Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 164 (2025)

Local Samplers for Product Distributions

Jordan Horacsek*
Simon Fraser University

Chin Ho Lee†
North Carolina State University

Igor Shinkar‡
Simon Fraser University

Emanuele Viola§
Northeastern University

Renfei Zhou¶
CMU

Abstract

We obtain several results on sampling product distributions in a local and randomness-efficient fashion:

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 rely on pseudorandom distributions that are bounded uniform on average. These distributions are obtained using various tools from low-density parity-check codes, and recent results on succinct and retrieval data structures by Hu, Liang, Yu, Zhang, and Zhou (STOC 2025).

*jordan_horacsek@sfu.ca
†chinho.lee@ncsu.edu
‡ishinkar@sfu.ca
§mathematicsoftheimpossible@gmail.com. Supported by NSF grant CCF-2430026.
¶renfeiz@andrew.cmu.edu. Supported in part by Jane Street Graduate Research Fellowship and MongoDB PhD Fellowship.

ISSN 1433-8092

1 Introduction

Shannon's source coding theorem [Sha48], see for example [CT91, Theorem 3.2.1], says that n i.i.d. samples from a source $\mathcal{D}$ can be compressed into about $nH(\mathcal{D})$ bits, from which the samples can be decoded with high probability. In Shannon's original result, the decoder is not explicit; efficient source decoding algorithms were developed in subsequent works (see, for example, [MCW15] and the references therein).

However, source coding places no requirement on the statistical distance between the decoder's output distribution and the source distribution. Later, Knuth and Yao [KY76] initiated the study of "source simulation," more commonly known as *sampling*. Formally, given a distribution $\mathcal{D}$, say supported on $\{0,1\}^w$, the goal is to design a mapping $f\colon \{0,1\}^m \to \{0,1\}^w$ such that for a uniformly random $\boldsymbol{x} \in \{0,1\}^m$, the distribution of $f(\boldsymbol{x})$ is equal (or close) to $\mathcal{D}$. Knuth and Yao proved optimal bounds on the *expected* number of uniform bits required for this task.

By a simple concentration argument, one can extend their result to construct samplers for n i.i.d. samples from any distribution $\mathcal{D}$ with worst-case randomness complexity close to the information-theoretic optimum $nH(\mathcal{D})$. The sampling question has since been studied extensively in both information theory and computer science theory, leading to time- and space-efficient samplers in various settings. We refer readers to [Li24, Chapter 9] for a brief survey, and to [DS26] for a summary of more recent developments.

However, achieving randomness complexity close to $nH(\mathcal{D})$ using samplers that run in *constant time* per output symbol has proved elusive. Somewhat surprisingly, we show that this is possible in various settings. For example, Theorem 1 establishes such a result in the bit-probe (a.k.a. NC^0) model for any *dyadic* source, a restriction that prior work has shown to be necessary.

More broadly, a main motivation for this paper arises from the study of the *complexity of distributions*. We now elaborate on this perspective and then present our results.

Recent years have witnessed substantial progress on understanding the complexity of sampling distributions, motivated in part by earlier work showing that *sampling* can be easier than *computing*. In contrast to the standard task of computing a function, sampling does not require f to output any specific value on a given input x; rather, the focus is solely on the distribution induced by $f(\boldsymbol{x})$ when $\boldsymbol{x}$ is chosen uniformly at random.

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 following 2-local function on $n-1$ bits:

$$(x_1, \ldots, x_{n-1}) \mapsto (x_1, x_1 \oplus x_2, x_2 \oplus x_3, \ldots, x_{n-2} \oplus x_{n-1}, x_{n-1}).$$

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 using restricted computational models and established several 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; KOW24; KOW25a;

KOW25b], 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; AGMRS26], low-distortion embeddings [BCS16; BS23], quantum and classical separation [WP26; GKMOW26], 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 blogpost [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 $\mathrm{Ber}(p)^n$, 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; KOW24; KOW25a]. 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 [Reg09] 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 [HH24] and the works [DILV24a; DILV24b]. Again, efficient implementations of these 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, the result by Knuth and Yao [KY76] implies that any p-biased distributions on n bits can be sampled with randomness complexity $h(p)n + o_n(1)$, where $h(p) \coloneqq p\log_2(1/p) + (1-p)\log_2(1/(1-p))$ is the binary entropy function, which is best possible. However, their result and follow-up works do not take locality of sampling algorithm into account. It is natural to ask whether one can simultaneously achieve small locality and low randomness complexity. This question was explicitly posed in a blog post [Vio14b] nearly a decade ago; yet, to the best of our knowledge, the tradeoffs involved remain poorly understood. 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. Somewhat surprisingly, we show that with constant locality we can sample *any* product distribution (in particular, $\mathrm{Ber}(1/4)^n$) with nearly optimal randomness complexity. This result requires the distribution to be *dyadic*, i.e., all probability masses are of the form $a/2^b$ for integers a, b. The dyadic requirement is necessary: for example, $\mathrm{Ber}(1/3)^n$ cannot be sampled locally, even approximately. This follows from the techniques in [Vio23], though the result there is stated for the Hamming slice; alternatively, see [KOW24, Theorem 1.10]. Thus, our results illustrate a stark contrast between sampling $\mathrm{Ber}(p)^n$ for dyadic and non-dyadic values of p. Henceforth, we denote statistical distance by dist.

Theorem 1 (Special case of Theorem 8)**.** *Let $D = (D_1, D_2, \ldots, D_n)$ be a product distribution where each D_i is dyadic and supported on $\{0,1\}^w$. For every $\epsilon > 0$, there is a $O_{w,\epsilon}(1)$-local f with input length $H(D) + \epsilon wn$ such that $\mathrm{dist}(f(U), D) \le e^{-\Omega_{w,\epsilon}(n)}$.*

By increasing the locality to $O(\log n)$, we can approximate any distribution by a dyadic one and sample any *arbitrary* product distribution to within distance $1/\mathrm{poly}(n)$ (see Corollary 9).

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

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.*

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 $1/\mathrm{poly}(n)$ 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 20 and Theorem 28)**.** *For $\epsilon > 0$ and $f\colon \{0,1\}^{(2-\epsilon)n} \to \{0,1\}^n$ be any d-local function. We have*

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

Theorem 4 (Theorem 24 and Theorem 31)**.** *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*

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

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 11) 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.

Broadly speaking, we obtain Theorems 1 and 2 by concatenating independent blocks of local samplers with *variable* input lengths that are close to optimal *on average*, followed by sampling their inputs locally and randomness-efficiently. The latter relies on sampling distributions that are bounded-uniform in an *average-case* sense. Henceforth, we call a sampler for each block a *block-sampler.*

Overview of Theorem 1. A building block of our construction is a (possibly inefficient) block-sampler of $\mathrm{Ber}(1/4)^b$ with *expected* randomness complexity close to the optimal $h(1/4)b$. Such construction dates back to the work of Knuth and Yao [KY76]. 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 realized using a prefix-free encoding so that the expected number of bits read is close to optimal.

Given such a block-sampler, we divide the n output bits in blocks of length $b = O_\epsilon(1)$, and consider sampling each block with an independent copy of the block-sampler. By concentration inequalities, with high probability over the randomness of the input bits, the actual number of random bits used to sample a typical output is close to optimal.

We next *derandomize* this construction. To do so, we sample the inputs to the block-samplers pseudorandomly via a local linear transformation. Specifically, we take a nearly optimal number of uniform bits and multiply them by a sparse matrix that expands them into input bits of the block-samplers.

The key property we need from the matrix is that *most* small subsets of its rows, corresponding to the coordinates read by the block-samplers, are linearly independent. This condition is strictly *weaker* than bounded uniformity, which requires *every* small subset of rows to be linearly independent. Indeed, the Plotkin bound implies that no matrix satisfying the stronger requirement can achieve optimal seed length. So exploiting this weaker condition is crucial in our construction.

Although the required matrix property seems relatively basic, we are not aware of any result in the literature that can be applied directly. So, we give a self-contained analysis showing that a suitable random construction satisfies this property with high probability.

Overview 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 below) and used in a number of following works, our connection is different. The work [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 [HLYZZ25], in particular the use of *retrieval data structures*.

To explain we begin with a key concept, originating in [Pat08] (see also [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

$$1 - \frac{\binom{U}{n}}{2^M \cdot K} \le \log_2\left(\frac{2^M \cdot K}{\binom{U}{n}}\right) \le \epsilon. \qquad \square$$

We divide the n bits into blocks of $B = \mathrm{polylog}(n)$ bits, as opposed to $O_\epsilon(1)$ bits in Theorem 1. To sample a block with constant word-probes, we 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 $1/\mathrm{poly}(n)$ small.

A critical point here is that to sample $\mathrm{Ber}(1/4)^B$, the weight s is not fixed, but a random variable distributed according to the binomial distribution $\mathrm{Bin}(B, 1/4)$. Consequently, both M and K are random variables induced by s.

To sample $\mathrm{Ber}(1/4)^B$, as in [HLYZZ25] we encode a distribution $\mathcal{B}$ that is close to $\mathrm{Bin}(B, 1/4)$ 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 $\boldsymbol{s} \sim \mathcal{B}$, followed by sampling a uniform string of Hamming weight $\boldsymbol{s}$ using the 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* over $\boldsymbol{s} \sim \mathcal{B}$. Now we can apply Claim 6 to obtain a $O(1)$-word-probe block-sampler for $\mathrm{Ber}(1/4)^B$.

Our plan is to concatenate the $L := n/B$ independent copies of the block-sampler to sample the n bits. However, as the size of a representation depends on $\boldsymbol{s} \sim \mathcal{B}$, 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 $M^{(\boldsymbol{s}_i)}$ and $K^{(\boldsymbol{s}_i)}$ are both 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 [HLYZZ25] 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* $M_{\min} := \log \binom{B}{B/4-B^{2/3}} = h(1/4)B - \Theta(B^{1/3})$, which is much larger than its deviation $\Delta_{\max} := \log_2 \binom{B}{B/4} - M_{\min} = O(B^{1/3})$. Based on this observation, [HLYZZ25] 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 for the block-samplers. However, unlike [HLYZZ25], our construction does not achieve $\mathrm{polylog}(n)$ redundancy, because in contrast to the data structure setting, a (local) sampler cannot first sample the sizes $M^{(\boldsymbol{s})}$ (and $K^{(\boldsymbol{s})}$) for the L block-samplers and then decide which portions of input bits are read by each block-sampler.

Also, the data structure in [HLYZZ25] 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 15).

We refer the readers to Section 3.3 for more details.

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 28) 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 [HH24]. Recall that the output of the trivial 2-local sampler is the bitwise AND $x \wedge_n y$ for two independent uniform n-bit strings x and y. The work [HLV18] suggested the following equivalent view of $x \wedge_n y$.

Claim 7. *Let $\boldsymbol{x}, \boldsymbol{y} \sim \{0,1\}^n$ be independent, where $\boldsymbol{x}$ is uniform. Then*

$$\mathrm{dist}\big(\boldsymbol{x} \wedge_n \boldsymbol{y}, \mathrm{Ber}(1/4)^n\big) \le \Pr_{\boldsymbol{S} \subseteq [n]}\big[\boldsymbol{y}_{\boldsymbol{S}} \text{ is not uniform on } \{0,1\}^{\boldsymbol{S}}\big].$$

Proof. We can think of sampling $\mathrm{Ber}(1/4)^n$ by first sampling a uniform $\boldsymbol{x}$ follows by sampling an independent uniform string on the subset $\boldsymbol{S}$ of the 1-coordinates of $\boldsymbol{x}$. □

Therefore, to get close to $\mathrm{Ber}(1/4)^n$, it suffices for $\boldsymbol{y}$ to be uniform on a uniform subset $\boldsymbol{S}$ of its n coordinates *with high probability*. To generate such $\boldsymbol{y}$, our 3-local sampler assigns 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 be $z_i \oplus z_i'$. To analyze the construction, we show that a random subgraph of G has no cycle with probability $1 - 1/\mathrm{poly}(n)$. 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 uniform bits to select a random subset $\boldsymbol{S} \subseteq [n]$. Then to sample $\boldsymbol{y}$, we use a 3-local LDPC code by Gallager [Gal62] 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 8. *Let q be an integer and $D_1, \dots, D_n$ be n distributions on $\{0,1\}^w$, where the probability $D_i(s)$ is an integer multiple of 2^{-q} for every $i \in [n]$ and $s \in \{0,1\}^w$. Let $D = D_1 \times \cdots \times D_n$ be the product distribution of the D_i's.*

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

The following is an almost immediate corollary from Theorem 8.

Corollary 9. *Let $D_1, \dots, D_n$ be n distributions on $\{0,1\}^w$ for $w = O(\log n)$. The product distribution $D := D_1 \times \cdots \times D_n$ can be sampled using $H(D) + O(n \log n)$ bits with locality $O(\log n)$ and error $1/\mathrm{poly}(n)$.*

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 $\mathrm{dist}(D_i, D'_i) \le \gamma/n$ 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 $D' \coloneqq D'_1 \times \cdots \times D'_n$ are $1/\mathrm{poly}(n)$-close in total variation distance. Note that for each $i \in [n]$ we have

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

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

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

Therefore, $H(D') \le H(D) + 1/\mathrm{poly}(n)$, and the corollary follows by applying Theorem 8 on D'. □

We now turn to the proof of Theorem 8. We first apply the following lemma, which is a special case of [KY76], which gives optimal bounds on the average-case randomness complexity of sampling an arbitrary distribution.

Lemma 10 ([KY76])**.** *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 exactly by a decision tree $f\colon \{0,1\}^q \to \{0,1\}^w$, where the expected depth of a leaf is at most $H(D)+2$.*

Lemma 10 and its more general statement in [KY76] follows from an application of Kraft's inequality. (A proof can be found in [Li24, Theorem 55].) For the proof of Theorem 8, we also need a sparse matrix with the following properties.

Lemma 11. *Fix $k \in \mathbb{N}$ and a sufficiently small $\alpha > 0$. Let $d \coloneqq \lceil \ln(1/\alpha)/\alpha\rceil$ and $m \ge \max\{\frac{k}{1-2h(\alpha)}, \alpha n\}$. Let $\mathcal{S}$ be any distribution supported on subsets $S \subseteq [n]$ of size 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 of its rows $\boldsymbol{S}$ 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*

$$\Pr_{\boldsymbol{S}\sim\mathcal{S}}\left[M_{\boldsymbol{S}} \text{ is full rank}\right] \ge 1 - 2^{-\Omega(\alpha\log(\frac{1}{\alpha})n)}.$$

Remark 12. In Lemma 11 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 $\mathcal{S}$. We note that with the required parameters we cannot possibly hope for a *binary* matrix where *any* k rows are linearly independent, as such matrix would correspond to a parity-check matrix of a linear error correcting codes with block length m, minimum distance k, and dimension at least $m-k$, which is impossible over small alphabet (e.g., by Plotkin's bound, stating that a linear code of length m with minimum distance k has dimension at most $m - 2k + o(1)$).

The proof of Lemma 11 is below in Section 2.1.

Proof of Theorem 8. Divide the D_i's into n/t blocks each of size $t = \lceil 8/\epsilon \rceil$. Let $D^{(j)}$ be the product of the D_i's in the j-th block. Note that $D^{(j)}$ is supported on $\{0,1\}^{tw}$ and its probability mass on each outcome is an integer multiple of 2^{-qt}. We first apply Lemma 10 to sample each $D^{(j)}$ independently with a qt-local block-sampler using qt bits. Let $f' \colon \{0,1\}^{qn} \to \{0,1\}^{wn}$ be the concatenation of the n/t block-samplers. We will sample the qn-bit input of f' pseudorandomly by applying the sparse matrix from Lemma 11 to a seed of some length m.

For each input $z \in \{0,1\}^{qn}$ of f', let $S_z \subseteq [qn]$ denote the subset of positions read by f' to evaluate $f'(z) \in \{0,1\}^{wn}$. We emphasize that S_z are the only positions read by f' to evaluate $f'(z)$. Note that S_z can be written as $S_z = S_z^1 \sqcup \cdots \sqcup S_z^{n/t}$, where $S_z^j \subseteq \{(j-1)t+1, (j-1)t+2, \ldots, jt\}$ is the subset of bits read by the j-th block-sampler. By Lemma 10, for a uniform $\boldsymbol{z} \sim \{0,1\}^{qn}$, we have $\mathbb{E}[|S_{\boldsymbol{z}}^j|] \le H(D^{(j)}) + 2$ for every $j \in [n/t]$. Therefore,

$$\mathbb{E}_{\boldsymbol{z}}\big[|S_{\boldsymbol{z}}|\big] \le \sum_{i=1}^{n/t} H(D^{(j)}) + 2 \cdot (n/t) \le H(D) + (\epsilon/4)n.$$

Let $k \coloneqq H(D) + (\epsilon/2)n$. By Hoeffding's inequality, we have

$$\Pr\big[|S_{\boldsymbol{z}}| \ge k\big] \ge \Pr\big[|S_{\boldsymbol{z}}| \ge \mathbb{E}_z[|S_z|] + \epsilon n/4\big] \le e^{-\Omega\left(\frac{(\epsilon n)^2}{(n/t)\cdot(qt)^2}\right)} \le e^{-\Omega\left(\frac{\epsilon^3}{q^2}n\right)}.$$

Define $\alpha > 0$ so that $\frac{1}{1-2h(\alpha)} = (1+\epsilon/2)$ (and therefore $\alpha = \Theta(\epsilon/\log(1/\epsilon))$. Let

$$m \coloneqq k \cdot \frac{1}{1-2h(\alpha)} = k \cdot \left(1+\frac{\epsilon}{2}\right) = \left(H(D) + \frac{\epsilon n}{2}\right)\left(1+\frac{\epsilon}{2}\right) \le H(D) + \epsilon qn,$$

where the last inequality follows from $H(D) \le \log_2(|\mathrm{supp}(D)|) \le qn$.

Let M be the $qn \times m$ matrix obtained by applying Lemma 11 with $\mathcal{S}$ equal to the distribution of $S_{\boldsymbol{z}}$ conditioned on $|S_{\boldsymbol{z}}| \le k$ and our choice of α. Our sampler $f \colon \{0,1\}^m \to \{0,1\}^{wn}$ takes an input $x \in \{0,1\}^m$ and output $f'(Mx)$.

Clearly the input length of f is $m = H(D) + \epsilon n$. The locality of f is at most $qt \cdot d = O(\frac{q}{\epsilon^2}\log^2\left(\frac{1}{\epsilon}\right))$, where $d = O(\log(1/\alpha)/\alpha) = O(\log^2(1/\epsilon)/\epsilon)$ is row-sparsity of M given by Lemma 11.

We now analyze the error. By Lemma 11, we have

$$\Pr_{\boldsymbol{z}}\big[M_{S_{\boldsymbol{z}}} \text{ is full rank} \mid |S_{\boldsymbol{z}}| \le k\big] \ge 1 - 2^{-\Omega(\alpha\log(\frac{1}{\alpha})qn)} \ge 1 - 2^{-\Omega(\epsilon qn)}.$$

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

$$\mathrm{dist}\big(f(U_m), D\big) \le \Pr_{\boldsymbol{z}\sim\{0,1\}^{qn}}\big[|S_{\boldsymbol{z}}| \ge k\big] + 2^{-\Omega(\epsilon qn)} \le e^{-\Omega(\frac{\epsilon^3}{q^2}n)},$$

as required. □

2.1 Proof of Lemma 11

We prove the lemma by considering a random matrix $\boldsymbol{M} \in \mathbb{F}_2^{n\times m}$, where each row of $\boldsymbol{M}$ is sampled independently according to the following distribution: Select d indices $\boldsymbol{i}_1, \boldsymbol{i}_2, \ldots, \boldsymbol{i}_d \sim [m]$ 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 $p_\ell \coloneqq \Pr[z\boldsymbol{M} = 0]$, where $z \in \mathbb{F}_2^n$ is an arbitrary vector of Hamming weight ℓ. (Note that p_ℓ only depends on $|z|$ due to the randomness of $\boldsymbol{M}$.) We start with the following expression for the probability that z is a null vector of $\boldsymbol{M}$.

Claim 13. *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 11.

Consider the random matrix $\boldsymbol{M}$ with at most d ones in each row as described above. Denote by L the event that any subset of αm rows of $\boldsymbol{M}$ are linearly dependent. Then, using the assumption about α being sufficiently small, Claim 13 implies that

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

We bound the first summand as follows. Using $\binom{n}{\ell} \le (en/\ell)^\ell$, and that $2x\log_2(1/x)$ is increasing in $[0, 1/e]$,

$$\begin{aligned}\sum_{\ell=1}^{\frac{m}{4d}}\binom{n}{\ell}2\left(\frac{2d\ell}{m}\log\left(\frac{m}{d\ell}\right)\right)^{\frac{d\ell}{2}} &\le 2\sum_{\ell=1}^{\frac{m}{4d}}\left(\frac{en}{\ell}\cdot\left(\frac{2d\ell}{m}\log\left(\frac{m}{d\ell}\right)\right)^{\frac{d}{2}}\right)^{\ell} \\ &\le 2\sum_{\ell=1}^{\frac{m}{4d}}\left(\frac{2d\ell}{\alpha n}\log\left(\frac{\alpha n}{d\ell}\right)\right)^{\frac{d\ell}{4}} \le n^{-\Omega(d)}.\end{aligned}$$

For the second summand, using $h(x) \le 2x\log_2(1/x)$ for $x \in [0, 1/2]$ and $m \ge \alpha n$, we have $h(\frac{\alpha m}{n})n \le 2\alpha m\log(\frac{n}{\alpha m}) \le 4\alpha m\log_2(1/\alpha)$. As $\alpha > 0$ is sufficiently small, we have

$$\sum_{\ell=\frac{m}{4d}+1}^{\alpha m}\binom{n}{\ell}2\cdot 2^{-\frac{m}{4}} \le 2\cdot 2^{h(\frac{\alpha m}{n})n}\cdot 2^{-\frac{m}{4}} \le 2\cdot 2^{-(4\alpha\log_2(\frac{1}{\alpha})-1/4)m} \le 2^{-\Omega(m)} \le 2^{-\Omega(\alpha n)}.$$

Putting the two bounds together gives $\Pr[L] \le 2^{-\Omega(m)} + n^{-\Omega(d)} \le n^{-\Omega(d)}$.

Fix a subset of the rows $S \subseteq [n]$ of size k. Then by Claim 13, using $k \le (1 - 2h(\alpha))m$, and $h(\alpha) \ge \alpha \log(1/\alpha)$,

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

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

$$\begin{aligned}&\Pr_{\boldsymbol{M}}\big[\text{there exists a subset } T \subseteq S \text{ of rows in } \boldsymbol{M} \text{ with } |T| \ge \alpha m \text{ whose sum is } 0 \mid \overline{L}\big] \\ &\quad\le \frac{2^{-\Omega(\alpha \log(\frac{1}{\alpha})n)}}{1 - \Pr[L]} \le 2^{-\Omega(\alpha \log(\frac{1}{\alpha})n)}.\end{aligned}$$

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

$$\Pr_{\boldsymbol{S} \sim \mathcal{S}}\left[M_{\boldsymbol{S}} \text{ is full rank}\right] \ge 1 - 2^{-\Omega(\alpha \log(\frac{1}{\alpha})n)}.$$

This completes the proof of Lemma 11. □

We now return to the proof of Claim 13.

Proof of Claim 13. 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

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

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

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

$$\begin{aligned} p_\ell &\le 2^{-m}\sum_{i=0}^{t}\binom{m}{i}\Big(1-\frac{2i}{m}\Big) + \Big(1-\frac{2t}{m}\Big)^{d\ell} \\ &\le 2^{-(1-h(\frac{t}{m}))m} + h\Big(\frac{d\ell}{m}\Big)^{\frac{d\ell}{2}}. \end{aligned}$$

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 $h(t/m) = h(1/2-\sqrt{h(d\ell/m)}/2) \le 1 - \frac{h(d\ell/m)}{2}$. So,

$$2^{-(1-h(\frac{t}{m}))m} \le 2^{-\frac{1}{2}h(\frac{d\ell}{m})m} \le \frac{1}{\binom{m}{d\ell}^{1/2}} \le \Big(\frac{d\ell}{m}\Big)^{\frac{d\ell}{2}}.$$

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,

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

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

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

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

$$\binom{m}{i}\Big(1-\frac{2i}{m}\Big)^{d\ell} \le 2^{h(\alpha)m}.$$

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

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

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,

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

This completes the proof of Claim 13. □

3 Sampling p-biased distributions from static dictionary

In this section we prove Theorem 2, restated below.

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.*

In Section 3.1, we establish several claims that will be used in our construction. In Section 3.2, we show how to construct a local block-sampler with near-optimal average-case randomness complexity. In Section 3.3, we show how to concatenate these block-samplers in a local fashion.

3.1 Changing bases

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

Claim 14. *The uniform distribution over* $[K]$ *can be sampled by* m *elements in* $[q]$ *with error* K/q^m.

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

Lemma 15. *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 14 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 16. *Let $B \coloneqq \mathrm{polylog}(n)$ and $C > 0$ be any constant. The distribution $\mathrm{Ber}(p)^B$ can be sampled 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 distribution $\mathcal{B}$ supported on $[pB - B^{2/3}, pB + B^{2/3}]$ that is $n^{-\Omega(C)}$-close to $\mathrm{Bin}(B,p)$ can be sampled using the first $t \coloneqq C \log_2 n$ bits of $\{0,1\}^{\boldsymbol{M}}$.*
- *Given a sample $\boldsymbol{s} \sim \mathcal{B}$, the lengths $M^{\boldsymbol{s}} \coloneqq \boldsymbol{M}$ and $K^{(\boldsymbol{s})} \coloneqq \boldsymbol{K}$ are fixed and*

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

- *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 16 follows [HLYZZ25], where we encode the distribution of $\boldsymbol{s}$ into the succinct data structure in [Yu22] with a small increase in redundancy.

Lemma 17 (Lemma 28 in [Yu22])**.** *Let $B \coloneqq \mathrm{polylog}(n)$, 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 16. As in [HLYZZ25, Lemma 4.2], we first instantiate Lemma 17 to represent a size-s subset of $[B]$ with a spillover representation $(m', k') \in \{0,1\}^{M'} \times [K']$. Then we encode a distribution $\mathcal{B}'$ that is $n^{-\Omega(C)}$-close to $\mathrm{Bin}(B,p)$ into the first $t \coloneqq \Theta(\log n)$ bits of m' as follows.

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

We now encode $s \in \mathrm{supp}(\mathcal{B})$ into the first t bits of m'. We first 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

$$m'_0 = \left\lfloor \frac{m'_0}{|T_s|} \right\rfloor \cdot |T_s| + m'_0 \bmod |T_s|.$$

Now, we first 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 next t bits of m', and encode $\left\lfloor \frac{m'_0}{|T_s|} \right\rfloor$ along with k' as the spill $k \in [K]$, where $K \coloneqq K' \cdot \lceil \frac{2^{2t}}{|T_s|} \rceil \le K' \cdot \frac{1}{\mathcal{B}(s)} \cdot \mathrm{poly}(n)$. A similar calculation as in [HLYZZ25] shows that the redundancy increases by $1/\mathrm{poly}(n)$.

We now explain how to (approximately) sample $\mathrm{Ber}(p)^B$. We first sample $\boldsymbol{s} \sim \mathcal{B}$ using t uniform bits. Given $\boldsymbol{s}$, we can determine the sizes $(M'^{(\boldsymbol{s})}, K'^{(\boldsymbol{s})})$. We can then sample the rest of $(\boldsymbol{m}', \boldsymbol{k}')$ from $\{0,1\}^{M'(\boldsymbol{s})-t} \times K^{(\boldsymbol{s})}$. This lets us recover $(\boldsymbol{m}', \boldsymbol{k}')$, which is then used to sample a uniform string of size $\boldsymbol{s}$.

Let $\mathcal{D}$ be the resulting distribution. By Lemma 17, we have

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

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

3.3 Concatenation

In this subsection, we explain how to sample the inputs of the $L = n/B$ copies of the block-sampler in Theorem 16. Specifically, we will sample the L spill representations $(\boldsymbol{m}_i, \boldsymbol{k}_i) \in \{0,1\}^{M(\boldsymbol{s}_i)} \times [K^{(\boldsymbol{s}_i)}]$ in a local and randomness-efficient way.

To illustrate the conceptual idea, let us for simplicity consider sampling only the $\boldsymbol{m}_i$'s but not the $\boldsymbol{k}_i$, using $L\,\mathbb{E}[M^{(\boldsymbol{s})}] + \widetilde{O}(\sqrt{n})$ bits.

Recall that in Theorem 16, $\mathcal{B}$ is a distribution that approximates $\mathrm{Bin}(B,p)$ and is supported on $[pB - B^{2/3}, pB + B^{2/3}]$. For every fixed s in $\mathrm{supp}(\mathcal{B})$, the sampler will sample from the uniform distribution on $\{0,1\}^{M^{(s)}} \times [K^{(s)}]$ a spill representation of a size-s subset in $[B]$. Without loss of generality, we will assume $M^{(s)}$ is an integer multiple of the word size $w = O(\log n)$ and $K^{(s)} \ge n^C$ for a large enough C. This can be achieved by moving $O(w)$ bits in $M^{(s)}$ to the spill, which can only change the spill size $K^{(s)}$ by a factor of at most $2^{O(w)} \le \mathrm{poly}(n)$.

We now make some observations about $M^{(s)}$ for $s \in \mathrm{supp}(\mathcal{B})$. Henceforth, we will treat bits as words, and view $\{0,1\}^{M^{(s)}}$ as $W^{(s)} \coloneqq \frac{M^{(s)}}{w}$ $\mathbb{F}_{2^w}$-elements. Note that we have $W^{(s)} \in [W_{\min}, W_{\max}]$, where

$$W_{\min} \coloneqq \frac{1}{w}\log_2 \binom{B}{pB - B^{2/3}} = h(p)(B/w) - \Theta(B^{1/3}/w) \text{ and}$$
$$W_{\max} \coloneqq \frac{1}{w}\log_2 \binom{B}{pB} \le h(p)(B/w).$$

Thus, we can write $W^{(s)} = W_{\min} + \Delta^{(s)}$, where $W_{\min}$ does not depend on s and $\Delta^{(s)} \le \Delta_{\max} \coloneqq W_{\max} - W_{\min} = \Theta(B^{1/3})$. Therefore, we can make the following conclusion on the concatenation of the L samplers:

- it always reads a *fixed* set of $\overline{W}_{\min} \coloneqq LW_{\min}$ coordinates;

- it reads at most $\overline{\Delta}_{\text{thr}} := \sqrt{L \log n}$ additional coordinates with probability $1 - 1/\text{poly}(n)$;
- it reads at most $\overline{\Delta}_{\max} := L\Delta_{\max}$ additional coordinates in the worst-case.

Moreover, the maximum deviation $\overline{\Delta}_{\max}$ is much smaller than $\overline{W}_{\min}$. Specifically, we have $\overline{\Delta}_{\max} \le \overline{W}_{\min}/\log n$. Based on this observation, we can apply the following sparse matrix used in the augmented retrieval data structure in [HLYZZ25] to sample the $W^{(\boldsymbol{s}_1)} + \cdots + W^{(\boldsymbol{s}_L)}$ input elements to the L block-samplers.

Lemma 18. *Let $\mathbb{F}$ be a finite field of size at least n^C. Let $\boldsymbol{S} \subseteq [\overline{\Delta}_{\max}]$ be a random subset of size at most $\overline{\Delta}_{\text{thr}}$. Suppose $\overline{W}_{\min} \ge \overline{\Delta}_{\max} \log n$. Then there exists a $(\overline{W}_{\min} + \overline{\Delta}_{\max}) \times (\overline{W}_{\min} + \overline{\Delta}_{\text{thr}})$ matrix G over $\mathbb{F}$ with $O(1)$ nonzeros in every row such that*

$$\Pr_{\boldsymbol{S}}\left[G_{[\overline{W}_{\min}]\times \boldsymbol{S}} \text{ is full rank}\right] \ge 1 - \frac{1}{n^{C/2}}.$$

The parameters in our lemma are slightly different from the one in [HLYZZ25], so, we give a proof sketch in Appendix C.

Here the random subset $\boldsymbol{S}$ in Lemma 18 corresponds to the random subset of the $LW_{\max}$ coordinates read by all L samplers modulo their first $W_{\min}$ elements. By concentration bounds, $\boldsymbol{S}$ has size at most $\sqrt{L \log n}$ with high probability. It follows from Lemma 18 that the number of uniform words used by the sampler is $\overline{W}_{\min} + \overline{\Delta}_{\max} \le L\,\mathbb{E}[W^{(\boldsymbol{s})}] + \widetilde{O}(\sqrt{n})$, as desired.

We now briefly discuss how to sample the L spills using the same idea. First, we will assume that $K^{(s)}$ is prime and treat $[K^{(s)}]$ as a field, by embedding $[K^{(s)}]$ into the closest prime field, which has little effect on the error and seed length (see Claim 14).

Note that the $K^{(s_i)}$'s depend on s_i's and therefore are not all identical. Nevertheless, as $\mathcal{B}$ is supported on $[pB - B^{2/3}, pB + B^{2/3}]$, there are at most $2B^{2/3} + 1 = \text{polylog}(n)$ many possible values for $K^{(s)}$. Let us denote these values by $K_1, \ldots, K_q$ for some $q \le 2B^{2/3} + 1$, and define $p_j := \Pr_{\boldsymbol{s}}[K^{(\boldsymbol{s})} = K_j] = \sum_{s:K^{(s)}=K_j} \mathcal{B}(s)$. Let $N_j = N_j(s_1, \ldots, s_L)$ denote the number of $K^{(s_i)}$'s equal to K_j. Over the random choice of $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L \sim \mathcal{B}$, the random variable $\boldsymbol{N}_j$ is distributed close to $\text{Bin}(L, p_j)$, and thus is at most $p_j L + O(\sqrt{L \log n})$ with probability at least $1 - 1/\text{poly}(n)$.

One complication is that $p_j L$ can be smaller than $\sqrt{L \log n}$. So, in order to apply Lemma 18, we will sample the $\boldsymbol{N}_j$ $\mathbb{F}_{K^j}$-elements together with a fraction of the $\overline{W}_{\min}$ $\mathbb{F}_{2^w}$-elements. (We will convert the $\mathbb{F}_{2^w}$-elements to $\mathbb{F}_{K_j}$-elements via Lemma 15.)

3.3.1 Proof of Theorem 2

Our goal is to sample random elements

$$\boldsymbol{w} := (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L) \in (\mathbb{F}_{2^w}^{W_{\max}})^L \quad \text{and} \quad \boldsymbol{k}_j \in [K_j]^L : j \in [q],$$

so that with probability $1 - 1/\text{poly}(n)$ over $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L) \sim \mathcal{B}^L$, each $\boldsymbol{w}_i$ is $1/\text{poly}(n)$-close to uniform on the first $W^{(\boldsymbol{s}_i)}$ elements, and $\boldsymbol{k}_j$ is $1/\text{poly}(n)$-close to uniform on the first $\boldsymbol{N}_j$ elements, using $h(p)n + \sqrt{n} \cdot \text{polylog}(n)$ uniform bits. Moreover, each output element depends on $O(1)$ words of size $w = O(\log n)$ bits.

We first partition these elements into $q+1$ parts and sample each part individually using Lemma 18. For each block $i \in [L]$, let I_i be the first $W_{\min} = \Theta(B)$ positions of w_i, and $\overline{I_i}$ its remaining $\Delta_{\max} = W_{\max} - W_{\min} = \Theta(B^{1/3})$ positions. We then partition their union $\bigcup_{i=1}^{L} I_i$ into $q+1$ sets $J_0 \sqcup J_1 \sqcup \cdots \sqcup J_q$, where $|J_i| := \frac{LW_{\min}}{4q}$ for $i \in [q]$ (and so $|J_0| = \frac{3LW_{\min}}{4}$). Note that the coordinates in J_i's are always read by the sampler, and the rest may not be. We will apply Lemma 18 to sample

1. (Words) the $\mathbb{F}_{2^w}$-elements in J_0 and $\bigcup_{i=1}^{q} \overline{I_i}$.
2. (Spills) $[K_j]^L$ and the $\mathbb{F}_{2^w}$-elements in J_j for each $j \in [q]$.

Let $\overline{W}_{\min} := LW_{\min}$ and $\overline{W}_{\max} := LW_{\max}$.

Sampling the words. Let $\overline{\Delta}_{\text{wd,max}} := L\Delta_{\max}$, $\overline{\Delta}_{\text{wd,thr}} := \sqrt{L \log n}$, and $\Delta^{(s)} := W^{(s)} - W_{\min}$ for every $s \in \text{supp}(\mathcal{B})$. Note that $\mathbb{E}[\sum_{i=1}^{L} \Delta^{(\boldsymbol{s}_i)}] = L\,\mathbb{E}[W^{(\boldsymbol{s})} - W_{\min}] = o(\sqrt{L \log n})$. By Hoeffding's inequality, we have

$$\Pr_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_L}\left[\sum_{i=1}^{L} \Delta^{(\boldsymbol{s}_i)} \le \overline{\Delta}_{\text{wd,thr}}\right] \ge 1 - 1/\text{poly}(n).$$

Conditioned on this event, let $\boldsymbol{S}$ be the subset of coordinates in $\bigcup_{i=1}^{L} \overline{I_i}$ (determined by $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L$) that are read by the sampler (in addition to $\bigcup_{i=1}^{L} I_i$). We have $|\boldsymbol{S}| \le \overline{\Delta}_{\text{wd,thr}}$, and

$$\sum_{i=1}^{q} |\overline{I_i}| \log n = \overline{\Delta}_{\text{wd,max}} \log n \le \tfrac{3}{4}\overline{W}_{\min} = |T_0|.$$

By Lemma 18, there is a $(\frac{3}{4}\overline{W}_{\min} + \overline{\Delta}_{\text{wd,max}}) \times (\frac{3}{4}\overline{W}_{\min} + \overline{\Delta}_{\text{wd,thr}})$ matrix M over $\mathbb{F}_{2^w}$ with $O(1)$ nonzeros in each row such that for a uniform $\boldsymbol{u}$, we have

$$\Pr_{\boldsymbol{S}}\left[M\boldsymbol{u} \text{ is uniform on } \bigcup_{i=1}^{L} I_i \cup \boldsymbol{S}\right] \ge 1 - 1/\text{poly}(n).$$

Note that $W_{\min} \le W^{(s)}]$ for any $s \in \text{supp}(\mathcal{B})$, and thus $\overline{W}_{\min} \le L\,\mathbb{E}[W^{(\boldsymbol{s})}]$. Therefore, the number of uniform bits used to sample this part is at most

$$\left(\frac{3\overline{W}_{\min}}{4} + \overline{\Delta}_{\text{wd,thr}}\right) \cdot w \le \frac{3L\,\mathbb{E}[M^{(\boldsymbol{s})}]}{4} + \sqrt{L}\log^{3/2} n. \tag{2}$$

Sampling the spills. We now explain how to sample the spills and the remaining words. We will use the following result from number theory [BHP01].

Lemma 19. *For every sufficiently large n, there is a prime between n and $n + n^{0.525}$.*

For each $j \in [q]$, let $\Delta_{j,\text{thr}} := p_j L + \sqrt{L \log n}$, and P_j be the smallest prime that is at least K_j, which, by Lemma 19, is at most $K_j + K_j^{0.525}$. The sampling procedure consists of 3 main steps:

1. use Claim 14 and Lemma 15 to sample $[K_j]^L$ and the $\frac{\overline{W}_{\min}}{4q}$ word-elements by a distribution D over a tuple of $\mathbb{F}_{P_j}$-elements.
2. use Lemma 18 to sample the distribution D using uniform $\mathbb{F}_{P_j}$ elements.
3. use Lemma 15 to sample these uniform $\mathbb{F}_{P_j}$ elements using uniform words.

We now describe each step in more detail. First, by Claim 14, for any subset $S \subseteq [L]$, if a distribution $\mathcal{D}$ is ϵ-close to uniform on $\mathbb{F}_{P_j}^S$, then $\mathcal{D}$ is $(\epsilon+\delta)$-close to uniform on $[K_j]^S$, where

$$\delta \le L \cdot \frac{K_j^{0.525}}{K_j + K_j^{0.525}} \le 2 \cdot L \cdot K_j^{-0.475} \le 1/\mathrm{poly}(n). \tag{3}$$

Next, we apply Lemma 15 to obtain a local-sampler mapping $\mathbb{F}_{P_j}^{F_j}$ to $\mathbb{F}_{2^w}^{\overline{W}_{\min}/(4q)}$ with error $1/\mathrm{poly}(n)$, where

$$F_j := \frac{\overline{W}_{\min}}{4q} \cdot \frac{w}{\log_2 P_j} + O(\log_{P_j} n) = \Theta(n/B^{2/3}).$$

Recall that $\boldsymbol{N}_j \sim \mathrm{Bin}(L, p_j)$ with $\mathbb{E}[\boldsymbol{N}_j] = p_j L$ and $\Delta_{j,\mathrm{thr}} = p_j L + \sqrt{L \log n}$. By Hoeffding's inequality, we have

$$\Pr_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_L}\left[\boldsymbol{N}_j \le \Delta_{j,\mathrm{thr}}\right] \ge 1 - 1/\mathrm{poly}(n).$$

Conditioned on this event, let $\boldsymbol{S} \subseteq [L]$ be the coordinates read by the sampler. We have $|\boldsymbol{S}| \le \Delta_{j,\mathrm{thr}}$. Note that $F_j \ge L \log L = \Theta(\frac{n}{B} \log n)$. By Lemma 18, there is a $(F_j + L) \times (F_j + \Delta_{j,\mathrm{thr}})$ matrix M over $\mathbb{F}_{P_j}$ with $O(1)$ nonzero elements in each row such that for a uniform $\boldsymbol{u} \sim \mathbb{F}_{P_j}^{F_j + \Delta_{j,\mathrm{thr}}}$, we have

$$\Pr_{\boldsymbol{S}}\left[M\boldsymbol{u} \text{ is uniform on } [F_j] \times \boldsymbol{S}\right] \ge 1 - 1/\mathrm{poly}(n).$$

Finally, we use Lemma 15 again to sample $\boldsymbol{u}$ from $\{0,1\}^{m_j}$ with error $1/\mathrm{poly}(n)$, where

$$\begin{aligned} m_j &\le (F_j + \Delta_{j,\mathrm{thr}}) \log_2 P_j + O(\log_2 n) \\ &\le \frac{\overline{W}_{\min}}{4q} \cdot w + p_j L \log_2 P_j + O(\sqrt{L} \log_2^{3/2} n) \\ &\le \frac{\overline{W}_{\min}}{4q} \cdot w + p_j L \log_2 K_j + O(\sqrt{L} \log_2^{3/2} n). \end{aligned}$$

Closeness to uniform follows from Lemma 15, and locality follows since each sampler is $O(1)$ word-local, and thus their composition is also $O(1)$ word-local.

We now analyze the number of uniform bits used to sample this part. Observe that $\sum_{j=1}^{L} p_j \log_2 K_j = \sum_s \mathcal{B}(s) \log_2 K^{(s)} = \mathbb{E}[K^{(\boldsymbol{s})}]$. Therefore, the number of bits used is

$$\begin{aligned} \sum_{j=1}^{q} m_j &\le \frac{L W_{\min}}{4} \cdot w + L \sum_{j=1}^{q} p_j \log_2 K_j + B^{2/3} \cdot O(\sqrt{L} \log_2^{3/2} n) \\ &\le \frac{L\,\mathbb{E}[M^{(\boldsymbol{s})}]}{4} + L\,\mathbb{E}[\log_2 K^{(\boldsymbol{s})}] + B^{2/3} \cdot O(\sqrt{L} \log_2^{3/2} n). \end{aligned} \tag{4}$$

Summing Equations (2) and (4), the number of bits used by the sampler is at most

$$L\,\mathbb{E}[M^{(s)} + \log_2 K^{(s)}] + B^{2/3} \cdot O(\sqrt{L}\log_2^{3/2} n) \leq h(p)n + \sqrt{n} \cdot \mathrm{polylog}(n).$$

This completes the proof of Theorem 2.

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

In this section, we prove Theorem 20, which says 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 20. *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*

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

for some absolute constant $c > 0$.

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

Proposition 21. *Let $f\colon \{0,1\}^m \to \{0,1\}^n$ be a 1-local mapping. Then*

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

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 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

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

- 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 39 we have

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

This completes the proof of Proposition 21. □

Claim 22. *Let $f\colon \{0,1\}^m \to \{0,1\}^2$ be a 2-local mapping. Let $i \in [m]$ be a coordinate 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 20.

Proof of Theorem 20. 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.

Case 1: $|N(I^*)| > n/d$**.** 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 21 the 1-local mapping satisfies

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

Applying Claim 36 with all $2^{|I^*|}$ assignments to the input bits in I^* we get

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

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 20 in case of $|N(I^*)| \geq n/d$.

Case 2: $|N(I^*)| \leq n/d$**.** 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 38 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 23. $|J| \geq \frac{\epsilon'}{2C} n'$.

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 22 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 38 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 20. □

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

In this section, we show that in contrast to Theorem 20, if we allow the sampler to be 3-local, we can approximate the distribution $\mathrm{Ber}(1/4)^n$ within distance of $1/\mathrm{poly}(n)$, and this is optimal up to constant factor in the exponent.

Theorem 24. *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*

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

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

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

Proof of Theorem 24. We start with a graph $G' = (V', E')$ that is 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 exist [Mor94, Theorem 5.13].

Claim 25. *Let $p = 2^{-t}$ for some $t \geq 3$, and let $\boldsymbol{G}'_p = (V, E_p)$ be a random subgraph of G' obtained by keeping each edge in E with probability p independently. Then $\Pr[\boldsymbol{G}'_p$ has a cycle$] < k^{-\frac{2t-5}{3}}$.*

Proof. By the assumption, G' has no cycle 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,

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

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

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

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 $m = |V| + |E| = (3t - 0.5)k = (2 - 1/3t)n = (2-\epsilon)n$. Define $g\colon \{0,1\}^{m-n} \to \{0,1\}^n$ by $g_i(y) = y_{u_i} \oplus y_{v_i}$, where (u_i, v_i) are the endpoints of the i'th edge in G. By Claim 7, we have

$$\mathrm{dist}\big(f(U_m), \mathrm{Ber}(1/4)^n\big) \leq \Pr_{\boldsymbol{S}}[g(\boldsymbol{y}) \text{ restricted to the coordinates in } \boldsymbol{S} \text{ is not uniform}],$$

where $\boldsymbol{S}$ is a uniform subset of $[n]$. We will show that

$$\Pr_{\boldsymbol{S}}[g(\boldsymbol{y})_{\boldsymbol{S}} \text{ is not uniform}] \leq \Pr[\boldsymbol{G}_{1/2} \text{ has a cycle}], \tag{6}$$

where $\boldsymbol{G}_{1/2}$ is a random subgraph of G obtained by keeping each edge with probability $1/2$. This proves Theorem 24 by combining Claim 25 with the observation that $\Pr[\boldsymbol{G}_{1/2} \text{ has a cycle}] = \Pr[\boldsymbol{G}'_{2^{-t}} \text{ has a cycle}]$.

To prove Eq. (6), we will index the edges in E with $[n]$, and view each $S \subseteq [n]$ as the corresponding subset E_S of edges in E. Denote by F the event that the edge set E_S induces a forest in G, i.e., the subgraph (V, E_S) does not contain a cycle. Observe that if $E_S \in F$, then for a uniform $\boldsymbol{y} \in \{0,1\}^{m-n}$, the coordinates $g(\boldsymbol{y})_S = (\boldsymbol{y}_{u_i} \oplus \boldsymbol{y}_{v_i})_{i \in E_S}$ are uniformly distributed. Using the natural correspondence between $E_{\boldsymbol{S}}$ and $\boldsymbol{G}_{1/2}$, this clearly implies Eq. (6). Therefore,

$$\Pr_{\boldsymbol{S}}[g(\boldsymbol{y})_{\boldsymbol{S}} \text{ is not uniform}] \leq \Pr[\boldsymbol{G}_{1/2} \text{ has a cycle}] = \Pr[\boldsymbol{G}'_{2^{-t}} \text{ has a cycle}] \leq \frac{1}{k^{\frac{2t-5}{3}}}.$$

This concludes the proof of Theorem 24. □

5.1 Theorem 24 is tight

Below we show that the analysis of the construction presented in the proof of Theorem 24 is tight up to the exponent of the polynomial. Specifically, we prove the following proposition.

Proposition 26. *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 24. Then*

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

The proposition relies on the following claim.

Claim 27. *Let G be the graph in the proof of Theorem 24. If G contains a cycle of length ℓ then,* $\operatorname{dist}(f(U), \operatorname{Ber}(1/4)^n) \geq (1/4)^\ell$.

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 $\Pr[\operatorname{Ber}(1/4)^\ell = \mathbf{1}] = (1/4)^\ell$. 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 $\operatorname{dist}(f(U), \operatorname{Ber}(1/4)^n) = |\Pr[\mathbf{b} = \mathbf{1}] - \Pr[\operatorname{Ber}(1/4)^\ell = \mathbf{1}]| \geq |0 - (1/4)^\ell| = (1/4)^\ell$.
- 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 $\operatorname{dist}(f(U), \operatorname{Ber}(1/4)^n) = |\Pr[\mathbf{b} = \mathbf{1}] - \Pr[\operatorname{Ber}(1/4)^\ell = \mathbf{1}]| \geq |2(1/4)^\ell - (1/4)^\ell| = (1/4)^\ell$.

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

We now prove Proposition 26.

Proof of Proposition 26. We consider $G' = (V', E')$ as in the proof of Theorem 24. 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 27, 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 28. *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*

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

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 28 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 29 (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 29 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 30. *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 $|N_G(O')| \leq |I'| + |O'| \leq (k-1) + k = 2k - 1$.
- $t \geq 1$: 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

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

This completes the proof of Theorem 28. □

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.

[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.

Theorem 31. *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*

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

for some absolute constant $c > 0$.

The proof of Theorem 31 relies on the following lemma.

Lemma 32. *Let m, n be parameters such that $n \le m \le 2n$, 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*

$$\operatorname{dist}\big(f(U_m), \operatorname{Ber}(1/4)^n\big) \le \Pr_{\boldsymbol{S}\subseteq[n]}[\textit{the rows of } M_{\boldsymbol{S}} \textit{ are linearly dependent}], \tag{7}$$

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

Proof. Define $f\colon \{0,1\}^n \times \{0,1\}^{m-n}$ by $f(x, y) = x \wedge_n My$. Clearly, f is $(d+1)$-local. By Claim 7, for a uniform subset $\boldsymbol{S} \subseteq [n]$,

$$\begin{aligned}\operatorname{dist}\big(\boldsymbol{x} \wedge_n (M\boldsymbol{y}), \operatorname{Ber}(1/4)^n\big) &\le \Pr\big[M\boldsymbol{y} \text{ is not uniform on } \{0,1\}^{\boldsymbol{S}}\big] \\ &= \Pr\big[M_{\boldsymbol{S}} \text{ is linear dependent}\big].\end{aligned}$$

□

Next we state a result about the existence of a sparse matrix which satisfies the conditions in Lemma 32. We show the existence of such matrix by adapting Gallager's result on random sparse matrices [Gal62]. Its proof can be found in Appendix B.

Theorem 33. *For a sufficiently large $n \in \mathbb{N}$ there exists a matrix $M \in \mathbb{F}_2^{n\times 0.75n}$ with exactly 3 ones in every row and 4 ones in each column such that for a uniform subset $\boldsymbol{S} \subseteq [n]$,*

$$\Pr[M_{\boldsymbol{S}} \textit{ is linearly dependent}] \le 2^{-0.05n}.$$

Theorem 31 immediately follows from applying Lemma 32 to Theorem 33.

Proof of Theorem 31. Let $M \in \mathbb{F}_2^{n\times 0.75n}$ be the matrix from Theorem 33, which has 3 ones in each row. By Lemma 32, there is a 4-local sampler $f\colon \{0,1\}^{1.75n} \to \{0,1\}^n$ such that $\operatorname{dist}(f(U_m), \operatorname{Ber}(1/4)^n) \le \Pr[M_{\boldsymbol{S}} \text{ is linear dependent}] \le 2^{-0.05n}$. □

Acknowledgements. We thank the conference reviewers for their comments. Chin Ho Lee is grateful to Cheuk Ting Li for helpful discussion, in particular for pointing out [KY76] and the references on source simulation.

References

[AGMRS26] Yaroslav Alekseev, Mika Göös, Konstantin Myasnikov, Artur Riazanov, and Dmitry Sokolov. "Sampling Permutations with Cell Probes is Hard". In: *58th Annual Symposium on Theory of Computing (STOC)*. 2026 (p. 3).

[Bab87] László Babai. "Random oracles separate PSPACE from the polynomial-time hierarchy". In: *Inform. Process. Lett.* 26.1 (1987), pp. 51–53. DOI: 10.1016/0020-0190(87)90036-6 (p. 2).

[BCS16] Itai Benjamini, Gil Cohen, and Igor Shinkar. "Bi-Lipschitz bijection between the Boolean cube and the Hamming ball". In: *Israel J. Math.* 212.2 (2016), pp. 677–703. DOI: 10.1007/s11856-016-1302-0 (p. 3).

[BHP01] R. C. Baker, G. Harman, and J. Pintz. "The difference between consecutive primes. II". In: *Proc. London Math. Soc. (3)* 83.3 (2001), pp. 532–562. DOI: 10.1112/plms/83.3.532 (p. 18).

[BIL12] Chris Beck, Russell Impagliazzo, and Shachar Lovett. "Large deviation bounds for decision trees and sampling lower bounds for AC0-circuits". In: *53rd Annual Symposium on Foundations of Computer Science (FOCS)*. 2012, pp. 101–110. DOI: 10.1109/FOCS.2012.82 (p. 3).

[BKW03] Avrim Blum, Adam Kalai, and Hal Wasserman. "Noise-tolerant learning, the parity problem, and the statistical query model". In: *J. ACM* 50.4 (2003), pp. 506–519. DOI: 10.1145/792538.792543 (p. 3).

[BL87] R. B. Boppana and J. C. Lagarias. "One-way functions and circuit complexity". In: *Inform. and Comput.* 74.3 (1987), pp. 226–240. DOI: 10.1016/0890-5401(87)90022-8 (p. 2).

[BS23] Lucas Boczkowski and Igor Shinkar. "On mappings on the hypercube with small average stretch". In: *Combin. Probab. Comput.* 32.2 (2023), pp. 334–348. DOI: 10.1017/s0963548322000281 (p. 3).

[CGZ22] Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. "The space complexity of sampling". In: *13th Innovations in Theoretical Computer Science Conference (ITCS)*. 2022, 40:1–40:23. DOI: 10.4230/LIPIcs.ITCS.2022.40 (p. 3).

[CS16] Gil Cohen and Leonard J. Schulman. "Extractors for near logarithmic min-entropy". In: *57rd Annual Symposium on Foundations of Computer Science (FOCS)*. 2016, pp. 178–187. DOI: 10.1109/FOCS.2016.27 (p. 3).

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley & Sons, Inc., New York, 1991, pp. xxiv+542. ISBN: 0-471-06259-6. DOI: 10.1002/0471200611 (p. 2).

[CZ19] Eshan Chattopadhyay and David Zuckerman. "Explicit two-source extractors and resilient functions". In: *Ann. of Math. (2)* 189.3 (2019), pp. 653–705. DOI: 10.4007/annals.2019.189.3.1 (p. 3).

[DILV24a] Harm Derksen, Peter Ivanov, Chin Ho Lee, and Emanuele Viola. "Pseudorandomness, symmetry, smoothing: I". In: *39th Computational Complexity Conference (CCC)*. Vol. 300. 2024, 18:1–18:27. DOI: 10.4230/lipics.ccc.2024.18 (p. 3).

[DILV24b] Harm Derksen, Peter Ivanov, Chin Ho Lee, and Emanuele Viola. *Pseudorandomness, symmetry, smoothing: II*. 2024. arXiv: 2407.12110 [cs.CC]. URL: https://arxiv.org/abs/2407.12110 (p. 3).

[DPT10] Yevgeniy Dodis, Mihai Pătraşcu, and Mikkel Thorup. "Changing base without losing space". In: *43rd Annual Symposium on Theory of Computing (STOC)*. 2010, pp. 593–602. DOI: 10.1145/1806689.1806771 (pp. 6, 14).

[DS26] Thomas L. Draper and Feras A. Saad. "Efficient online random sampling via randomness recycling". In: *37st Annual Symposium on Discrete Algorithms (SODA)*. 2026, pp. 2473–2511. DOI: 10.1137/1.9781611978971.89 (p. 2).

[FLRS23] Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. "Sampling and certifying symmetric functions". In: *27th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*. 2023, 36:1–36:21. DOI: 10.4230/lipics.approx/random.2023.36 (pp. 2, 3).

[Gal62] R. G. Gallager. "Low-density parity-check codes". In: *IRE Trans.* IT-8 (1962), pp. 21–28. DOI: 10.1109/tit.1962.1057683 (pp. 8, 27).

[GGS23] Alexander Golovnev, Tom Gur, and Igor Shinkar. "Derandomization of cell sampling". In: *2023 Symposium on Simplicity in Algorithms (SOSA)*. 2023, pp. 278–284. DOI: 10.1137/1.9781611977585.ch26 (p. 26).

[GKMOW26] Daniel Grier, Daniel M. Kane, Jackson Morris, Anthony Ostuni, and Kewen Wu. "Quantum advantage from sampling shallow circuits: beyond hardness of marginals". In: *17th Innovations in Theoretical Computer Science Conference (ITCS)*. 2026, 73:1–73:14. DOI: 10.4230/lipics.itcs.2026.73 (p. 3).

[GW20] Mika Göös and Thomas Watson. "A lower bound for sampling disjoint sets". In: *ACM Trans. Comput. Theory* 12.3 (2020), 20:1–20:13. DOI: 10.1145/3404858 (p. 3).

[Hag91] Torben Hagerup. "Fast parallel generation of random permutations". In: *18th International Colloquium on Automata, Languages, and Programming (ICALP)*. 1991, pp. 405–416. DOI: 10.1007/3-540-54233-7_151 (p. 2).

[Has86] Johan Torkel Hastad. "Computational Limitations of Small Depth Circuits". Massachusetts Institute of Technology, 1986 (p. 2).

[HH24] Pooya Hatami and William Hoza. "Paradigms for unconditional pseudorandom generators". In: *Found. Trends Theor. Comput. Sci.* 16.1-2 (2024), pp. 1–210. DOI: 10.1561/0400000109 (pp. 3, 8).

[HLV18] Elad Haramaty, Chin Ho Lee, and Emanuele Viola. "Bounded independence plus noise fools products". In: *SIAM J. Comput.* 47.2 (2018), pp. 493–523. DOI: 10.1137/17M1129088 (p. 8).

[HLYZZ25] Yang Hu, Jingxun Liang, Huacheng Yu, Junkai Zhang, and Renfei Zhou. "Optimal static dictionary with worst-case constant query time". In: *57th Annual Symposium on Theory of Computing (STOC)*. 2025, pp. 278–289. DOI: 10.1145/3717823.3718278 (pp. 6, 7, 15–17, 39, 40).

[IN96] Russell Impagliazzo and Moni Naor. "Efficient cryptographic schemes provably as secure as subset sum". In: *J. Cryptology* 9.4 (1996), pp. 199–216. DOI: 10.1007/s001459900012 (p. 2).

[KOW24] Daniel M. Kane, Anthony Ostuni, and Kewen Wu. "Locality bounds for sampling Hamming slices". In: *56th Annual Symposium on Theory of Computing (STOC)*. 2024, pp. 1279–1286. DOI: 10.1145/3618260.3649670 (pp. 2–4, 32).

[KOW25a] Daniel M. Kane, Anthony Ostuni, and Kewen Wu. "Locally sampleable uniform symmetric distributions". In: *57th Annual Symposium on Theory of Computing (STOC)*. 2025, pp. 1807–1816. DOI: 10.1145/3717823.3718243 (pp. 2, 3).

[KOW25b] Daniel M. Kane, Anthony Ostuni, and Kewen Wu. *Symmetric Distributions from Shallow Circuits*. 2025. arXiv: 2511.14127 [cs.CC]. URL: https://arxiv.org/abs/2511.14127 (p. 3).

[KY76] Donald E. Knuth and Andrew C. Yao. "The complexity of nonuniform random number generation". In: *Algorithms and complexity (Proc. Sympos., Carnegie-Mellon Univ., Pittsburgh, Pa., 1976)*. Academic Press, New York-London, 1976, pp. 357–428 (pp. 2, 3, 5, 9, 27).

[Li24] Cheuk Ting Li. "Channel Simulation: Theory and Applications to Lossy Compression and Differential Privacy". In: *Foundations and Trends in Communications and Information Theory* 21.6 (Dec. 2024), pp. 847–1106. DOI: 10.1561/0100000141 (pp. 2, 9).

[LV12] Shachar Lovett and Emanuele Viola. "Bounded-depth circuits cannot sample good codes". In: *Comput. Complexity* 21.2 (2012), pp. 245–266. DOI: 10.1007/s00037-012-0039-3 (pp. 3, 5).

[MCW15] Arya Mazumdar, Venkat Chandar, and Gregory W. Wornell. "Local recovery in data compression for general sources". In: *2015 IEEE International Symposium on Information Theory (ISIT)*. 2015, pp. 2984–2988. DOI: 10.1109/ISIT.2015.7283004 (p. 2).

[Mor94] Moshe Morgenstern. "Existence and explicit constructions of $q + 1$ regular Ramanujan graphs for every prime power q". In: *J. Combin. Theory Ser. B* 62.1 (1994), pp. 44–62. DOI: 10.1006/jctb.1994.1054 (p. 23).

[MV91] Yossi Matias and Uzi Vishkin. "Converting high probability into nearly-constant time—with applications to parallel hashing". In: *23th Annual Symposium on Theory of Computing (STOC)*. 1991, pp. 307–316. DOI: 10.1145/103418.103453 (p. 2).

[Pat08] Mihai Patrascu. "Succincter". In: *49th Annual Symposium on Foundations of Computer Science (FOCS)*. 2008, pp. 305–313. DOI: 10.1109/FOCS.2008.83 (p. 6).

[Reg09] Oded Regev. "On lattices, learning with errors, random linear codes, and cryptography". In: *J. ACM* 56.6 (2009), 34:1–34:40. DOI: 10.1145/1568318.1568324 (p. 3).

[Sha48] C. E. Shannon. "A mathematical theory of communication". In: *Bell System Tech. J.* 27 (1948), pp. 379–423, 623–656. DOI: 10.1002/j.1538-7305.1948.tb01338.x (p. 2).

[Smo87] Roman Smolensky. "Algebraic methods in the theory of lower bounds for Boolean circuit complexity". In: *23th Annual Symposium on Theory of Computing (STOC)*. 1987, pp. 77–82. DOI: 10.1145/28395.28404 (p. 2).

[SS24] Ronen Shaltiel and Jad Silbak. "Explicit codes for poly-size circuits and functions that are hard to sample on low entropy distributions". In: *56th Annual Symposium on Theory of Computing (STOC)*. 2024, pp. 2028–2038. DOI: 10.1145/3618260.3649735 (p. 3).

[Vio12] Emanuele Viola. "The complexity of distributions". In: *SIAM J. Comput.* 41.1 (2012), pp. 191–218. DOI: 10.1137/100814998 (pp. 2, 3, 6, 9, 32).

[Vio14a] Emanuele Viola. "Extractors for circuit sources". In: *SIAM J. Comput.* 43.2 (2014), pp. 655–672. DOI: 10.1137/11085983X (p. 3).

[Vio14b] Emanuele Viola. *Is Nature a Low-Complexity Sampler?* 2014. URL: https://emanueleviola.wordpress.com/2014/11/09/is-nature-a-low-complexity-sampler/ (visited on 04/02/2025) (p. 3).

[Vio23] Emanuele Viola. "New sampling lower bounds via the separator". In: *38th Computational Complexity Conference (CCC)*. 2023, 26:1–26:23. DOI: 10.4230/lipics.ccc.2023.26 (pp. 2–4, 32).

[Vio24] Emanuele Viola. *15 Years of Complexity of Distributions*. 2024. URL: https://emanueleviola.wordpress.com/2024/11/11/15-years-of-complexity-of-distributions/ (visited on 05/03/2026) (p. 3).

[VWY20] Emanuele Viola, Omri Weinstein, and Huacheng Yu. "How to store a random walk". In: *31st Annual Symposium on Discrete Algorithms (SODA)*. 2020, pp. 426–445. DOI: 10.5555/3381089.3381115 (p. 5).

[WP26] Adam Bene Watts and Natalie Parham. "Unconditional quantum advantage for sampling with shallow circuits". In: *17th Innovations in Theoretical Computer Science Conference (ITCS)*. 2026, 17:1–17:12. DOI: 10.4230/lipics.itcs.2026.17 (p. 3).

[Yu22] Huacheng Yu. "Nearly optimal static Las Vegas succinct dictionary". In: *SIAM J. Comput.* 51.3 (2022), STOC20-174–STOC20-249. DOI: 10.1137/20M1363649 (pp. 6, 15).

[YZ24] Huacheng Yu and Wei Zhan. "Sampling, flowers and communication". In: *15th Innovations in Theoretical Computer Science Conference (ITCS)*. 2024, 100:1–100:11. DOI: 10.4230/lipics.itcs.2024.100 (p. 3).

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 34. *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 35. *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; KOW24]. We prove them here for completeness.

Claim 36. *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 37. *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 38. *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 a 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 38. □

Claim 39. *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 35 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 40. *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 41. *The binary entropy function is defined as $h(x) = x \log_2(\frac{1}{x}) + (1-x) \log_2(\frac{1}{1-x})$.*

Fact 42. *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 43 (Chernoff's 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 44 (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 33

In this section, we prove Theorem 33, restated below.

Theorem 33. *For a sufficiently large $n \in \mathbb{N}$ there exists a matrix $M \in \mathbb{F}_2^{n \times 0.75n}$ with exactly 3 ones in every row and 4 ones in each column such that for a uniform subset $\boldsymbol{S} \subseteq [n]$,*

$$\Pr[M_{\boldsymbol{S}} \text{ is linearly dependent}] \leq 2^{-0.05n}.$$

We first prove a claim relating the probability we wish to bound to the weight distribution of vectors in the (left) null space of M. One can think of M as the parity-check matrix of a code of block length n, and the null space is the set of codewords in the code.

Claim 45. *Let $M \in \mathbb{F}_2^{n\times m}$ be a matrix. For a uniformly chosen subset $\boldsymbol{S} \subseteq [n]$, we have*

$$\Pr\left[M_{\boldsymbol{S}} \text{ is linearly dependent}\right] \leq \sum_{\ell=1}^{n} \frac{w_\ell}{2^\ell},$$

where $w_\ell \coloneqq \{z \in \mathbb{F}_2^n : |z| = \ell \text{ and } zM = 0\}$. Moreover, letting $\delta \coloneqq \min\{|z|/n : zM = 0\}$, we have

$$\Pr\left[M_{\boldsymbol{S}} \text{ is linearly dependent}\right] \leq \frac{2^{n-\mathrm{rank}(M)}}{2^{\delta n}}.$$

Proof. Let $z \in \mathbb{F}_2^n$ be of Hamming weight ℓ such that $zM = 0$. Let $T \subseteq [n]$ be the subset of its nonzero coordinates. For every $S \supseteq T$, we have that M_S is linear dependent. There are $2^{n-\ell}$ such subsets S. Therefore,

$$\Pr\left[M_{\boldsymbol{S}} \text{ is linearly independent}\right] \leq 2^{-n} \sum_{z\in\mathbb{F}_2^n} 2^{n-|z|} \mathbb{1}(zM = 0) = \sum_{\ell=1}^{n} \frac{w_\ell}{2^\ell}.$$

The "Moreover" part follows from

$$\sum_{\ell=1}^{n} \frac{w_\ell}{2^\ell} \leq \sum_{\ell=\delta n}^{n} \frac{w_s}{2^{\delta n}} = \frac{1}{2^{\delta n}} \cdot \sum_{\ell=1}^{n} w_\ell = \frac{2^{n-\mathrm{rank}(M)}}{2^{\delta n}}. \qquad \square$$

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 33 follows from applying the following lemma in Claim 45.

Lemma 46. *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 46 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

$$g(s) = 1 + \binom{4}{2} \cdot 2^{2s} + 2^{4s}. \tag{8}$$

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 47. *Let $0 \leq \ell \leq n$, 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, $N_{\mathcal{B}}[\ell] \leq \frac{g(s)^{n/4}}{2^{s\ell}}$ for any $s \in \mathbb{R}$.*

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

$$(g(s))^{n/4} = \sum_{\ell=0}^{n} Q(\ell) 2^{\ell s}, \tag{9}$$

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 $Q(\ell) \geq 0$ and $2^{\ell s} \geq 0$, it follows that $g(s)^{n/4} \geq Q(\ell) \cdot 2^{\ell s}$ for all $\ell \geq 0$ and any $s \in \mathbb{R}$. Therefore,

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

as required. □

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

$$f(\lambda) = \frac{(1 + 6 \cdot 2^{2s} + 2^{4s})^{3/4}}{2^{3s\lambda} \cdot 2^{2h(\lambda)}}, \tag{10}$$

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

Claim 48. *For $\ell = 1, \ldots, n$ let $\lambda = \ell/n$. Then $\mathbb{E}[w_\ell] \leq 8\lambda n f(\lambda)^n$ 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

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

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

Claim 49. *Let $\delta = 0.03$ as in Lemma 46. For any $\lambda \in (0, \delta]$ there exists $s = s(\lambda)$ such that $f(\lambda) < 0.75^\lambda$.*

Claim 50. *Let $\delta = 0.03$ as in Lemma 46. 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 46.

Proof of Lemma 46. 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 51. *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

$$\Pr[\sum_{\ell=1}^{20} w_\ell = 0] > 1 - O(1/n) > 0.99. \tag{11}$$

Next we use Claim 48 and Claim 49 to bound $\mathbb{E}[\sum_{\ell=22}^{\delta n} w_\ell]$. Let $p = 0.75$ be the base of the exponent in Claim 49. 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

$$\Pr[\sum_{\ell=22}^{\delta n} w_\ell = 0] = \Pr[\sum_{\ell=22}^{\delta n} w_\ell < 1] > 1 - 0.89 = 0.11. \tag{12}$$

Combining Eqs. (11) and (12) we get

$$\Pr[\sum_{\ell=1}^{\delta n} w_\ell = 0] > 0.1, \tag{13}$$

assuming that n is sufficiently large.

Then, by Claim 48 and Claim 50 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

$$\Pr\left[\sum_{\ell=\delta n}^{n} \frac{w_\ell}{2^\ell} \geq 0.96^n\right] \leq \frac{2n^2 \cdot 0.95^n}{0.96^n} < 0.02, \tag{14}$$

assuming that n is sufficiently large. By combining Eqs. (13) and (14) 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 46. □

We now return to proving Claim 49 and Claim 50.

Proof of Claim 49. 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 50. 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 $s(\lambda) = \frac{2\log_2(\frac{1}{1-0.75\lambda})}{3}$. Then

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

  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 $(1+6y^4+y^8)^{3/4} \le 1+5y^{3.6}$.

  Denote $F(\lambda) = \frac{1+5\left(\frac{1}{1-0.75\lambda}\right)^{1.2}}{2^{2\lambda\log_2(\frac{1}{1-0.75\lambda})}\cdot 2^{2h(\lambda)+\lambda}}$. 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 $\frac{f(\lambda)}{2^\lambda} \le F(\lambda) < 0.95$ for all $\lambda \in [0.6, 1]$.

This completes the proof of Claim 50. □

C Proof of Lemma 18

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

Claim 52. *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 [HLYZZ25, Lemma 3.3] is the success probability. One can verify this can be improved to $1/n^{C/2}$ by adjusting the field size $|\mathbb{F}|$ from $2n$ to n^C and t from $10\log n$ to $C\log n$. □

Proof of Lemma 18. 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 52, we have

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

So we can fix a choice of G with the desired property. The lemma follows from sparsifying G using elementary operations as in [HLYZZ25, 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 18 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 18, 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 [HLYZZ25, Section 3].

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