k-1$.*

Proof. Since $|E'| \geq |V'|+1$, there must be at least one vertex in G' of degree ≥ 3. Recall that all vertices in O' have degree either 2 or 3, and denote by t the number of vertices in O' of degree 3. Consider the following two cases.

- $t = 0$: Since all vertices in O' have degree 2, the set I' must have a vertex of degree ≥ 3 in G'. Furthermore, $|E'| = 2 \cdot |O'| = 2k$, and hence by counting degrees of the vertices in I', we have $|I'| \leq k-1$. Finally, note that each $v \in O'$ has at most one neighbour outside C, and thus $|N_G(O')| \leq |I'| + |O'| \leq (k-1) + k = 2k-1$.

[1] Otherwise, if C has a vertex with $\deg_C(v) \leq 1$, we can remove v from C, and the remaining subset C' will also satisfy the property that $|C'| \leq 8\log(|V|) \cdot \lceil 1/\epsilon' \rceil$ and it spans at least $|C'|+1$ edges.

- $t \geq 1$: By counting the degrees of O' in G' note that $|E'| = 2(k-t) + 3t = 2k + t$. Similarly, by counting the degrees of I' in G', we have $|E'| \geq 2|I'|$. Finally, there are exactly $k - t$ nodes $v \in O'$ with one neighbour outside C, and t nodes $v \in O'$ with no neighbours outside C. Therefore,

$$|N_G(O')| \leq |I'| + (k-t) \leq |E'|/2 + (k-t) \leq (k + t/2) + (k-t) \leq 2k - t/2.$$

  Since $|N_G(O')|$ is an integer and $t \geq 1$, it follows that $|N_G(O')| \leq 2k - 1$.

In both cases we showed that $|N_G(O')| \leq 2k - 1$, as required. □

Therefore, letting $S = C \cap O$, we get a set of size $k = |S| \leq |C| \leq 8\log(|V|) \cdot \lceil 1/\epsilon' \rceil \leq 8\log(3n) \cdot 3/\epsilon$ such that $|N(S)| \leq 2k - 1$. By the discussion above this implies that

$$\mathrm{dist}(f(U), \mathrm{Ber}(1/4)^n) \geq |\Pr[f(U_m)_{|S} \equiv 0] - \Pr[(\mathrm{Ber}(1/4)^n)_{|S} \equiv 0]| \geq 4^{-k} \geq (3n)^{-48/\epsilon}.$$

This completes the proof of Theorem 29. □

7 A 4-local construction with $m = 1.75n$ that is $\exp(-cn)$-close to $\mathrm{Ber}(1/4)^n$

In this section we prove that 4-local samplers can approximate the distribution $\mathrm{Ber}(1/4)^n$ within exponentially small distance.

Theorem 32. *Let $n \in \mathbb{N}$ be sufficiently large, and let $m = (2 - 1/4)n$. Then, there exists a 4-local mapping $f\colon \{0,1\}^m \to \{0,1\}^n$ such that*

$$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \leq 2^{-cn},$$

for some absolute constant $c > 0$.

The proof of Theorem 32 relies on the following lemma and its corollary below.

Lemma 33. *Let m, n be parameters such that $n \leq m \leq 2n$, and let $M \in \mathbb{F}_2^{n \times (m-n)}$ be a matrix such that every row of M has exactly D ones. Then, there exists a $(D+1)$-local sampler $f\colon \{0,1\}^m \to \{0,1\}^n$ satisfying*

$$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \leq \Pr_{S \subseteq [n]}[\textit{the rows of } M_S \textit{ are linearly dependent}], \tag{10}$$

where M_S is the submatrix of M obtained by taking only the rows of M with indices in S.

Corollary 34. *Let m, n be parameters such that $n \leq m \leq 2n$, and let $M \in \mathbb{F}_2^{n \times (m-n)}$ be a matrix such that every row of M has exactly D ones, and let $C = \{x \in \{0,1\}^n : xM = 0\}$.*

1. *There exists a* $(D+1)$*-local sampler* $f\colon \{0,1\}^m \to \{0,1\}^n$ *satisfying*

$$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \leq \sum_{\ell=1}^{n} \frac{w_\ell}{2^\ell},$$

where $w_\ell = |\{x \in C : |x| = \ell\}|$ *for all* $1 \leq \ell \leq n$.

2. *There exists a* $(D+1)$*-local sampler* $f\colon \{0,1\}^m \to \{0,1\}^n$ *satisfying*

$$\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \leq \frac{2^{n-\mathrm{rank}(M)}}{2^{\delta n}},$$

where $\delta = \min\{|x| : x \in C\}$ *is the minimum weight of a vector in* C.

Next we state a result about the existence of a sparse matrix satisfying the conditions in Corollary 34. We show the existence of such matrix by adapting Gallager's result on random sparse matrices [Gal62]. The proof of Theorem 35 can be found in Appendix B.

Theorem 35. *For a sufficiently large* $n \in \mathbb{N}$ *there exists a matrix* $M \in \mathbb{F}_2^{n \times 0.75n}$ *such that every row of* M *has exactly* 3 *ones and*

$$\sum_{\ell=1}^{n} \frac{w_\ell}{2^\ell} < 2^{-0.05n},$$

where $w_\ell = |\{x \in \mathbb{F}_2^n : xM = 0, |x| = \ell\}|$ *for* $\ell = 1, \ldots, n$.

Let us now show how Theorem 32 follows from Corollary 34.

Proof of Theorem 32. Let $M \in \mathbb{F}_2^{n \times 0.75n}$ be the matrix from Theorem 35. The matrix has 3 ones in each row and satisfies $\sum_{\ell=1}^{n} \frac{w_\ell}{2^\ell} < 2^{-0.05n}$. Using the first part of Corollary 34 we get a 4-local sampler $f\colon \{0,1\}^{1.75n} \to \{0,1\}^n$ such that $\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n) \leq 2^{-0.05n}$, as required. □

We now return to proving Lemma 33 and Corollary 34.

Proof of Lemma 33. Given a matrix $M \in \mathbb{F}_2^{n \times (m-n)}$ as in the assumption of the lemma, define the $(D+1)$-local sampler $f\colon \{0,1\}^m \to \{0,1\}^n$ as follows. Treat the input to f as $x \circ y$, where $x \in \{0,1\}^n$ and $y \in \{0,1\}^{m-n}$, and define

$$f(x \circ y) = x \wedge_n My, \tag{11}$$

where My is the matrix-vector multiplication modulo 2, and $\wedge_n$ is the coordinate-wise AND operation.

Next we prove that f satisfies the guarantees of the lemma. For any $x \in \{0,1\}^n$ let $S_x = \{i \in [n] : x_i = 1\}$, and define

$$Good = \{x \in \{0,1\}^n \backslash \{0\} : \text{the rows of } M_{S_x} \text{ are linearly independent}\} \text{ and } Bad = \{0,1\}^n \backslash Good. \tag{12}$$

Given the natural correspondence between the sets $S \subseteq [n]$ and vectors $x \in \{0,1\}^n$, we need to prove that

$$\Pr[\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n)] \leq \Pr[x \in Bad].$$

We observe first that for any fixed $x \in Good$ the distributions $x \wedge_n My$ and $x \wedge_n z$ are *identical*. That is, for any $a \in \{0,1\}^n$ it holds that

$$\Pr_{y \sim \mathrm{Ber}(1/2)^{m-n}}[x \wedge_n My = a] = \Pr_{z \sim \mathrm{Ber}(1/2)^n}[x \wedge_n z = a]. \tag{13}$$

Indeed, note first that if $S_a \not\subseteq S_x$, then both sides of Eq. (13) are equal to zero. Next assume that $S_a \subseteq S_x$.

- For $z \sim \mathrm{Ber}(1/2)^n$ we have

$$\Pr[x \wedge z = a] = \Pr[x|_{S_x} \wedge z|_{S_x} = a|_{S_x}] = \Pr[z|_{S_x} = a|_{S_x}] = \frac{1}{2^{|S_x|}}.$$

- For $y \sim \mathrm{Ber}(1/2)^{m-n}$ we have

$$\Pr[x \wedge My = a] = \Pr[x|_{S_x} \wedge My|_{S_x} = a|_{S_x}] = \Pr[(M|_{S_x})y = a|_{S_x}] = \frac{1}{2^{|S_x|}},$$

  where the last equality uses the assumption that the rows of $M|_{S_x}$ are linearly independent, and hence for each $b \in \{0,1\}^{S_x}$ the equation $(M|_{S_x})y = b$ has exactly $2^{m-n-|S_x|}$ solutions. This proves Eq. (13).

Using Eq. (13), we can complete our proof. Indeed, let $A \subseteq \{0,1\}^n$ be any event on n bit strings.

$$\begin{aligned}
|\Pr[x \wedge_n My \in A] - \Pr[x \wedge z \in A]| =& |\Pr[(x \wedge_n My \in A) \wedge x \in Good] + \Pr[(x \wedge_n My \in A) \wedge x \in Bad] \\
&- \Pr[(x \wedge_n z \in A) \wedge x \in Good] - \Pr[(x \wedge_n z \in A) \wedge x \in Bad]| \\
\leq& |\Pr[(x \wedge_n My \in A) \wedge x \in Good] - \Pr[(x \wedge_n z \in A) \wedge x \in Good]| \\
&+ |\Pr[(x \wedge_n My \in A) \wedge x \in Bad] - \Pr[(x \wedge_n z \in A) \wedge x \in Bad]|.
\end{aligned}$$

By Eq. (13) we have

$$|\Pr[(x \wedge_n My \in A) \wedge x \in Good] - \Pr[(x \wedge_n z \in A) \wedge x \in Good]| = 0,$$

and

$$|\Pr[(x \wedge_n My \in A) \wedge x \in Bad] - \Pr[(x \wedge_n z \in A) \wedge x \in Bad]| \leq \Pr[x \in Bad].$$

Therefore $|\Pr[\mathrm{dist}(f(U_m), \mathrm{Ber}(1/4)^n)] \leq \Pr[x \in Bad]$, as required. □

Proof of Corollary 34. Using the definition of Bad in Eq. (12), it suffices to show that

$$\Pr_{x \in \mathbb{F}_2^n}[x \in Bad] \leq \sum_{s=1}^{n} \frac{w_s}{2^s}.$$

Indeed, note that $x \in Bad$ if and only if there exists some $y \in C \setminus \{0\}$ such that $S_y \subseteq S_x$. Furthermore, if $y \in C$ has weight $|y| = s$, then there there are exactly 2^{n-s} many x's satisfying $S_y \subseteq S_x$. Therefore,

$$|Bad| \leq \sum_{x \in C} 2^{n-|x|} = \sum_{s=1}^{n} w_s \cdot 2^{n-s}.$$

This implies

$$\Pr_{x \in \mathbb{F}_2^n}[x \in Bad] = \frac{|Bad|}{2^n} \leq \sum_{s=1}^{n} \frac{w_s}{2^s},$$

as required.

For the second item note that $w_s = 0$ for all $s \leq \delta n$, and hence

$$\Pr_{x \in \mathbb{F}_2^n}[x \in Bad] \leq \sum_{s=1}^{n} \frac{w_s}{2^s} \leq \sum_{s=\delta n}^{n} \frac{w_s}{2^{\delta n}} = \frac{1}{2^{\delta n}} \cdot \left(\sum_{s=1}^{n} w_s \right) = \frac{1}{2^{\delta n}} \cdot |C| = \frac{2^{n-\mathrm{rank}(M)}}{2^{\delta n}}.$$

This completes the proof of Corollary 34. □

References

[Bab87] László Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Inform. Process. Lett.*, 26(1):51–53, 1987. 1

[BCS16] Itai Benjamini, Gil Cohen, and Igor Shinkar. Bi-Lipschitz bijection between the Boolean cube and the Hamming ball. *Israel J. Math.*, 212(2):677–703, 2016. 1

[BHP01] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes, ii. *Proceedings of the London Mathematical Society*, 83(3):532–562, 2001. 3.3.1

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

[BKW03] Avrim Blum, Adam Kalai, and Hal Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *J. ACM*, 50(4):506–519, 2003. 1

[BL87] R. B. Boppana and J. C. Lagarias. One-way functions and circuit complexity. *Inform. and Comput.*, 74(3):226–240, 1987. 1

[BS23] Lucas Boczkowski and Igor Shinkar. On mappings on the hypercube with small average stretch. *Combin. Probab. Comput.*, 32(2):334–348, 2023. 1

[CGZ22] Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The space complexity of sampling. In *13th Innovations in Theoretical Computer Science Conference*, volume 215 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 40, 23. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2022. 1

[CS16] Gil Cohen and Leonard J. Schulman. Extractors for near logarithmic min-entropy. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*, pages 178–187. IEEE Computer Soc., Los Alamitos, CA, 2016. 1

[CT06] Thomas Cover and Joy Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006. 1

[CZ19] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. *Ann. of Math. (2)*, 189(3):653–705, 2019. 1

[DILV24a] Harm Derksen, Peter Ivanov, Chin Ho Lee, and Emanuele Viola. Pseudorandomness, symmetry, smoothing: I. In *Conf. on Computational Complexity (CCC)*, 2024. 1

[DILV24b] Harm Derksen, Peter Ivanov, Chin Ho Lee, and Emanuele Viola. Pseudorandomness, symmetry, smoothing: II. 2024. 1

[DPT10] Yevgeniy Dodis, Mihai Pătraşcu, and Mikkel Thorup. Changing base without losing space. In *STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 593–602. ACM, New York, 2010. 1.2, 3.1

[FLRS23] Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. Sampling and certifying symmetric functions. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 275 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 36, 21. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2023. 1

[Gal62] R. G. Gallager. Low-density parity-check codes. *IRE Trans.*, IT-8:21–28, 1962. 7

[GGS23] Alexander Golovnev, Tom Gur, and Igor Shinkar. *Derandomization of Cell Sampling*, pages 278–284. 2023. 30

[GW20] Mika Göös and Thomas Watson. A lower bound for sampling disjoint sets. *ACM Trans. Comput. Theory*, 12(3):Art. 20, 13, 2020. 1

[Hag91] Torben Hagerup. Fast parallel generation of random permutations. In *Automata, languages and programming (Madrid, 1991)*, volume 510 of *Lecture Notes in Comput. Sci.*, pages 405–416. Springer, Berlin, 1991. 1

[Has86] Johan Torkel Hastad. *COMPUTATIONAL LIMITATIONS OF SMALL DEPTH CIRCUITS*. ProQuest LLC, Ann Arbor, MI, 1986. Thesis (Ph.D.)–Massachusetts Institute of Technology. 1

[HH23] Pooya Hatami and William Hoza. Theory of unconditional pseudorandom generators. In *Electron. Colloquium Comput. Complex., TR23-019*, volume 1, page 1, 2023. 1, 1.2

[HLY+24] Yang Hu, Jingxun Liang, Huacheng Yu, Junkai Zhang, and Renfei Zhou. Optimal static dictionary with worst-case constant query time. *arXiv preprint arXiv:2412.10655*, 2024. 1.2, 1.2, 3.2, 3.2, 3.3, 3.3.1, C, D

[IN96] Russell Impagliazzo and Moni Naor. Efficient cryptographic schemes provably as secure as subset sum. *J. Cryptology*, 9(4):199–216, 1996. 1

[KOW24a] Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locally sampleable uniform symmetric distributions. *arXiv preprint arXiv:2411.08183*, 2024. 1

[KOW24b] Daniel M. Kane, Anthony Ostuni, and Kewen Wu. Locality bounds for sampling Hamming slices. In *STOC'24—Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1279–1286. ACM, New York, [2024] ©2024. 1, 1.1, A.1

[LV12] Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. *Comput. Complexity*, 21(2):245–266, 2012. 1, 1.1

[Mor94] Moshe Morgenstern. Existence and explicit constructions of $q + 1$ regular Ramanujan graphs for every prime power q. *J. Combin. Theory Ser. B*, 62(1):44–62, 1994. 5

[MV91] Yossi Matias and Uzi Vishkin. Converting high probability into nearly-constant time—with applications to parallel hashing. In *Proceedings of the twenty-third annual ACM symposium on Theory of Computing*, pages 307–316, 1991. 1

[Păt08] Mihai Pătraşcu. Succincter. In *49th IEEE Symp. on Foundations of Computer Science (FOCS)*. IEEE, 2008. 1.2

[Reg05] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '05, pages 84–93, New York, NY, USA, 2005. Association for Computing Machinery. 1

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 77–82, 1987. 1

[SS24] Ronen Shaltiel and Jad Silbak. Explicit codes for poly-size circuits and functions that are hard to sample on low entropy distributions. In *STOC'24—Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 2028–2038. ACM, New York, [2024] ©2024. 1

[Vio12] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012. 1, 1.2, 1.2, 2, A.1

[Vio14a] Emanuele Viola. Extractors for circuit sources. *SIAM J. Comput.*, 43(2):655–672, 2014. 1

[Vio14b] Emanuele Viola. Is nature a low-complexity sampler?, 2014. Blog post, accessed April 2, 2025. 1

[Vio23] Emanuele Viola. New sampling lower bounds via the separator. In *38th Computational Complexity Conference*, volume 264 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 26, 23. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2023. 1, 1.1, A.1

[Vio24] Emanuele Viola, 2024. https://emanueleviola.wordpress.com/2024/11/11/15-years-of-complexity-of-distributions/. 1

[VWY20] Emanuele Viola, Omri Weinstein, and Huacheng Yu. How to store a random walk. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, 2020. Available at http://www.ccs.neu.edu/home/viola/. 1.1

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

[Yu22] Huacheng Yu. Nearly optimal static Las Vegas succinct dictionary. *SIAM J. Comput.*, 51(3):STOC20–174–STOC20–249, 2022. 1.2, 3.2, 16

[YZ24] Huacheng Yu and Wei Zhan. Sampling, flowers and communication. In *15th Innovations in Theoretical Computer Science Conference*, volume 287 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 100, 11. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2024. 1

A Background facts

In this section for completeness we state and prove several basic facts that are used in deriving our main results.

A.1 On the total variation distance between two distributions

Definition 36. *Given two distribution μ, ν over a finite set X, the* total variation distance *between μ and ν is defined as*

$$\mathrm{dist}(\mu, \nu) = \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)|.$$

We will also refer to it as the statistical distance *between μ and ν.*

The following is a standard fact about the total variation distance between two distributions.

Fact 37. *Given two distribution μ, ν over a finite set X, we have*

$$\mathrm{dist}(\mu, \nu) = \max_{A \subseteq X} |\mu(A) - \nu(A)| = \mu(A^*) - \nu(A^*),$$

where $A^* = \{x \in X : \mu(x) > \nu(x)\}$.

We will need the following claims. Similar claims have been shown, e.g., in [Vio12, Vio23, KOW24b]. We prove them here for completeness.

Claim 38. *Let μ, ν be two distributions over a finite domain X. Let $\nu_1, \ldots, \nu_k$ be k distributions over X, such that $\nu = \frac{1}{k} \sum_{i=1}^{k} \nu_i$, and suppose that $\mathrm{dist}(\mu, \nu_i) = 1 - \epsilon_i$ for some $\epsilon_i \in [0, 1/2]$. Then*

$$1 - 2 \sum_{i=1}^{k} \epsilon_i \leq \mathrm{dist}(\mu, \nu) \leq 1 - \frac{1}{k} \sum_{i=1}^{k} \epsilon_i.$$

Proof. For the upper bound let $A \subseteq X$ be such that $\mathrm{dist}(\mu, \nu) = \mu(A) - \nu(A)$. Then

$$\mathrm{dist}(\mu, \nu) = \mu(A) - \nu(A) = \frac{1}{k} \sum_{i=1}^{k} \big(\mu(A) - \nu_i(A)\big) \leq \frac{1}{k} \sum_{i=1}^{k} (1 - \epsilon_i) = 1 - \frac{1}{k} \sum_{i=1}^{k} \epsilon_i.$$

For the lower bound, let $A_i = \{x \in X : \mu(x) > \nu_i(x)\}$. Then we have $\mathrm{dist}(\mu, \nu_i) = \mu(A_i) - \nu_i(A_i)$. In particular $\mu(A_i) \geq 1 - \epsilon_i$ and $\nu_i(A_i) \leq \epsilon_i$. Consider the set $A = \cap_{i=1}^{k} A_i$, and note that $\mu(A) \geq 1 - \sum_{i=1}^{k} \mu(X \setminus A_i) \geq 1 - \sum_{i=1}^{k} \epsilon_i$. On the other hand $\nu(A) = \frac{1}{k} \sum_{i=1}^{k} \nu_i(A) \leq \frac{1}{k} \sum_{i=1}^{k} \nu_i(A_i) \leq \frac{1}{k} \sum_{i=1}^{k} \epsilon_i$. Therefore,

$$\mu(A) - \nu(A) \geq \Big(1 - \sum_{i=1}^{k} \epsilon_i\Big) - \frac{1}{k} \sum_{i=1}^{k} \epsilon_i = 1 - \Big(1 + \frac{1}{k}\Big) \cdot \sum_{i=1}^{k} \epsilon_i,$$

as required. □

Claim 39. *Let μ_X, ν_X be two distributions over X, and let μ_Y, ν_Y two distributions over Y. Consider the product distributions $\mu_X \times \mu_Y$ and $\nu_X \times \nu_Y$. Then*

$$\mathrm{dist}(\mu_X \times \mu_Y, \nu_X \times \nu_Y) \leq \mathrm{dist}(\mu_X, \nu_X) + \mathrm{dist}(\mu_Y, \nu_Y).$$

Proof. For each $x \in X$ define $\delta_X(x) = \mu_X(x) - \nu_X(x)$. Similarly, for each $y \in Y$ define $\delta_Y(y) = \mu_Y(y) - \nu_Y(y)$. Note that $\mathrm{dist}(\mu_X, \nu_X) = \frac{1}{2}\sum_{x\in X}|\delta_X(x)|$ and $\mathrm{dist}(\mu_Y, \nu_Y) = \frac{1}{2}\sum_{y\in Y}|\delta_Y(y)|$. Then

$$\begin{aligned}
\mathrm{dist}(\mu_X \times \mu_Y, \nu_X \times \nu_Y) &= \frac{1}{2}\sum_{x\in X}\sum_{y\in Y}|\mu_X(x)\cdot\mu_Y(y) - \nu_X(x)\cdot\nu_Y(y)| \\
&= \frac{1}{2}\sum_{x\in X}\sum_{y\in Y}|\nu_X(x)\delta_Y(y) + \mu_Y(y)\delta_X(x)| \\
&\leq \frac{1}{2}\sum_{x\in X}\sum_{y\in Y}|\nu_X(x)\cdot\delta_Y(y)| + \frac{1}{2}\sum_{x\in X}\sum_{y\in Y}|\mu_Y(y)\cdot\delta_X(x)| \\
&= \frac{1}{2}\Big(\sum_{x\in X}\nu_X(x)\Big)\Big(\sum_{y\in Y}|\delta_Y(y)|\Big) + \frac{1}{2}\Big(\sum_{x\in X}|\delta_X(x)|\Big)\Big(\sum_{y\in Y}\mu_Y(y)\Big) \\
&= \mathrm{dist}(\mu_X, \nu_X) + \mathrm{dist}(\mu_Y, \nu_Y),
\end{aligned}$$

as required. □

Claim 40. *Let $n \in \mathbb{N}$ and for each $i \in [n]$ let μ_i and ν_i be two distributions over a domain X_i. Suppose that for each $i \in [n]$ it holds that $\mathrm{dist}(\mu_i, \nu_i) \geq \epsilon$. Define the product distributions $\mu = \mu_1 \times \mu_2 \times \cdots \times \nu_n$ and $\nu = \nu_1 \times \nu_2 \times \cdots \times \nu_n$ over the domain $X = X_1 \times X_2 \times \cdots \times X_n$.*

$$\mathrm{dist}(\mu, \nu) \geq 1 - 2e^{-\frac{\epsilon^2 n}{12}}.$$

Proof. For each $i \in [k]$ let $B_i \subseteq X_i$ be such that $\mu_i(B_i) \geq \nu_i(B_i) + \epsilon$. Define $p_i = (\mu_i(B_i) + \nu_i(B_i))/2$. Given a random $x = (x_1, \ldots, x_n) \in X_1 \times X_2 \times \cdots \times X_n$, define $S_x = |\{i \in [n] : x_i \in B_i\}|$. Define $A = \{S_x \geq \sum_{i=1}^n p_i\}$. Then using Chernoff bound we have

$$\begin{aligned}
&\mathrm{dist}(\mu, \nu) \\
&\quad\geq \mu(A) - \nu(A) \\
&\quad= \Pr_{\mu}\left[S_x \geq \left(1 - \frac{\epsilon}{2\cdot\frac{1}{n}\sum\mu_i(B_i)}\right)\sum\mu_i(B_i)\right] - \Pr_{\nu}\left[S_x \geq \left(1 + \frac{\epsilon}{2\cdot\frac{1}{n}\sum\nu_i(B_i)}\right)\sum\nu_i(B_i)\right] \\
&\quad\geq \left(1 - e^{-\frac{\epsilon^2 n}{8\frac{1}{n}\sum\mu_i(B_i)}}\right) - e^{-\frac{\epsilon^2 n}{12\frac{1}{n}\sum\nu_i(B_i)}} \\
&\quad\geq 1 - 2e^{-\frac{\epsilon^2 n}{12}}.
\end{aligned}$$

This proves Claim 40. □

Claim 41. *Let μ be a distribution over X such that for any subset $X' \subseteq X$ of size $|X'| = k$ it holds that $\mu(X') \leq \epsilon$. Let ν be a distribution over X such that $\mathrm{supp}(\nu) \leq k$. Then $\mathrm{dist}(\mu, \nu) \geq 1 - \epsilon$.*

Proof. Let $A = \mathrm{supp}(\nu)$. Then by Fact 37 we have $\mathrm{dist}(\mu, \nu) \geq \nu(A) - \mu(A) \geq 1 - \epsilon$. □

A.2 Entropy, binomial coefficients, concentration inequalities, etc

We start with the definition of the entropy of a distribution.

Definition 42. *Given a distribution $\mathcal{D}$ over a finite domain X, we define the entropy of $\mathcal{D}$ as* $H(\mathcal{D}) = \sum_{x \in X} \mathcal{D}(x) \log_2(\frac{1}{\mathcal{D}(x)})$.

Next we define the *binary entropy function*, which corresponds to the entropy of a Bernoulli random variable with the appropriate parameter.

Definition 43. *The binary entropy function is defined as* $h(x) = x \log_2(\frac{1}{x}) + (1-x) \log_2(\frac{1}{1-x})$.

Fact 44. *For all* $1 \leq k \leq n-1$ *it holds that*

$$\sqrt{\frac{n}{8k(n-k)}} \cdot 2^{h(\frac{k}{n})n} \leq \binom{n}{k} \leq \sum_{i=0}^{k} \binom{n}{i} \leq 2^{h(\frac{k}{n})n}.$$

Theorem 45 (Chernoff bound)**.** *Let $X_1, \ldots, X_n$ be independent random variables with $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$ for each i. Let $X = \sum_{i=1}^{n} X_i$, and let $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$. Then*

1. $\Pr[X \geq (1+\epsilon)\mu] \leq e^{-\frac{\epsilon^2 \mu}{3}}$ *for all* $\epsilon > 0$.
2. $\Pr[X \leq (1-\epsilon)\mu] \leq e^{-\frac{\epsilon^2 \mu}{2}}$ *for all* $\epsilon \in (0, 1)$.

Theorem 46 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables such that $a \leq X_i \leq b$ for each i. Let $X = \sum_{i=1}^{n} X_i$, and let $\mu = \mathbb{E}[X]$. Then*

1. $\Pr[X \geq \mu + t] \leq e^{-\frac{2t^2}{(b-a)n}}$ *for all* $t > 0$.
2. $\Pr[|X - \mu| \geq t] \leq 2e^{-\frac{2t^2}{(b-a)n}}$ *for all* $t > 0$.

B Proof of Theorem 35

For a parameter $n \in \mathbb{N}$ define $\mathcal{B}$ to be the uniform distribution over matrices in $\mathbb{F}_2^{n \times n/4}$ with exactly one entry equal to 1 in each row and exactly four 1's in each column. Define a random matrix $M = [B_1, B_2, B_3] \in \mathbb{F}_2^{n \times 3n/4}$, where each $B_i \in \mathbb{F}_2^{n \times n/4}$ is distributed according to $\mathcal{B}$ independently.

Theorem 35 is proved using the following lemma.

Lemma 47. *For a sufficiently large n, let $M \in \mathbb{F}_2^{n \times 3n/4}$ be a random matrix from the distribution described above, and let $\delta = 0.03$. For $C = \{x \in \mathbb{F}_2^n : xM = 0\}$ let $w_\ell = |\{x \in C : |x| = \ell\}|$ be the weight distribution of C. Then*

- $\Pr[\sum_{\ell=1}^{\delta n} w_\ell = 0] > 0.1$ *and*
- $\Pr[\sum_{\ell=\delta n}^{n} \frac{w_\ell}{2^\ell} < 0.96^n] > 0.98$.

In particular, there exists a matrix $M \in \mathbb{F}_2^{n \times 3n/4}$ *such that* $\sum_{\ell=1}^{n} \frac{w_\ell}{2^\ell} < 0.96^n < 2^{-0.05n}$.

The key idea in the proof of Lemma 47 is to understand $\mathbb{E}[w_\ell]$, the expected number of vectors $x \in \mathbb{F}_2^n$ of weight ℓ satisfying $xM = 0$. In order to do it, define

$$g(s) = 1 + \binom{4}{2} \cdot 2^{2s} + 2^{4s}. \tag{14}$$

Note that the coefficient of $2^{\ell s}$ is equal to the size of the set $\{x \in \mathbb{F}_2^4 : |x| = \ell \wedge x_1 + x_2 + x_3 + x_4 = 0\}$.

Claim 48. *Let* $0 \le \ell \le n$*, and denote by* $N_{\mathcal{B}}[\ell]$ *the expected number of vectors* $x \in \mathbb{F}_2^n$ *of weight* ℓ *satisfying* $xB = 0$*. Then,* $N_{\mathcal{B}}[\ell] \le \frac{g(s)^{n/4}}{2^{s\ell}}$ *for any* $s \in \mathbb{R}$*.*

Proof. Consider the function $g(s)^{n/4}$, and write it as

$$(g(s))^{n/4} = \sum_{\ell=0}^{n} Q(\ell) 2^{\ell s}, \tag{15}$$

and observe that by definition of $g(s)$ we have $Q(\ell) = |\{x \in \mathbb{F}_2^n : x\mathcal{B} = 0 \wedge |x| = \ell]\}|$. Now since $Q(\ell) \ge 0$ and $2^{\ell s} \ge 0$, it follows that $g(s)^{n/4} \ge Q(\ell) \cdot 2^{\ell s}$ for all $\ell \ge 0$ and any $s \in \mathbb{R}$. Therefore,

$$N_{\mathcal{B}}[\ell] = Q(\ell) \le \frac{g(s)^{n/4}}{2^{s\ell}},$$

as required. □

Define a function $f\colon [0,1] \to \mathbb{R}$ as

$$f(\lambda) = \frac{(1 + 6 \cdot 2^{2s} + 2^{4s})^{3/4}}{2^{3s\lambda} \cdot 2^{2h(\lambda)}}, \tag{16}$$

for any parameter $s \in \mathbb{R}$.

Claim 49. *For* $\ell = 1, \ldots, n$ *let* $\lambda = \ell/n$*. Then* $\mathbb{E}[w_\ell] \le 8\lambda n f(\lambda)^n$ *for all values of* s *in the definition of* f*.*

Proof. Since in the definition of $M = [B_1, B_2, B_3]$ the B_i's are independent, it follows that

$$\mathbb{E}[w_\ell] = \binom{n}{\ell} \cdot \left(\frac{N_B[\ell]}{\binom{n}{\ell}} \right)^3 = \frac{(N_B[\ell])^3}{\binom{n}{\ell}^2} \le \frac{8\ell(n-\ell)}{n} \cdot \frac{g(s)^{3n/4}}{2^{3s\ell} \cdot 2^{2h(\ell/n)n}} \le 8\lambda(1-\lambda)n \cdot f(\lambda)^n,$$

for $\lambda = \ell/n$, where the first inequality uses the fact that $\binom{n}{\lambda n} \ge \frac{1}{\sqrt{8\lambda(1-\lambda)n}} 2^{h(\lambda)n}$. □

Claim 50. *Let* $\delta = 0.03$ *as in Lemma 47. For any* $\lambda \in (0, \delta]$ *there exists* $s = s(\lambda)$ *such that* $f(\lambda) < 0.75^\lambda$*.*

Claim 51. *Let $\delta = 0.03$ as in Lemma 47. For all $\lambda \in [\delta, 1]$ there exists $s = s(\lambda)$ such that $\frac{f(\lambda)}{2^\lambda} < 0.95$.*

We postpone the proof of the claims until later, and show below how the two claims above imply Lemma 47.

Proof of Lemma 47. Observe that since each row of M has an odd number of 1's, it follows that $w_\ell = 0$ for all odd values of ℓ. Note also that for any constant even k it holds that $\Pr[w_k > 0] = O_k(n^{-k/2})$.

Claim 52. *For any constant even k it holds that $\Pr[w_k > 0] \leq \frac{(6k)^{3k}}{n^{k/2}}$*

Proof. Note that if $w_k > 0$, then there are k rows and at most $3k/2$ columns of M such that the ones of the k rows are all contained in these $3k/2$ columns. Therefore

$$\Pr[w_k > 0] \leq \binom{n}{k} \cdot \binom{3n/4}{3k/2} \times \left(\frac{3k/2}{n/4}\right)^{3k} \leq n^k \cdot n^{1.5k} \times \frac{(6k)^{3k}}{n^{3k}} = \frac{(6k)^{3k}}{n^{k/2}},$$

as required. □

In particular, for a sufficiently large n we have

$$\Pr[\sum_{\ell=1}^{20} w_\ell = 0] > 1 - O(1/n) > 0.99. \tag{17}$$

Next we use Claim 49 and Claim 50 to bound $\mathbb{E}[\sum_{\ell=22}^{\delta n} w_\ell]$. Let $p = 0.75$ be the base of the exponent in Claim 50. Then

$$\begin{aligned}
\mathbb{E}[\sum_{\ell=22}^{\delta n} w_\ell] &\leq \sum_{\substack{\ell=22 \\ \ell \text{ even}}}^{\infty} \frac{8\ell(n-\ell)}{n} \cdot p^\ell < \sum_{\substack{\ell=22 \\ \ell \text{ even}}}^{\infty} 8\ell \cdot p^\ell \\
&\leq 16 \cdot \sum_{j=11}^{\infty} j \cdot p^{2j} \\
&= 16 \cdot \left(\frac{p^2}{(1-p^2)^2} - \sum_{j=1}^{10} j \cdot p^{2j} \right) \\
&= 16 \cdot \left(\frac{p^2}{(1-p^2)^2} - \frac{p^2(1 + 10p^{22} - 11p^{20})}{(1-p^2)^2} \right) \\
&= 16 \cdot \frac{p^2(11p^{20} - 10p^{22})}{(1-p^2)^2} < 0.89,
\end{aligned}$$

where the last inequality holds for all $p < 0.75$. Hence, by Markov's inequality

$$\Pr[\sum_{\ell=22}^{\delta n} w_\ell = 0] = \Pr[\sum_{\ell=22}^{\delta n} w_\ell < 1] > 1 - 0.89 = 0.11. \tag{18}$$

Combining Eqs. (17) and (18) we get

$$\Pr[\sum_{\ell=1}^{\delta n} w_\ell = 0] > 0.1, \tag{19}$$

assuming that n is sufficiently large.

Then, by Claim 49 and Claim 51 we have

$$\mathbb{E}\left[\sum_{\ell=\delta n}^{n} \frac{w_\ell}{2^\ell}\right] \leq \sum_{\ell=\delta n}^{n} \frac{8\ell(n-\ell)}{n}\left(\frac{f(\ell/n)}{2^{\ell/n}}\right)^n < 2n^2 \cdot 0.95^n.$$

Hence, by Markov's inequality, we have

$$\Pr\left[\sum_{\ell=\delta n}^{n} \frac{w_\ell}{2^\ell} \geq 0.96^n\right] \leq \frac{2n^2 \cdot 0.95^n}{0.96^n} < 0.02, \tag{20}$$

assuming that n is sufficiently large. By combining Eqs. (19) and (20) we get that $\sum_{\ell=1}^{n} \frac{w_\ell}{2^\ell} < 0.96^n$ with probability at least 0.08. This completes the proof of Lemma 47. □

We now return to proving Claim 50 and Claim 51.

Proof of Claim 50. For $0 < \lambda \leq \delta$ define $s(\lambda) = \frac{2\log_2(\lambda)}{3}$. Then

$$f(\lambda) = \frac{(1+6\cdot 2^{2s}+2^{4s})^{3/4}}{2^{3\lambda s}\cdot 2^{2h(\lambda)}} = \frac{(1+6\cdot\lambda^{4/3}+\lambda^{8/3})^{3/4}}{2^{2\lambda\log_2(\lambda)}\cdot 2^{2h(\lambda)}} \leq \frac{1+1.5\lambda}{2^{2(1-\lambda)\log_2\left(\frac{1}{1-\lambda}\right)}},$$

where last inequality holds for all $\lambda \leq \delta$, which is easy to verify by comparing the polynomials in the denominators.

Next, we let $F(\lambda) = \frac{1+1.5\lambda}{2^{2(1-\lambda)\log_2\left(\frac{1}{1-\lambda}\right)}}$, and show that

$$F(\lambda) < 0.75^\lambda$$

for $\lambda \in [0, \delta]$. Letting $G(\lambda) = 0.75^\lambda$, we show below that $F(0) = G(0)$ and $F'(0) < G'(0)$.

Indeed $F(0) = 1 = G(0)$. We show below that $F'(0) = -0.5 < -0.29 < \ln(0.75) = G'(0)$. Indeed,

$$\begin{aligned} F'(\lambda) &= \frac{1.5 - (1+1.5\lambda)\times\ln(2)\left(2\log_2(\lambda) + \frac{2}{\ln(2)} - 2\log_2(\frac{\lambda}{1-\lambda})\right)}{2^{2(1-\lambda)\log_2\left(\frac{1}{1-\lambda}\right)}} \\ &= \frac{1.5 - (1+1.5\lambda)\times\left(2\ln(\lambda) + 2 - 2\ln(\frac{\lambda}{1-\lambda})\right)}{2^{2(1-\lambda)\log_2\left(\frac{1}{1-\lambda}\right)}} \\ &= \frac{1.5 - (1+1.5\lambda)\times(2+2\ln(1-\lambda))}{2^{2(1-\lambda)\log_2\left(\frac{1}{1-\lambda}\right)}}, \end{aligned}$$

and hence $F'(0) = -0.5$. For the derivative of G, we have $G'(0) = \ln(0.75) \cdot (0.75^\lambda)|_{\lambda=0} = \ln(0.75) > -0.29$.

We have $F(0) = G(0)$ and $F'(0) < G'(0)$, and hence by continuity, $F(\lambda) < 0.75^\lambda$ in some small neighborhood of 0. Numerical calculations show that $F'(\lambda) < G'(\lambda)$ for $\lambda \in (0, 0.2]$. In particular, $f(\lambda) \leq F(\lambda) < 0.75^\lambda$ for all $\lambda \in (0, \delta]$, as required. □

Proof of Claim 51. We break the proof into two painful (but tolerable) cases depending on the value of $\lambda \in [\delta, 1]$.

- For $\lambda \in [\delta, 0.6]$ let $s(\lambda) = \frac{2\log_2(\lambda)}{3}$. Then

$$\frac{f(\lambda)}{2^\lambda} = \frac{(1 + 6 \cdot 2^{2s} + 2^{4s})^{3/4}}{2^{3s\lambda} \cdot 2^{2h(\lambda)+\lambda}} = \frac{(1 + 6 \cdot \lambda^{4/3} + \lambda^{8/3})^{3/4}}{2^{2\lambda\log_2(\lambda)} \cdot 2^{2h(\lambda)+\lambda}} = \frac{(1 + 6 \cdot \lambda^{4/3} + \lambda^{8/3})^{3/4}}{2^{2(1-\lambda)\log_2\left(\frac{1}{1-\lambda}\right)+\lambda}}.$$

  By computing the derivative of $\frac{f(\lambda)}{2^\lambda}$, it is straightforward to check that the function has a unique minimum in the interval $[\delta, 0.6]$ and obtains its maximum at the boundaries of the interval. Verifying that $F(\delta) < 0.95$ and $F(0.6) < 0.95$, it follows that $f(\lambda) \leq F(\lambda) < 0.95$ for all $\lambda \in [\delta, 0.6]$.

- For $\lambda \in [0.6, 1]$ let $s(\lambda) = \frac{2\log_2\left(\frac{1}{1-0.75\lambda}\right)}{3}$. Then

$$\begin{aligned}\frac{f(\lambda)}{2^\lambda} &= \frac{(1 + 6 \cdot 2^{2s} + 2^{4s})^{3/4}}{2^{3s\lambda} \cdot 2^{2h(\lambda)+\lambda}} \\ &= \frac{\left(1 + 6 \cdot \left(\frac{1}{1-0.75\lambda}\right)^{4/3} + \left(\frac{1}{1-0.75\lambda}\right)^{8/3}\right)^{3/4}}{2^{2\lambda\log_2\left(\frac{1}{1-0.75\lambda}\right)} \cdot 2^{2h(\lambda)+\lambda}} \\ &\leq \frac{1 + 5\left(\frac{1}{1-0.75\lambda}\right)^{1.2}}{2^{2\lambda\log_2\left(\frac{1}{1-0.75\lambda}\right)} \cdot 2^{2h(\lambda)+\lambda}}.\end{aligned}$$

  The last inequality can be verified by letting $y = \frac{1}{(1-0.75\lambda)^{1/3}} \in [1.2, 1.6]$, and checking that $(1 + 6y^4 + y^8)^{3/4} \leq 1 + 5y^{3.6}$.

  Denote $F(\lambda) = \frac{1+5\left(\frac{1}{1-0.75\lambda}\right)^{1.2}}{2^{2\lambda\log_2\left(\frac{1}{1-0.75\lambda}\right)} \cdot 2^{2h(\lambda)+\lambda}}$. Then, by computing the derivative of F, it is not difficult to check that F has a unique minimum in the interval $[0.6, 1]$ and obtains its maximum at the boundaries of the interval. Verifying that $F(0.6) < 0.95$ and $F(1) < 0.86$, it follows that $\frac{f(\lambda)}{2^\lambda} \leq F(\lambda) < 0.95$ for all $\lambda \in [0.6, 1]$.

This completes the proof of Claim 51. □

C Proof of Lemma 17

We start with stating the fact that a random matrix over a large enough finite field with $O(\log n)$ non-zeros in each row is full rank with high probability.

Claim 53. *Let $\mathbb{F}$ be a finite field of size n^C. Let M be a random $n \times n$ matrix obtained by picking $t := C \log n$ random positions (with repetition) in each row and set them to uniform random. Then M is full rank with probability $1 - 1/n^{C/2}$.*

Proof. The only difference from [HLY+24, Lemma 3.3] is the success probability. One can verify this can be improved to $1/n^{C/2}$ by replacing the field size $|\mathbb{F}|$ from $2n$ to n^C and t from $10 \log n$ to $C \log n$. □

Proof of Lemma 17. We first construct the block matrix

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_1 & 0 \\ 0 & I_{M_{\text{fixed}}} \end{bmatrix},$$

where $\boldsymbol{G}_1$ is a $U \times r$ random matrix obtained by picking $t := C \log n$ random positions (with repetition) in each row and set them to uniform. By Claim 53, we have

$$\Pr_{\boldsymbol{G}} \Pr_{\boldsymbol{S}} \left[\boldsymbol{G}_{\boldsymbol{S} \cup \{U+1,\ldots,U+M_{\text{fixed}}\}} \text{ is full rank} \right] = \Pr_{\boldsymbol{S}} \Pr_{\boldsymbol{G}} \left[\boldsymbol{G}_{\boldsymbol{S} \cup \{U+1,\ldots,U+M_{\text{fixed}}\}} \text{ is full rank} \right] \geq 1 - 1/n^C.$$

So we can fix a choice of G with the desired property. The lemma follows from sparsifying G using elementary operations as in [HLY+24, Section 3]. □

D Deterministic sampler in the cell-probe model

In this section, we provide a sketch of a deterministic construction of the random matrix in Lemma 17 in the cell-probe model. Our construction uses $O(n^{2/3+0.02})$ random bits, and thus increases the redundancy in Theorem 2 to this amount.

In the proof of Lemma 17, we generate a random $U \times n$ matrix $\mathbf{A}$, such that:

1. Fixing a set of n rows of $\mathbf{A}$, with high probability in n, these rows are linearly independent;
2. Every row of $\mathbf{A}$ has at most $C \log n$ nonzero elements.

Instead of generating this matrix directly, which costs more than $\Omega(U)$ random field elements, we do the following to generate a random matrix $\mathbf{B}$ with a similar guarantee.

We first create $n^{1/3}$ buckets, and use a $O(1)$-wise independent hash function to map each row $i \in [U]$ to a random bucket. With high probability in n, each bucket contains at most $n^{2/3} \cdot n^{1/3+0.01}$ valid rows. Note that to simulate such a hash function, we only need access to $O(1)$ random field elements.

We let the matrix $\mathbf{B}$ have $n^{1/3} \cdot (n^{2/3} + n^{1/3+0.01}) = n + n^{2/3+0.01}$ columns, which is slightly more than n columns. These columns are partitioned into groups each of which has $n^{2/3} + n^{1/3+0.01}$ columns. Each bucket occupies a group of columns. If row i is hashed to bucket j, then it can only have nonzero entries in the j-th column group. This structure will cause $n^{2/3+0.01+o(1)}$ of redundancy.

Next, we sample an $\widetilde{O}(n^{2/3})$-wise independent hash function $\mathbf{h}$, which maps every row i to the positions and values of the $O(\log n)$ nonzero entries in the column group it hashes to. This part

is similar to the original construction. Again, to simulate this hash function, we need access to a sequence of random cells of length $\widetilde{O}(n^{2/3})$. We take part of the input bits to do this.

Fixing a set of n valid rows, the submatrix of $\mathbf{B}$ formed by these rows is a block matrix, where the j-th block consists of all valid rows hashed to the j-th bucket and all columns in the j-th group. We only need to prove that each of these blocks has full row-rank with high probability. In each group j, there are at most $n^{2/3} + n^{1/3+0.01}$ such rows, which is less than the independence of the hash function $\mathbf{h}$. Therefore, the positions and values of the nonzero entries in this block are fully independent from each other. As such, the original analysis of the matrix $\mathbf{A}$ applies, implying that this block has full row-rank with high probability.

This construction has redundancy $n^{2/3+0.02}$ bits, where the 0.02 can be replaced with an arbitrary small constant, which comes from two sources: one is that we need to spend $n^{2/3+0.01+o(1)}$ cells from the input tape to simulate a high-independence hash function. The other is that the matrix $\mathbf{B}$ now use more than n columns, which causes some waste.

After the matrix $\mathbf{B}$ is constructed, we can do the same analysis as before, using elementary operations to sparsify $\mathbf{B}$ as in [HLY+24, Section 3].

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il