# Randomness Efficient Noise Stability and Generalized Small Bias Sets

Dana Moshkovitz [*] Justin Oh[†] David Zuckerman [‡]

**Abstract**

We present a randomness efficient version of the linear noise operator $T_\rho$ from boolean function analysis by constructing a sparse linear operator on the space of boolean functions $\{0,1\}^n \to \{0,1\}$ with similar eigenvalue profile to $T_\rho$. The linear operator we construct is a direct consequence of a generalization of $\epsilon$-biased sets to the product distribution $\mathcal{D}_p$ on $\{0,1\}^n$ where the marginal of each coordinate is $p = \frac{1}{2} - \frac{1}{2}\rho$. Such a generalization is a small support distribution that fools linear tests when the input of the test comes from $\mathcal{D}_p$ instead of the uniform distribution. We give an explicit construction of such a distribution that requires $\log n + O_p(\log\log n + \log\frac{1}{\epsilon})$ bits of uniform randomness to sample from, where the $p$ subscript hides $O(\log^2\frac{1}{p})$ factors. When $p$ and $\epsilon$ are constant, this yields a support size nearly linear in $n$, whereas previous best known constructions only guarantee a size of poly$(n)$. Furthermore, our construction implies an explicitly constructible "sparse" noisy hypercube graph that is a small set expander.

## 1 Introduction

Most constructions in pseudorandomness aim to simulate the behavior of a class of tests when the input to the tests are drawn from the *uniform* distribution on $\{0,1\}^n$. Simulating the uniform distribution has been the main subject of attention because many algorithmic problems ultimately boil down to finding a solution to a problem in an input space where a large fraction of inputs are correct. Thus a uniform sample from the space will find a correct solution with high probability. However, other distributions have also proved to be incredibly useful in solving important problems in computer science. One example of such a distribution is the product distribution with marginals $p$:

**Definition 1** (product distribution with marginals $p$)**.** *Let $p \in [0,1]$. The product distribution with marginals $p$ is the distribution $\mathcal{D}_{p,n}$ on $\{0,1\}^n$ where each bit $x_i$ is picked independently with*

[*] University of Texas at Austin, Department of Computer Science. Supported in part by NSF Grant CCF-1705028 and CCF-1648712. danama@cs.utexas.edu

[†] University of Texas at Austin, Department of Computer Science. Supported by NSF Grant CCF-1705028. sjo@cs.utexas.edu

[‡] University of Texas at Austin, Department of Computer Science. Supported in part by NSF Grant CCF-1705028 and a Simons Investigator Award (#409864). diz@cs.utexas.edu.

*$Pr[x_i = 1] = p$. When the length of the string $n$ is clear from context we simply denote the distribution as $\mathcal{D}_p$.*

Apart from being one of the simplest deviations from the uniform distribution, $\mathcal{D}_p$ in particular serves an integral role in the concept of the noise stability boolean functions. Noise stability is a fundamental concept in boolean function analysis that is pervasive in many branches of mathematics such as social choice theory [O'D14], and has a crucial application in celebrated results in hardness of approximation [Hås96, KKMO07]. Roughly speaking, the stability of a boolean function $f : \{0,1\}^n \to \{0,1\}$ is a measure of how likely the output is to change when each input bit is independently flipped with some small probability $p$. The bit flipping is generally thought of as noise, where input $\mathbf{x} \in \{0,1\}^n$ is perturbed to $\mathbf{x} + \boldsymbol{\mu}$ for $\boldsymbol{\mu} \sim \mathcal{D}_p$. If we instead draw $\mathbf{z} \sim \mathcal{Z}$ and perturb $\mathbf{x}$ to $\mathbf{x} + \mathbf{z}$ for $\mathcal{Z}$ that is a randomness efficient approximation of $\mathcal{D}_p$ (under the right notion of approximation), we can then define a randomness efficient notion of noise. In addition to suggesting a randomness efficient noise test, we believe that the existence of such a notion of noise is of independent interest.

An alternative view of the concept of noise stability relates to the *noise operator* $T_p$,[1] which is a linear operator that acts on truth tables of functions $\mathbf{f} : \{-1,1\}^n \to \{-1,1\}$. The matrix corresponding to $T_p$ is simply the $2^n \times 2^n$ transition matrix of the graph on $\{0,1\}^n$ where a random step from $\mathbf{x}$ moves to $\mathbf{x} + \mathbf{n}$ for $n \sim \mathcal{D}_{p,n}$. Many important properties of the noise operator and noise stability stem from the eigenvalues of $T_p$. Thus we focus on defining a linear noise operator with similar eigenvalue profile to $T_p$. We show that in order to do so it suffices to study a generalization of $\epsilon$-biased sets.

Small bias sets are a fundamental object in pseudorandomness, with applications to error-correcting codes, derandomization, and PCPs [NN93, TS17, BSSVW03]. An $\epsilon$-biased set is a small subset $S \subset \{0,1\}^n$ such that a uniform random sample from $S$ behaves similarly to a uniform random sample from all of $\{0,1\}^n$ with respect to linear tests. More formally, $S$ is an $\epsilon$-biased set if for any nonempty subset of indices $I \subset [n]$, the *bias* of $I$ is small: if $\mathcal{U}(S)$ is the uniform distribution on $S$ then:

$$\left| Pr_{\mathbf{x} \sim \mathcal{U}(S)} \left( \bigoplus_{i \in I} x_i = 0 \right) - Pr_{\mathbf{x} \sim \mathcal{U}(S)} \left( \bigoplus_{i \in I} x_i = 1 \right) \right| \leq \epsilon$$

In other words, the parity of any subset of indices has almost equal probability of being 0 or 1. Notice that in the case of a uniform random sample over $\{0,1\}^n$, the parity of any nonempty subset is equally likely to be 0 or 1. Hence $\epsilon$-biased sample spaces approximate the uniform distribution in the sense that parities of subsets of indices behave almost the way they should. Classic results show that there are $\epsilon$-biased sets that require $O(\log \frac{n}{\epsilon})$ bits of uniform randomness to sample from. In other words there are explicit constructions where the size of $S$ is polynomial in $n$, and optimal constructions even have size linear in $n$ [NN93, TS17]. In addition to having applications in ran-

[1] In mainstream literature, the noise operator that we denote $T_p$ is instead denoted as $T_\rho$ for $\rho = 1 - 2p$. We stray from the standard notation in this paper for convenience with our own notation

domness efficient noise, it is a natural question to ask whether there are small sample spaces that approximate distributions on $\{0,1\}^n$ other than the uniform distribution.

### 1.1 Our Contribution

We generalize $\epsilon$-biased sets for the distribution $\mathcal{D}_p$ on $\{0,1\}^n$. The sample space $\mathcal{Z}$ we construct approximates $\mathcal{D}_p$ in the sense that if $\mathbf{z} \sim \mathcal{Z}$ then for every $I \subset [n]$ the parity of $\mathbf{z}_I$ has approximately the same distribution as when $\mathbf{z}$ is drawn from $\mathcal{D}_p$.

**Theorem** (Main Result)**.** *Let $p$ be a power of 2. There exists a distribution $\mathcal{Z}$ on $\{0,1\}^n$ such that for every $I \subset [n]$ we have:*

$$\left| Pr_{\mathbf{z}\sim\mathcal{Z}}\left(\bigoplus_{i\in I} z_i = 1\right) - Pr_{\mathbf{r}\sim\mathcal{D}_{p,n}}\left(\bigoplus_{i\in I} r_i = 1\right)\right| \le \epsilon$$

*$\mathcal{Z}$ requires $\log n + O(\log\frac{1}{p}\log\log n + \log^2\frac{1}{p} + \log\frac{1}{p}\log\frac{1}{\epsilon})$ bits of uniform randomness to sample from. Moreover, the support of $\mathcal{Z}$ (along with the corresponding probability of each point) can be explicitly constructed in time $n \cdot poly(\log n, \frac{1}{\epsilon})$ for constant $p$.*

The main takeaway from our result is that there is a simple *explicit* construction of a distribution that approximates $\mathcal{D}_{p,n}$ with support size *nearly linear* in $n$ when $p$ and $\epsilon$ are constant. This roughly matches the size of an optimal $\epsilon$-biased set, although the size blows up for nonconstant $p$.

### 1.2 Application to Randomness Efficient Noise

The main application of our generalization of $\epsilon$-biased sets is in the definition of a "randomness efficient" version of noise stability. The stability of the function $\mathbf{f}$ is defined as:

$$\mathrm{Stab}_{1-2p}(\mathbf{f}) = \langle \mathbf{f}, T_p\mathbf{f}\rangle$$

Our construction of $\epsilon$-biased sets for $\mathcal{D}_{p,n}$ naturally suggests a new noise operator $T_{p,\epsilon}^{sparse}$ that is the transition matrix of the graph where a random step from $\mathbf{x}$ moves to $\mathbf{x}+\mathbf{z}$ for $\mathbf{z}$ a sample from our constructed distribution $\mathcal{Z}$. We can then define a new notion of stability:

$$\mathrm{Stab}_{1-2p}^{sparse}(\mathbf{f}) = \langle \mathbf{f}, T_{p,\epsilon}^{sparse}\mathbf{f}\rangle$$

Through analysis of the eigenvalues of $T_p$ and $T_{p,\epsilon}^{sparse}$, we can show that our new notion of stability is the same as the original up to an additive error of $2\epsilon$:

**Theorem** (Randomness Efficient Approximate Noise Stability)**.** *Let $\mathbf{f} : \{-1,1\}^n \to [0,1]$. Let $Stab_{1-2p}(\mathbf{f}) = \langle \mathbf{f}, T_p\mathbf{f}\rangle$ be the stability of $f$ under the noise operator $T_p$. Let $Stab_{1-2p}^{sparse}(\mathbf{f}) = \langle \mathbf{f}, T_{p,\epsilon}^{sparse}\mathbf{f}\rangle$ be the stability of $\mathbf{f}$ under the noise operator $T_{p,\epsilon}^{sparse}$ defined by our $\epsilon$-biased set for $\mathcal{D}_p$. Then:*

$$|Stab_{1-2p}(\mathbf{f}) - Stab_{1-2p}^{sparse}(\mathbf{f})| \le 2\epsilon$$

An immediate consequence of the above theorem is that the majority is stablest theorem, which is a crucial ingredient in hardness of approximation results, is also true for our randomness efficient noise operator up to an additive error of $2\epsilon$. We state the original majority is stablest theorem below:

**Theorem** (Majority Is Stablest [MOO05])**.** *Let $\mathbf{f} : \{-1,1\}^n \to [0,1]$ be a function with $E[\mathbf{f}] = \mu$. Suppose $Inf_i^{\leq 10\log(1/\tau)}(f) \leq \tau$ for all $i \in [n]$. Then:*

$$\langle \mathbf{f}, T_p \mathbf{f} \rangle \leq \Gamma_{1-2p}(\mu) + \frac{10 \log\log(1/\tau)}{(2p)\log(1/\tau)}$$

*where $\Gamma_{1-2p}$ is the Gaussian noise stability curve.*

Our result shows that the stability of a function under our randomness efficient noise operator, $\langle \mathbf{f}, T_{p,\epsilon}^{sparse} \mathbf{f} \rangle$ also obeys the same upper bound, with an extra additive error of $2\epsilon$.

As a secondary application, our construction also implies an explicitly constructible small set expander with large eigenvalues. We say that a graph $G = (V, E)$ is a small set expander if for sufficiently small constant $\delta$ and all subsets of vertices of size $\delta|V|$, the probability of leaving the set in one step of a random walk is at least some constant (say .9). Finding an efficient algorithm for deciding whether a graph is a small set expander remains an open problem. Arora, Barak, and Steurer [ABS15] observed that there is an algorithm that can solve the small set expansion problem in time exponential in the number of eigenvectors of $G$ that have eigenvalue greater than $1 - \xi$. Thus a natural question is how many such eigenvectors could a small set expander have? The noisy hypercube is one of the few "counterexamples" to the efficiency of the above mentioned algorithm, as it is an $N$-vertex graph that can have polylog$(N)$ such eigenvectors. Our construction implies the existence of a "sparse" noisy hypercube with similar spectrum and small set expansion properties.

**Theorem.** *For every $\xi > 0$, there is an explicit $N$-vertex small set expander with polylog$(N)$ eigenvectors with eigenvalue $1 - \xi$. Moreover the graph contains*

$$O\left(N \log N \cdot poly\left(\left(\frac{1}{\xi}\log\log N\right)^{\log\frac{1}{\xi}}\right)\right)$$

*edges.*

The main interest in small set expansion is the relationship between the number of *vertices* and the number of large eigenvalues. Our construction does not improve on any lower bounds on the number of such eigenvalues a small set expander could have. However, we do note that our graph is sparse in the number of edges, containing about $N \log N$ edges as opposed to the $O(N^2)$ needed for the original noisy hypercube.

### 1.3 Background and Related Work

The idea of approximating nonuniform distributions such as $\mathcal{D}_p$ is not entirely new in pseudorandomness. In fact, the linear tests on $\mathcal{D}_p$ that we aim to fool are a special case of combinatorial shapes. An $(m, n)$ combinatorial shape is a function $f : [m]^n \to \{0, 1\}$ that can be expressed as

$$f(x_1, \ldots, x_n) = h(1_{A_1}, \ldots, 1_{A_n})$$

for some symmetric function $h : \{0, 1\}^n \to \{0, 1\}$ and subsets $A_1, \ldots, A_n \subset [m]$. By setting $m = 1/p$ and $h$ as the parity of all its inputs, we can express the parity of any $I \subset [n]$ if we set $A_i = \{1\}$ if $i \in I$ and $A_i = \emptyset$ otherwise. Gopalan, Meka, Reingold, and Zuckerman [GMRZ13] give a PRG that fools all $(m, n)$-combinatorial shapes using seed length $O(\log m + \log n + \log^2(1/\epsilon)) = O(\log 1/p + \log n + \log^2(1/\epsilon))$. The main drawback of [GMRZ13] that we improve on is that the seed length is only guaranteed to be $O(\log n)$, which implies only a polynomial sized construction. On the other hand, when $p$ is a power of 2, our construction guarantees a nearly linear sized construction, with a slightly worse dependence on $p$, and a slightly better dependence on $\epsilon$.

In a previous work, Even, Goldreich, Luby, Nisan, and Veličković [EGL$^+$92] study the approximation of distributions on $[m]^n$ where each coordinate is an independent (and not necessarily identical) distribution. For any distribution $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ where each $\mathcal{D}_i$ is independent, their constructions give sample spaces that have size $(n/\epsilon)^{\log(1/\epsilon)}$ and $(n/\epsilon)^{\log n}$ such that for any $I \subset [n]$, the marginal distribution of the sample space restricted to $I$ is $\epsilon$-close to the marginal distribution of $\mathcal{D}$ in max-norm.

We mention that Meka, Reingold, and Tal [MRT18] define a notion of "$\delta$-biased distributions with marginals $p$." However, their definition of approximation is ad hoc for their main goal of constructing PRGs for width-3 branching programs.

Our application of sparsifying the noisy hypercube is related to the classic result of Spielman and Teng in the edge sparsification of graphs [ST11]. Indeed, their sparsification algorithm, when run on the noisy hypercube, should produce a sparsified graph with the properties we aim to preserve. However, the main drawback to this approach is that the sparsification algorithm runs in time $m\text{polylog}(m)$ where $m$ is the number of edges. In the case of the noisy hypercube, which is a dense graph defined on $\{0, 1\}^n$, this algorithm is much less efficient than the explicit construction we provide.

Barak et al. previously explored the idea of reducing the size of the noisy hypercube, which has close ties to hardness of approximation [BGH$^+$11]. Their work presents a "derandomized noisy hypercube" along with the appropriate analogues of small set expansion and the majority is stablest theorem. As their interest was in the relationship between the number of *vertices* and the number of large eigenvalues of a small set expander, their constructed graph contains a reduced number of vertices. On the other hand, our construction keeps the same $2^n$ vertices of the original noisy hypercube and reduces the number of edges.

### 1.4 Overview of Techniques

The construction of the randomness efficient noise operator and small set expanders are essentially direct applications of our construction of generalized small bias sets. Thus here we focus on the intuition behind our construction. It's easy to see that the bitwise product of $\log_2(1/p)$ independent uniform samples from $\{0,1\}^n$ is exactly equivalent to $\mathcal{D}_p$ for $p$ a power of 2. Thus intuitively, if $\epsilon$-biased sets approximate the uniform distribution on $\{0,1\}^n$, then the bitwise product of $\log_2(1/p)$ random draws from an $\epsilon$-biased set should approximate $\mathcal{D}_p$. Our main construction formalizes this intuition by showing via a hybrid argument that such a bitwise product indeed fools linear tests when the input is drawn from $\mathcal{D}_p$. This simple idea is not sufficient however, as the final seed length will be roughly $\log_2(1/p) \log n$ which implies at least a polynomial sized support for small $p$.

To improve the dependence on $n$, we observe that the parities of sufficiently large $I \subset [n]$ will be close to uniform on $\{0,1\}$. More specifically, the probability that the parity of a subset of indices $I$ under the distribution $\mathcal{D}_p$ is 1 is $\frac{1}{2} - \frac{1}{2}(1-2p)^{|I|}$. Thus for $|I| \geq \frac{1}{2p}\ln(\frac{1}{\epsilon})$ the probability of the parity being 1 is $\epsilon/2$ close to $1/2$. This means that we only need to accurately simulate the behavior of $\mathcal{D}_p$ for $|I|$ smaller than $k = \frac{1}{2p}\ln(\frac{1}{\epsilon})$. For large $|I|$ we simply need to simulate the uniform distribution. To do so, we can take the bitwise AND of $\log_2(1/p) - 1$ independent samples from a $k$-wise $\epsilon$-biased set (using seed length only $\log\log n$). This simulates $\mathcal{D}_{p/2}$. Finally we take the bitwise product of this with a final $\epsilon$-biased set with seed length $\log n$. For small $|I|$, the behavior of the parities under $\mathcal{D}_p$ are preserved, and for large $|I|$, the product of the $k$-wise $\epsilon$-biased sets will contain at least one 1, so the final probability the parity is 1 will be the probability that the final $\epsilon$-biased set outputs 1 on a specific coordinate, which is roughly $1/2$.

### 1.5 Paper Organization

In Section 2 we define the necessary preliminaries and notation. Section 3 presents and proves the correctness of our construction and Section 4 presents the applications of our result to randomness efficient noise and small set expansion. Finally in Section 5 we discuss lower bounds for our generalization of $\epsilon$-biased sets and further directions for research.

## 2 Preliminaries and Notation

In general we denote random variables as capital letters such as $X$ and $Y$. We denote fixed values using lowercase such as $x, y$. Distributions are denoted with calligraphic capital letters such as $\mathcal{D}$, and the uniform distribution on a set $S$ is denoted via $\mathcal{U}(S)$. We distinguish vector-valued random variables from scalars via boldface: $\mathbf{X}, \mathbf{x}$, and refer to a value at a specific index of a vector via the corresponding nonbolded symbol with subscript: $X_i, x_i$. Vectors in this paper generally take on values in the field $\mathbb{F}_2$ and thus arithmetic is generally done modulo 2. We use $\langle \cdot, \cdot \rangle$ to denote the inner product of two vectors modulo 2. Finally, we define the binary operation "$\odot$" between two vectors as the entrywise product modulo 2. For example, for $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$, we have: $\mathbf{X} \odot \mathbf{Y} = (X_1Y_1, \ldots, X_nY_n)$. It is straightforward to verify that for any vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \{0,1\}^n$, we have: $\langle \mathbf{x}, \mathbf{y} \odot \mathbf{z} \rangle = \langle \mathbf{x} \odot \mathbf{y}, \mathbf{z} \rangle$

We first define the bias of a subset according to a distribution.

**Definition 2** (Bias)**.** *Let $I \subset [n]$ and $\mathcal{D}$ be any distribution on $\{0,1\}^n$. Then the bias of $I$ according to $\mathcal{D}$ is defined as*

$$b_{I,\mathcal{D}} = Pr_{\mathbf{x} \sim \mathcal{D}} \left[ \bigoplus_{i \in I} x_i = 0 \right] - Pr_{\mathbf{x} \sim \mathcal{D}} \left[ \bigoplus_{i \in I} x_i = 1 \right]$$

*Equivalently, if $\boldsymbol{\alpha} \in \{0,1\}^n$ then we say that the bias is:*

$$b_{\boldsymbol{\alpha},\mathcal{D}} = Pr_{\mathbf{x} \sim \mathcal{D}} \left[ \langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 0 \right] - Pr_{\mathbf{x} \sim \mathcal{D}} \left[ \langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 1 \right]$$

When the probability distribution is clear from context, we denote the bias of $I$ as $b_I$.

Next, we define the concept of $\epsilon$-biased sets and $k$-wise independent $\epsilon$-biased sets, both of which have already well known constructions, and are crucial for our construction of $\epsilon$-biased product distributions with marginals $p$.

**Definition 3** ($\epsilon$-biased set)**.** *An $\epsilon$-biased set is a small set $S \subset \{0,1\}^n$ such that for every $\boldsymbol{\alpha} \in \{0,1\}^n$ we have:*

$$|b_{\boldsymbol{\alpha},\mathcal{U}(S)}| = \left| Pr_{\mathbf{x} \sim \mathcal{U}(S)}[\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 0] - Pr_{\mathbf{x} \sim \mathcal{U}(S)}[\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 1] \right| \leq \epsilon$$

*or equivalently:*

$$\left| Pr_{\mathbf{x} \sim \mathcal{U}(S)}[\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 1] - Pr_{\mathbf{x} \sim \mathcal{U}(\{0,1\}^n)}[\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 1] \right| \leq \epsilon/2$$

Numerous works [NN93,TS17] show that there are explicit constructions of $\epsilon$-biased sets that require $\log n + O(\log \frac{1}{\epsilon})$ random bits to specify a random point in $S$, or in other words, the size of $S$ is linear in $n$. A weaker notion of $\epsilon$-biased sets only considers the parity of subsets of indices of size at most $k$:

**Definition 4** ($k$-wise $\epsilon$-biased set)**.** *A $k$-wise $\epsilon$-biased set is a small set $S \subset \{0,1\}^n$ such that for any $\boldsymbol{\alpha} \in \{0,1\}^n$ with hamming weight $|\boldsymbol{\alpha}| \leq k$. We have:*

$$|b_{\boldsymbol{\alpha},\mathcal{U}(S)}| = \left| Pr_{\mathbf{x} \sim \mathcal{U}(S)}[\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 0] - Pr_{\mathbf{x} \sim \mathcal{U}(S)}[\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 1] \right| \leq \epsilon$$

*or equivalently:*

$$\left| Pr_{\mathbf{x} \sim \mathcal{U}(S)}[\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 1] - Pr_{\mathbf{x} \sim \mathcal{U}(\{0,1\}^n)}[\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 1] \right| \leq \epsilon/2$$

Naor and Naor show that there are explicit constructions of $k$-wise $\epsilon$-biased sets that require $O(\log k + \log \log n + \log \frac{1}{\epsilon})$ random bits to specify a random point in $S$.

Our notion of approximating a product distribution with marginals $p$ is the natural extension of the notion of approximation given by $\epsilon$-biased sets: the parity of any subset of coordinates from our approximate distribution should look like the parity of the subset of coordinates from $\mathcal{D}_p$.

**Definition 5** ($(p, \epsilon)$-biased sample space)**.** *Let $p \in [0,1]$. A $(p, \epsilon)$-biased sample space is a distribution $\mathcal{Z}$ on $\{0,1\}^n$ with small support $S \subset \{0,1\}^n$ such that for every $\boldsymbol{\alpha} \in \{0,1\}^n$ we have:*

$$\left| Pr_{\mathbf{z} \sim \mathcal{Z}}[\langle \boldsymbol{\alpha}, \mathbf{z} \rangle = 1] - Pr_{\mathbf{r} \sim \mathcal{D}_p}[\langle \boldsymbol{\alpha}, \mathbf{r} \rangle = 1] \right| \leq \epsilon$$

Historically, the definition of $\epsilon$-biased sets and $k$-wise independent $\epsilon$-biased sets use small bias as their notion of approximation. As stated in their definitions above, this notion is equivalent (up to constant factors) with the alternate notion that the distribution of the outputs of any linear function on input $x \sim U(S)$ is close to the distribution when $x \sim U(\{0,1\}^n)$. This equivalence no longer holds in the generalized notion of $\epsilon$-biased sets for $\mathcal{D}_p$. For example, if $p$ is small, then the bias $b_{I,\mathcal{D}_p}$ is almost 1 for any singleton subset $I$. The nonequivalence of these notions makes some simple facts about standard $\epsilon$-biased sets more tedious to prove for $\epsilon$-biased sets for $\mathcal{D}_p$. For completeness, we now state the facts important for our analysis, and defer their proofs to the appendix.

First, there is a well known relationship between the biases of a random $x \in \{0,1\}^n$ (over any distribution) and the probability mass function for the distribution.

**Proposition 6.** *Let $\mathcal{D}$ be any distribution. For any $\mathbf{a} \in \{0,1\}^n$, let $p_{\mathbf{a},\mathcal{D}}$ be the probability of sampling $\mathbf{a}$ under $\mathcal{D}$. Let $\mathbf{p}$ be the $2^n$ length vector of probabilities $p_{\mathbf{a},\mathcal{D}}$ for each $\mathbf{a}$. Let $\mathbf{b}$ be the $2^n$ length vector of biases $b_{\boldsymbol{\alpha},\mathcal{D}}$ indexed by $\boldsymbol{\alpha} \in \{0,1\}^n$. Let the Hadamard matrix $H$ be the $2^n \times 2^n$ matrix where each entry is defined as $(-1)^{\langle \boldsymbol{\alpha}, \mathbf{a} \rangle}$ then:*

$$\mathbf{p} = 2^{-n} H^T \mathbf{b}$$

Given this proposition, we can prove a necessary fact for the analysis of our construction that if $\mathcal{Z}$ is an $(p, \epsilon)$-biased space for $\mathcal{D}_p$, then $\mathcal{Z}$ is close in max-norm to $\mathcal{D}_p$.

**Corollary 7** ($\epsilon$-biased implies close in max norm)**.** *Let $\mathcal{Z}$ be an $(p, \epsilon)$-biased sample space. Then $\mathcal{Z}$ is $2\epsilon$-close to $\mathcal{D}_p$ in max-norm. That is, for any $\mathbf{a} \in \{0,1\}^n$ we have:*

$$|p_{\mathbf{a},\mathcal{Z}} - p_{\mathbf{a},\mathcal{D}_p}| \leq 2\epsilon$$

Finally, we note a useful fact that the distribution of the parity of $k$ independent random variables in $\{0,1\}$ with marginals $p$ is close to uniform on $\{0,1\}$ for sufficiently large $k$. The proof is again deferred to the appendix.

**Proposition 8.** *Consider $k$ independent tosses of a biased coin with $Pr[Heads] = p$. Then the probability of an odd number of heads is $\frac{1}{2} - \frac{1}{2}(1-2p)^k$.*

## 3 Construction

Our construction of a $(p, \epsilon)$-biased space for $\mathcal{D}_p$ is as follows:

**Construction.** *Let $k = \frac{1}{p} \ln \frac{100}{\epsilon}$ and $t = \log_2 \frac{1}{2p}$. Let $\epsilon' = \frac{1}{100} \frac{2\epsilon}{t+1} = \frac{1}{100} \frac{2\epsilon}{\log_2 \frac{1}{p}} < \frac{\epsilon}{4} \leq \epsilon$.*

*For $1 \leq i \leq t$, let $\mathbf{X}_i$ be $t$ independent draws from a $k$-wise $\epsilon'$-biased set of $\{0,1\}^n$. We let $\mathbf{X} = \bigodot_{i=1}^t \mathbf{X}_i$. Let $\mathbf{Y}$ be drawn from an $\epsilon'$-biased set of $\{0,1\}^n$. Our final distribution is then $\mathbf{Z} = \mathbf{X} \odot \mathbf{Y}$.*

We first state a main lemma that the product of $\epsilon$-biased spaces approximates $\mathcal{D}_p$ with the right notion of approximation. We defer the proof to the appendix.

**Lemma 9** (Coordinate-wise product of $\epsilon$-biased sets is $\epsilon$-biased for $\mathcal{D}_p$)**.** *Let $k \leq n$ and let $\mathbf{X}_1, \ldots, \mathbf{X}_t$ be independent draws from $k$-wise $\epsilon$-biased sets on $\{0,1\}^n$. Then $\mathbf{X} = \bigodot_i \mathbf{X}_i$ is a $k$-wise $(\frac{1}{2^t}, t\epsilon/2)$-biased sample space.*

Given the lemma, we can then prove the correctness of our construction.

**Theorem 10** (Main Result)**.** *Let $0 < p < 1/2$. For any $\epsilon > 0$, $\mathbf{Z}$ is a $(p,\epsilon)$-biased sample space requiring $\log n + O(\log^2 \frac{1}{p} + \log \frac{1}{p} \log \frac{1}{\epsilon} + \log \frac{1}{p} \log\log n)$ uniform random bits to sample from.*

*Proof.* We first note that using the constructions mentioned above, generating $\mathbf{Z}$ requires $\log n + O(t(\log k + \log\log n + \log \frac{1}{\epsilon'}) + \log \frac{1}{\epsilon'}) = \log n + O(\log^2 \frac{1}{p} + \log \frac{1}{p} \log \frac{1}{\epsilon} + \log \frac{1}{p} \log\log n)$ bits. Moreover, since the original constructions are explicit, we can construct the support of $\mathbf{Z}$ via enumeration of all elements in each used $\epsilon$-biased set.

We claim that $\mathbf{Z}$ is an $\epsilon$-biased distribution for $\mathcal{D}_p$. We show that for any $\boldsymbol{\alpha} \in \{0,1\}^n$:

$$\left| Pr_{\mathbf{z} \sim \mathbf{Z}}[\langle \boldsymbol{\alpha}, \boldsymbol{z} \rangle = 1] - Pr_{\boldsymbol{r} \sim \mathcal{D}_p}[\langle \boldsymbol{\alpha}, \boldsymbol{r} \rangle = 1] \right| \leq \epsilon$$

The proof splits into two cases. For the first case, assume $|\alpha| \leq k$. Since $\mathbf{Y}$ and the $\mathbf{X}_i$'s are $k$-wise $\epsilon'$-biased, by Lemma 9 we have immediately that:

$$\left| Pr_{\mathbf{z} \sim \mathbf{Z}}[\langle \boldsymbol{\alpha}, \boldsymbol{z} \rangle = 1] - Pr_{\boldsymbol{r} \sim \mathcal{D}_p}[\langle \boldsymbol{\alpha}, \boldsymbol{r} \rangle = 1] \right| \leq (t+1)\frac{\epsilon'}{2} \leq \epsilon$$

In the second case, assume $|\boldsymbol{\alpha}| > k$. Let $I \subset [n]$ be any subset of the indices of size exactly $k$ for which $\boldsymbol{\alpha}$ is 1. Consider the first component in the construction of $\mathbf{Z}$:

$$\mathbf{X} = \bigodot_{i=1}^{t} \mathbf{X}_i$$

where each $\mathbf{X}_i \in \{0,1\}^n$ is drawn from a $k$-wise $\epsilon'$-biased set. By Lemma 9, we know that the substring of $\mathbf{X}$ restricted only to indices in $I$, denoted $\mathbf{X}_I \in \{0,1\}^k$, is $(\frac{p}{2}, \gamma)$-biased for $\mathcal{D}_{\frac{p}{2},k}$ for $\gamma \leq t\epsilon'/2 \leq \epsilon/100$. Thus by Corollary 7, the distribution of $\mathbf{X}_I$ is $\frac{\epsilon}{50}$-close to $\mathcal{D}_{\frac{p}{2},k}$ in max-norm. In particular, this means that:

$$P(\mathbf{X}_I = 0^k) \leq (1-p)^k + \frac{\epsilon}{50} \leq (1-p)^{\frac{1}{p}\ln\frac{100}{\epsilon}} + \frac{\epsilon}{50} = \frac{\epsilon}{100} + \frac{\epsilon}{50} \leq \frac{\epsilon}{4}$$

Thus with probability at least $1 - \epsilon/4$, the string $\mathbf{X}$ will contain at least one 1 on an index where $\boldsymbol{\alpha}$ is 1. This means that we have:

$$\begin{aligned} Pr_{\mathbf{z} \sim \mathbf{Z}}(\langle \boldsymbol{\alpha}, \mathbf{z} \rangle = 1) &= P(\langle \boldsymbol{\alpha}, \mathbf{z} \rangle = 1 \wedge \mathbf{X}_{\boldsymbol{\alpha}} = 0^{|\boldsymbol{\alpha}|}) + P(\langle \boldsymbol{\alpha}, \mathbf{z} \rangle = 1 \wedge \mathbf{X}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}) \\ &\leq \frac{\epsilon}{4} + \sum_{\mathbf{x}: \mathbf{x}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}} P(\langle \boldsymbol{\alpha}, \mathbf{z} \rangle = 1 \wedge \mathbf{X} = \mathbf{x}) \end{aligned}$$

$$
\begin{aligned}
&= \frac{\epsilon}{4} + \sum_{\mathbf{x}:\mathbf{x}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}} P(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1 \mid \mathbf{X} = \mathbf{x}) P(\mathbf{X} = \mathbf{x}) \\
&= \frac{\epsilon}{4} + \sum_{\mathbf{x}:\mathbf{x}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}} P(\langle \boldsymbol{\alpha} \odot \mathbf{x}, \mathbf{y}\rangle = 1) P(\mathbf{X} = \mathbf{x}) \\
&\leq \frac{\epsilon}{4} + \sum_{\mathbf{x}:\mathbf{x}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}} (\frac{1}{2} + \epsilon') P(\mathbf{X} = \mathbf{x}) \\
&\leq \frac{1}{2} + \frac{\epsilon}{4} + \epsilon' \leq \frac{1}{2} + \frac{\epsilon}{2}
\end{aligned}
$$

Similarly for a lower bound we have:

$$
\begin{aligned}
Pr_{\mathbf{z}\sim\mathbf{Z}}(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1) &= P(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1 \wedge \mathbf{X}_{\boldsymbol{\alpha}} = 0^{|\boldsymbol{\alpha}|}) + P(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1 \wedge \mathbf{X}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}) \\
&\geq 0 + \sum_{\mathbf{x}:\mathbf{x}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}} P(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1 \wedge \mathbf{X} = \mathbf{x}) \\
&= \sum_{\mathbf{x}:\mathbf{x}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}} P(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1 \mid \mathbf{X} = \mathbf{x}) P(\mathbf{X} = \mathbf{x}) \\
&= \sum_{\mathbf{x}:\mathbf{x}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}} P(\langle \boldsymbol{\alpha} \odot \mathbf{x}, \mathbf{y}\rangle = 1) P(\mathbf{X} = \mathbf{x}) \\
&\geq (\frac{1}{2} - \epsilon') \sum_{\mathbf{x}:\mathbf{x}_{\boldsymbol{\alpha}} \neq 0^{|\boldsymbol{\alpha}|}} P(\mathbf{X} = \boldsymbol{x}) \\
&\geq (\frac{1}{2} - \epsilon')(1 - \frac{\epsilon}{4}) \\
&= \frac{1}{2} + \epsilon' \frac{\epsilon}{4} - \epsilon' - \frac{\epsilon}{8} \\
&\geq \frac{1}{2} - \epsilon' - \frac{\epsilon}{4} \\
&\geq \frac{1}{2} - \frac{\epsilon}{2}
\end{aligned}
$$

Combining the upper and lower bound shows that $Pr_{\mathbf{z}\sim\mathbf{Z}}(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1)$ is $\epsilon/2$ close to $1/2$. Since by Proposition 8 we know that $Pr_{\mathbf{r}\sim\mathcal{D}_p}(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1)$ is also $\epsilon/2$ close to $1/2$ we must have that:

$$
|Pr_{\mathbf{z}\sim\mathbf{Z}}(\langle \boldsymbol{\alpha}, \mathbf{z}\rangle = 1) - Pr_{\mathbf{r}\sim\mathcal{D}_p}(\langle \boldsymbol{\alpha}, \mathbf{r}\rangle = 1)| \leq \epsilon
$$

□

## 4 Applications

We first define the noisy hypercube, which is a crucial graph in our applications and also an important graph in many areas of theoretical computer science.

**Definition 11** (Noisy Hypercube Graph)**.** *The $p$-noisy hypercube graph, which we denote $T_p$, is the graph on vertex set $\{0,1\}^n$ such that a random step from node $\mathbf{a} \in \{0,1\}^n$ is equivalent to picking $\mathbf{r} \sim \mathcal{D}_p$ and moving to $\mathbf{a} + \mathbf{r}$.*

Note that the transition matrix of $T_p$ has no nonzero entries since there is a nonzero probability of reaching any node from any other node and is thus very dense. Our $\epsilon$-biased distribution for $\mathcal{D}_p$ allows us to construct a spare noisy hypercube that has similar properties to the original noisy hypercube but with fewer edges.

**Definition 12** (Sparse Noisy Hypercube Graph)**.** *Let $\mathbf{Z}$ be an $(p, \epsilon)$-biased sample space. The sparse $(p, \epsilon)$-noisy hypercube graph, which we denote $T_{p,\epsilon}^{sparse}$, is the graph on vertex set $\{0,1\}^n$ such that a random step from node $\mathbf{a} \in \{0,1\}^n$ is equivalent to picking $\mathbf{z} \sim \mathbf{Z}$ and moving to $\mathbf{a} + \mathbf{z}$.*

Because of the size of our construction's seed length, each row and column of the $2^n \times 2^n$ transition matrix of $T_{p,\epsilon}^{sparse}$ has $\tilde{O}(n)$ nonzero entries when $p$ and $\epsilon$ are constant.

We first show that the noise operator defined by $T_{p,\epsilon}^{sparse}$ has similar eigenvalues to that of the original noise operator. This leads to the fact that our randomness efficient notion of stability approximates the original notion of stability, and also implies that the graph $T_{p,\epsilon}^{sparse}$ is our desired sparse small set expander.

### 4.1 Eigenvalues

The main feature about $T_{p,\epsilon}^{sparse}$ from which our applications arise is that it has a similar spectrum to $T_p$. We first give a well known (and easily verifiable) fact about the eigenvalues and eigenvectors of graphs on the boolean hypercube that are defined like above.

**Theorem 13.** *Let $\mathbf{Z}$ be any distribution on $\{0,1\}^n$. Define $G = (V, E)$ on vertices $V = \{0,1\}^n$ as the graph on which a random step starting at $\mathbf{a} \in \{0,1\}^n$ is equivalent to drawing $\mathbf{z} \in \mathbf{Z}$ and moving to $\mathbf{a} + \mathbf{z}$. Let $M$ be the $2^n \times 2^n$ transition matrix of $G$. For every subset of indices $I \subset [n]$, define the vector $\mathbf{v}_I \in \{-1,1\}^{2^n}$ to be $1$ if the parity of the $i$th bitstring in $\{0,1\}^n$ restricted to $I$ is 0 and $-1$ if the parity is 1. Each $\mathbf{v}_I$ is an eigenvector of $M$ with eigenvalue $b_{I,Z}$.*

Given this well known fact it is straightforward to see that the eigenvalue profiles of $T_p$ and $T_{p,\epsilon}^{sparse}$ are close:

**Corollary 14.** *The graphs $T_p$ and $T_{p,\epsilon}^{sparse}$ have the same eigenvectors. For every eigenvector $\mathbf{v}$ of both graphs, the corresponding eigenvalues differ by at most $2\epsilon$.*

*Proof.* By Theorem 13, both $T_p$ and $T_{p,\epsilon}^{sparse}$ have the same eigenvectors $\mathbf{v}_I \in \{-1,1\}^{2^n}$. For any $I$, $\mathbf{v}_I$ has eigenvalue $b_{I,\mathcal{D}_p}$ in $T_p$ and $b_{I,\mathbf{Z}}$ in $T_{p,\epsilon}^{sparse}$ where $\mathbf{Z}$ is an $\epsilon$-biased distribution for $\mathcal{D}_p$. However we know by definition of $(p, \epsilon)$-biased distribution that:

$$|b_{I,\mathbf{Z}} - b_{I,\mathcal{D}_p}| \leq 2\epsilon$$

□

## 4.2 Randomness Efficient Noise

The stability of a boolean function $\mathbf{f}$ on $\{-1,1\}^n$ is a fundamental concept in the analysis of boolean functions that measures the tendency of the output of a function to change when each bit of the input is flipped independently with probability $p$. In our context, the stability is equivalent to

$$\mathrm{Stab}_{1-2p} = \langle \mathbf{f}, T_p \mathbf{f} \rangle$$

where we think of $\mathbf{f}$ as a $2^n$ length truth table, and $T_p$ is the transition matrix of the noisy hypercube above (here we no longer think of $\langle \cdot, \cdot \rangle$ as the inner product modulo 2).

We can show that the stability of a function under our notion of "derandomized noise", where noise is added to the input via a sample from a $(p, \epsilon)$-biased space for $\mathcal{D}_p$ is close to the original notion of stability.

**Theorem 15** (Randomness Efficient Noise Stability is Close to Noise Stability)**.** *Let* $\mathbf{f} : \{-1,1\}^n \to [0,1]$ *be a function with* $E[\mathbf{f}] = \mu$. *Then:*

$$Stab_{1-2p}(\mathbf{f}) - 2\epsilon \leq Stab_{1-2p}^{sparse}(\mathbf{f}) \leq Stab_{1-2p}(\mathbf{f}) + 2\epsilon$$

*Proof.* We can write $\mathbf{f}$ in the fourier basis as:

$$\mathbf{f} = \sum_I f_I \mathbf{v}_I$$

It is a well know fact in fourier analysis that:

$$\langle \mathbf{f}, T_p \mathbf{f} \rangle = \sum_I b_{I,\mathcal{D}_p} f_I^2$$

Similarly we can derive the corresponding expression for $T_{p,\epsilon}^{sparse}$:

$$\begin{aligned}
&\langle \mathbf{f}, T_{p,\epsilon}^{sparse} \mathbf{f} \rangle \\
&= \left\langle \sum_I f_I \mathbf{v}_I, T_{p,\epsilon}^{sparse} \sum_I f_I \mathbf{v}_I \right\rangle \\
&= \left\langle \sum_I f_I \mathbf{v}_I, \sum_I b_{I,\mathbf{Z}} f_I \mathbf{v}_I \right\rangle \\
&= \sum_I b_{I,\mathbf{Z}} f_I^2 \langle \mathbf{v}_I, \mathbf{v}_I \rangle \\
&\leq \sum_I (b_{I,\mathcal{D}_p} + 2\epsilon) f_I^2 \\
&= \langle \mathbf{f}, T_p \mathbf{f} \rangle + 2\epsilon
\end{aligned}$$

For the lower bound, we replace the inequality with $b_{I,\mathbf{Z}} \geq b_{I,\mathcal{D}_p} - 2\epsilon$ □

### 4.3 Small Set Expansion

We now show that our sparse noisy hypercube is our desired sparse small set expander with large eigenvalues. We first define the expansion of a graph.

**Definition 16** (Expansion)**.** *Given graph $G = (V, E)$, let $S$ be any subset of vertices of $G$. The expansion of $S$, denote $\Phi(S)$ is the probability that a randomly chosen edge $(u, v)$ has $v \notin S$ conditioned on $u \in S$. Equivalently, if $G$ is a regular undirected graph, we have:*

$$\Phi_G(S) = \frac{E(S, V \setminus S)}{\sum_{v \in S} deg(v)}$$

In the context of small set expansion, we are typically interested in the expansion of sets that contain a small constant fraction $\delta$ of vertices. We say that a graph is a small set expander if for sufficiently small $\delta$, all subsets containing $\delta$-fraction of vertices have expansion at least some constant (such as 0.9). We know that the noisy hypercube has $n$ eigenvalues that are at least $1 - 2p$. As a consequence of Corollary 14, we know that $T_{p,\epsilon}^{sparse}$ has at least $n$ eigenvalues that are at least $1-2p-2\epsilon$.

It remains to verify that the sparse noisy hypercube is also a small set expander. The following theorem relates the top eigenvectors of a graph to the expansion of sets [BGH+11].

**Theorem 17.** *For any vector space $\mathcal{V}$, define the $p \to q$ norm of a subspace $\mathcal{U}$ of $\mathcal{V}$ as:*

$$||\mathcal{U}||_{p\to q} = \max_{v \in \mathcal{V}} \frac{||P_{\mathcal{U}} v||_q}{||v||_p}$$

*Where $P_{\mathcal{U}}$ is the projection operator onto subspace $\mathcal{U}$.*

*For graph $G = (V, E)$, let $\mathcal{U}$ be the subspace spanned by all eigenvectors of $G$ with eigenvalue larger than $\lambda$. Then for any $S \subset V$ containing $\delta$ fraction of vertices we have:*

$$\Phi(S) \geq 1 - \lambda - ||\mathcal{U}||_{2\to 4}^2 \sqrt{\delta}$$

In the case of the noisy hypercube, once can show via the Bonami Lemma that $||\mathcal{U}||_{2\to 4}$ is bounded. This implies via Theorem 17 that for sufficiently small $\delta$, the expansion of $S$ is large. Finally, the next corollary relates the expansion of sets in $T_{p,\epsilon}^{sparse}$ to those in $T_p$.

**Corollary 18.** *Let $\mathcal{U}_{true}$ be the subspace spanned by all eigenvectors of $T_p$ with eigenvalue larger than $\lambda$. Let $\mathcal{U}_{pseudo}$ be the subspace spanned by all eigenvectors of $T_{p,\epsilon}^{sparse}$ with eigenvalue larger than $\lambda + 2\epsilon$. Then for any $S \subset V$ that contains $\delta$ fraction of vertices we have:*

$$\Phi_{T_{p,\epsilon}^{sparse}}(S) \geq 1 - \lambda - ||\mathcal{U}_{true}||_{2\to 4}^2 \sqrt{\delta} - 2\epsilon$$

*Proof.* Observe that since the eigenvalues of $T_p$ are at most $2\epsilon$ away from the eigenvalues of $T_{p,\epsilon}^{sparse}$, we have $U_{pseudo} \subset U_{true}$. This implies that $||U_{pseudo}||_{2\to 4} \leq ||U_{true}||_{2\to 4}$. Thus by Theorem 17 we have:

$$\Phi_{T_{p,\epsilon}^{sparse}}(S) \geq 1 - (\lambda + 2\epsilon) - ||\mathcal{U}_{pseudo}||_{2\to 4}^2 \sqrt{\delta} \geq 1 - \lambda - ||\mathcal{U}_{true}||_{2\to 4}^2 \sqrt{\delta} - 2\epsilon$$

□

Thus sets in $T_{p,\epsilon}^{sparse}$ have similar expansion to those in $T_p$. As mentioned earlier, by the Bonami Lemma [O'D14], we have that when $\lambda = (1-2p)^k$ then $||\mathcal{U}_{true}||_{2\to 4}^2 \leq 3^k$. Thus we have:

$$\Phi_{T_{p,\epsilon}^{sparse}}(S) \geq 1 - (1-2p)^k - 3^k\sqrt{\delta} - 2\epsilon$$

Thus if we want expansion at least $1-\gamma$ for some small $\gamma$, we can set $\epsilon < \frac{\gamma}{6}$, $k > O\left(\frac{\ln 1/\gamma}{p}\right)$, and $\delta < \gamma^{O\left(\frac{1}{p}\right)}$.

## 5 Lower Bounds and Discussion

A natural question is how the size of our construction compares to an optimal, possibly nonexplicit construction. We first note that a simple probabilistic argument shows that any collection of $2^n$ tests from $\{0,1\}^n$ to $\{0,1\}$ under the uniform distribution can be $\epsilon$-fooled by some function $G : \{0,1\}^s \to \{0,1\}^n$ for $s = \log n + 2\log(1/\epsilon) + O(1)$. The probabilistic construction is to simply pick each output of $G$ independently and uniformly at random from $\{0,1\}^n$. Using an analogous argument, picking each output of $G$ independently from $\mathcal{D}_p$ shows that there is a distribution $\mathcal{Z}$ using the same seed length $s$ that fools all $2^n$ linear tests under $\mathcal{D}_p$. Thus, non-explicitly there exists a construction of an $(p,\epsilon)$-biased distribution whose size does not depend on $p$. Moreover, the distribution is *uniform* on its support, which is not the case for our explicit construction.

Alon et al [AGHP92] prove a lower bound of $\Omega\left(\frac{n}{\epsilon^2 \log 1/\epsilon}\right)$ on the size of $\epsilon$-biased sets. We note that as a whole, since our construction works for $p = 1/2$ this lower bound is also a lower bound in general for $\epsilon$-biased sets for $\mathcal{D}_p$. However, the story changes dramatically for small $p$. The previously mentioned lower bound is a result on the equivalence of $\epsilon$-biased sets with $\epsilon$-balanced linear error correcting codes. In an $\epsilon$-balanced linear error correcting codes with message length $n$ and block length $m$, every codeword has weight between $(1/2-\epsilon)m$ and $(1/2+\epsilon)m$. The equivalence between such codes and $\epsilon$-biased sets breaks down when generalizing to $(p,\epsilon)$-biased sample spaces. Under the assumption that we wish to construct an $(p,\epsilon)$-biased distribution for $\mathcal{D}_{p,n}$ of size $m$ that is uniform on its support, we would require a linear error correcting code with basis $\mathbf{a}_1,\ldots,\mathbf{a}_n \in \{0,1\}^m$ such that the weight of the codeword $\sum_I \mathbf{a}_I$ for every $I \subset [n]$ is between $\frac{1}{2} - \frac{1}{2}(1-2p)^{|I|} - \epsilon$ and $\frac{1}{2} - \frac{1}{2}(1-2p)^{|I|} + \epsilon$.

We note that our construction worsens in comparison to the optimal as $p$ gets small. Indeed, as $p$ approaches $1/n$, the amount of entropy in $\mathcal{D}_p$ approaches 1, however, our seed length approaches $\log^2 n$. Thus, our construction illuminates a peculiar question about simulating a unfair coin: in order to simulate a coin with bias $p$, we require $\log\frac{1}{p}$ flips of a fair coin, or in other words $\log\frac{1}{p}$ bits of Shannon entropy. This is an extremely wasteful amount of randomness needed to simulate a distribution that has only $H(p) \ll 1$ bits of entropy. However, it is unclear how to simulate an unfair coin using fair coins in a more efficient way. We note that the reverse direction of simulating a fair coin with a biased coin is a well known riddle attributed to von Neumann [von63].

One reason that the efficiency of our construction depends on $p$ is because of an asymmetry between the nature of the seed and the output. We aimed to use $O(\log n)$ independent *fair* coin flips to approximate the distribution of $n$ independent *unfair* coin flips. A more apt comparison would be to stretch $O(\log n)$ unfair coins to approximate $n$ unfair coins. It would be interesting to see whether there are simple constructions that can do so.

# References

[ABS15] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. *J. ACM*, 62(5):42:1–42:25, 2015.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[BGH$^+$11] Boaz Barak, Parikshit Gopalan, Johan Håstad, Raghu Meka, Prasad Raghavendra, and David Steurer. Making the long code shorter, with applications to the unique games conjecture. *CoRR*, abs/1111.0405, 2011.

[BSSVW03] Ben-Sasson, Sudan, Vadhan, and Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *STOC: ACM Symposium on Theory of Computing (STOC)*, 2003.

[EGL$^+$92] Even, Goldreich, Luby, Nisan, and Velickovic. Approximations of general independent distributions. In *STOC: ACM Symposium on Theory of Computing (STOC)*, 1992.

[GMRZ13] Parikshit Gopalan, Raghu Meka, Omer Reingold, and David Zuckerman. Pseudorandom generators for combinatorial shapes. *SIAM J. Comput*, 42(3):1051–1076, 2013.

[Hås96] Johan Håstad. Testing of the long code and hardness for clique. In *Proceedings of The Twenty-Eighth Annual ACM Symposium On The Theory Of Computing (STOC '96)*, pages 11–19, New York, USA, May 1996. ACM Press.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput*, 37(1):319–357, 2007.

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *CoRR*, abs/math/0503503, 2005.

[MRT18] Raghu Meka, Omer Reingold, and Avishay Tal. Pseudorandom generators for width-3 branching programs. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:112, 2018.

[NN93] Naor and Naor. Small-bias probability spaces: Efficient constructions and applications. *SICOMP: SIAM Journal on Computing*, 22, 1993.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[ST11] Daniel A. Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM J. Comput*, 40(4):981–1025, 2011.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:41, 2017.

[von63] J. von Neumann. Various techniques for use in connection with random digits. In *von Neumann's Collected Works*, volume 5, pages 768–770. Pergamon, 1963.

# Appendix A Omitted Proofs

*Proof of Proposition 6.* We note that $H^{-1} = 2^{-n}H^T$ and show that $H\mathbf{p} = \mathbf{b}$. For any fixed entry of $\mathbf{b}$, we have:

$$\begin{aligned} b_{\boldsymbol{\alpha},\mathcal{D}} &= Pr_{\mathbf{a}\sim\mathcal{D}}\left[\langle\boldsymbol{\alpha},\mathbf{a}\rangle = 0\right] - Pr_{\mathbf{a}\sim\mathcal{D}}\left[\langle\boldsymbol{\alpha},\mathbf{a}\rangle = 1\right] \\ &= \sum_{a:\langle\boldsymbol{\alpha},\mathbf{a}\rangle=0} p_{\mathbf{a},\mathcal{D}} - \sum_{a:\langle\boldsymbol{\alpha},\mathbf{a}\rangle=1} p_{\mathbf{a},\mathcal{D}} \\ &= \sum_{a}(-1)^{\langle\boldsymbol{\alpha},\mathbf{a}\rangle} p_{\mathbf{a},\mathcal{D}} \\ &= (H\mathbf{p})_{\boldsymbol{\alpha}} \end{aligned}$$

□

*Proof of Corollary 7.* For any $\boldsymbol{\alpha} \in \{0,1\}^n$, we have that:

$$\begin{aligned} &|b_{\boldsymbol{\alpha},\mathcal{Z}} - b_{\boldsymbol{\alpha},\mathcal{D}_p}| \\ &= |P_{\mathbf{z}\sim\mathcal{Z}}[\langle\boldsymbol{\alpha},\mathbf{z}\rangle = 0] - Pr_{\mathbf{z}\sim\mathcal{Z}}[\langle\boldsymbol{\alpha},\mathbf{z}\rangle = 1] - (P_{\mathbf{z}\sim\mathcal{D}_p}[\langle\boldsymbol{\alpha},\mathbf{z}\rangle = 0] - (P_{\mathbf{z}\sim\mathcal{D}_p}[\langle\boldsymbol{\alpha},\mathbf{z}\rangle = 1])| \\ &\le |P_{\mathbf{z}\sim\mathcal{Z}}[\langle\boldsymbol{\alpha},\mathbf{z}\rangle = 0] - P_{\mathbf{z}\sim\mathcal{D}_p}[\langle\boldsymbol{\alpha},\mathbf{z}\rangle = 0]| + |Pr_{\mathbf{z}\sim\mathcal{Z}}[\langle\boldsymbol{\alpha},\mathbf{z}\rangle = 1] - (P_{\mathbf{z}\sim\mathcal{D}_p}[\langle\boldsymbol{\alpha},\mathbf{z}\rangle = 1])| \\ &\le 2\epsilon. \end{aligned}$$

Fix any $\mathbf{a} \in \{0,1\}^n$. By the formula from Proposition 6, we know that:

$$p_{\mathbf{a},\mathcal{Z}} = 2^{-n} \sum_{\boldsymbol{\alpha}\in\{0,1\}^n} (-1)^{\langle\mathbf{a},\boldsymbol{\alpha}\rangle} b_{\boldsymbol{\alpha},\mathcal{Z}}$$

Similarly:

$$p_{\mathbf{a},\mathcal{D}} = 2^{-n} \sum_{\boldsymbol{\alpha}\in\{0,1\}^n} (-1)^{\langle\mathbf{a},\boldsymbol{\alpha}\rangle} b_{\boldsymbol{\alpha},\mathcal{D}}$$

Thus:

$$\begin{aligned}
&|p_{\mathbf{a},\mathcal{Z}} - p_{\mathbf{a},\mathcal{D}}| \\
&= 2^{-n} \left| \sum_{\boldsymbol{\alpha}\in\{0,1\}^n} (-1)^{\langle \mathbf{a},\boldsymbol{\alpha}\rangle} (b_{\boldsymbol{\alpha},\mathcal{Z}} - b_{\boldsymbol{\alpha},\mathcal{D}}) \right| \\
&\leq 2^{-n} \sum_{\boldsymbol{\alpha}\in\{0,1\}^n} |(b_{\boldsymbol{\alpha},Z} - b_{\boldsymbol{\alpha},\mathcal{D}})| \\
&\leq 2\epsilon
\end{aligned}$$

□

*Proof of Proposition 8.* Let $X_i$ be a random variable with value $-1$ if the $i$th coin toss is heads and 1 otherwise. Then the probability of an odd number of heads is equal to $Pr[\prod_{i=1}^k X_i = -1]$. Note that the random variable $\frac{1}{2} + \frac{1}{2}\prod_{i=1}^k X_i$ is an indicator random variable that is 1 when there is an even number of heads. Thus

$$Pr(\text{even number of heads}) = \mathbb{E}\left[\frac{1}{2} + \frac{1}{2}\prod_{i=1}^k X_i\right] = \frac{1}{2} + \frac{1}{2}\prod_{i=1}^k \mathbb{E}[X_i] = \frac{1}{2} + \frac{1}{2}(1-2p)^k$$

Thus the probability of an odd number of heads is

$$\frac{1}{2} - \frac{1}{2}(1-2p)^k$$

□

*Proof of Lemma 9.* We wish to show that for any $\boldsymbol{\alpha} \in \{0,1\}^n$ with $|\boldsymbol{\alpha}| \leq k$:

$$\left| Pr_{\mathbf{x}\sim\mathbf{X}}[\langle \boldsymbol{\alpha}, \mathbf{x}\rangle = 1] - Pr_{\mathbf{x}\sim\mathcal{D}_{\frac{1}{2^t}}}[\langle \boldsymbol{\alpha}, \mathbf{x}\rangle = 1] \right| \leq t\epsilon$$

We prove this via a hybrid argument.

Consider random variables $\mathbf{X}_1, \ldots, \mathbf{X}_t, \mathbf{R}_1, \ldots, \mathbf{R}_t$ taking on values in $\{0,1\}^n$ where the $X_i$'s are independent draws from a $k$-wise $\epsilon$-biased set, $R_i$'s are chosen independently and uniformly at random from $\{0,1\}^n$. We then define for $0 \leq \ell \leq t$ the $t+1$ hybrid distributions:

$$\mathcal{H}_\ell = \left\langle \boldsymbol{\alpha}, \left( \bigodot_{i=1}^{\ell} \mathbf{R}_i \right) \odot \left( \bigodot_{i=\ell+1}^{t} \mathbf{X}_i \right) \right\rangle$$

Notice that $\mathcal{H}_0 = \langle \boldsymbol{\alpha}, \mathbf{x}\rangle$ when $\mathbf{x} \sim \mathbf{X}$, while $\mathcal{H}_{t+1} = \langle \boldsymbol{\alpha}, \mathbf{x}\rangle$, when $x \sim \mathcal{D}_p$. We show that $|\mathcal{H}_\ell - \mathcal{H}_{\ell+1}| \leq \epsilon$ for every $0 \leq \ell \leq t$. Since each $\mathcal{H}_\ell$ is a distribution on $\{0,1\}$ we can write the probability that distribution $\mathcal{H}_\ell$ outputs 1 as:

$$
\begin{aligned}
&\underset{\substack{\mathbf{R}_1,\ldots,\mathbf{R}_\ell,\\ \mathbf{X}_{\ell+1},\ldots,\mathbf{X}_t}}{Pr}\left[\left\langle \boldsymbol{\alpha}, \left(\bigodot_{i=1}^{\ell}\mathbf{R}_i\right)\odot\left(\bigodot_{i=\ell+1}^{t}\mathbf{X}_i\right)\right\rangle = 1\right]\\
&= \underset{\substack{\mathbf{R}_1,\ldots,\mathbf{R}_\ell,\\ \mathbf{X}_{\ell+1},\ldots,\mathbf{X}_t}}{Pr}\left[\left\langle \boldsymbol{\alpha}, \left(\bigodot_{i=1}^{\ell}\mathbf{R}_i\right)\odot\mathbf{X}_{\ell+1}\odot\left(\bigodot_{i=\ell+2}^{t}\mathbf{X}_i\right)\right\rangle = 1\right]\\
&= \underset{\substack{\mathbf{R}_1,\ldots,\mathbf{R}_\ell,\\ \mathbf{X}_{\ell+1},\ldots,\mathbf{X}_t}}{Pr}\left[\left\langle \boldsymbol{\alpha}, \left(\bigodot_{i=1}^{\ell}\mathbf{R}_i\right)\odot\left(\bigodot_{i=\ell+2}^{t}\mathbf{X}_i\right)\odot\mathbf{X}_{\ell+1}\right\rangle = 1\right]\\
&= \underset{\substack{\mathbf{R}_1,\ldots,\mathbf{R}_\ell,\\ \mathbf{X}_{\ell+1},\ldots,\mathbf{X}_t}}{Pr}\left[\left\langle \boldsymbol{\alpha}\odot\left(\bigodot_{i=1}^{\ell}\mathbf{R}_i\right)\odot\left(\bigodot_{i=\ell+2}^{t}\mathbf{X}_i\right), \mathbf{X}_{\ell+1}\right\rangle = 1\right]\\
&= \underset{\substack{\mathbf{R}_1,\ldots,\mathbf{R}_\ell,\\ \mathbf{X}_{\ell+2},\ldots,\mathbf{X}_t}}{\mathbb{E}}\left[Pr_{\mathbf{X}_{\ell+1}}\left[\left\langle \boldsymbol{\alpha}\odot\left(\bigodot_{i=1}^{\ell}\mathbf{R}_i\right)\odot\left(\bigodot_{i=\ell+2}^{t}\mathbf{X}_i\right), \mathbf{X}_{\ell+1}\right\rangle = 1\right]\right]
\end{aligned}
$$

Where the last equality makes use of the fact that all the $\mathbf{X}_i$'s and $\mathbf{R}_i$'s are independent from each other. Similarly, we can write the probability that $H_{\ell+1}$ outputs 1 as:

$$
\begin{aligned}
&\underset{\substack{\mathbf{R}_1,\ldots,\mathbf{R}_{\ell+1},\\ \mathbf{X}_{\ell+2},\ldots,\mathbf{X}_t}}{Pr}\left[\left\langle \boldsymbol{\alpha}, \left(\bigodot_{i=1}^{\ell+1}\mathbf{R}_i\right)\odot\left(\bigodot_{i=\ell+2}^{t}\mathbf{X}_i\right)\right\rangle = 1\right]\\
&= \underset{\substack{\mathbf{R}_1,\ldots,\mathbf{R}_\ell,\\ \mathbf{X}_{\ell+2},\ldots,\mathbf{X}_t}}{\mathbb{E}}\left[Pr_{\mathbf{R}_{\ell+1}}\left[\left\langle \boldsymbol{\alpha}\odot\left(\bigodot_{i=1}^{\ell}\mathbf{R}_i\right)\odot\left(\bigodot_{i=\ell+2}^{t}\mathbf{X}_i\right), \mathbf{R}_{\ell+1}\right\rangle = 1\right]\right]
\end{aligned}
$$

For fixed $\mathbf{R}_1, \ldots, \mathbf{R}_\ell$ and $\mathbf{X}_{\ell+2}, \ldots, \mathbf{X}_t$, we know that $\boldsymbol{\beta} = \boldsymbol{\alpha}\odot\left(\bigodot_{i=1}^{\ell}\mathbf{R}_i\right)\odot\left(\bigodot_{i=\ell+2}^{t}\mathbf{X}_i\right)$ is a vector with at most $k$ 1's. Thus since $\mathbf{X}_{\ell+1}$ is $k$-wise $\epsilon$-biased, we know that:

$$
\left|Pr_{\mathbf{X}_{\ell+1}}\left[\langle\boldsymbol{\beta}, \mathbf{R}_{\ell+1}\rangle = 1\right] - Pr_{\mathbf{X}_{\ell+1}}\left[\langle\boldsymbol{\beta}, \mathbf{X}_{\ell+1}\rangle = 1\right]\right| \leq \epsilon/2
$$

Since expectation is just a weighted average, and each $\mathcal{H}_\ell$ is a distribution over $\{0,1\}$, we can conclude that $|\mathcal{H}_\ell - \mathcal{H}_{\ell+1}| \leq \epsilon/2$. Combining all the hybrid steps via triangle inequality gives us that $|\mathcal{H}_0 - \mathcal{H}_\ell| \leq t\epsilon/2$ □