Electronic Colloquium on Computational Complexity, Report No. 3 (2025)

Fooling Near-Maximal Decision Trees

William M. Hoza
Department of Computer Science
The University of Chicago
williamhoza@uchicago.edu

Abstract

For any constant $\alpha > 0$, we construct an explicit pseudorandom generator (PRG) that fools n-variate decision trees of size m with error ε and seed length $(1+\alpha) \cdot \log_2 m + O(\log(1/\varepsilon) + \log\log n)$. For context, one can achieve seed length $(2 + o(1)) \cdot \log_2 m + O(\log(1/\varepsilon) + \log\log n)$ using well-known constructions and analyses of small-bias distributions, but such a seed length is trivial when $m \geq 2^{n/2}$. By combining our new PRG with work by Chen and Kabanets (TCS 2016), we get an explicit PRG that fools circuits of size $2.99 \cdot n$ over the U_2 basis with error $2^{-\Omega(n)}$ and seed length $(1 - \Omega(1)) \cdot n$.

Our approach for fooling decision trees is to develop a new variant of the classic concept of almost k-wise independence, which might be of independent interest. We say that a distribution X over $\{0,1\}^n$ is *k-wise ε-probably uniform* if every Boolean function f that depends on only k variables satisfies $\mathbb{E}[f(X)] \geq (1-\varepsilon) \cdot \mathbb{E}[f]$. We show how to sample a k-wise ε-probably uniform distribution using a seed of length $(1+\alpha) \cdot k + O(\log(1/\varepsilon) + \log\log n)$.

1 Introduction

How many coin flips does it take to sample n bits that appear random from the perspective of an observer who only looks at $0.9 \cdot n$ of the bits?

1.1 Almost k-wise uniformity and k-wise probable uniformity

Almost k-wise uniformity is a well-studied concept that provides one possible way of formalizing the question posed above.

Definition 1.1 (Almost k-wise uniformity)**.** Let X be a distribution over $\{0,1\}^n$, let $k \in [n]$, and let $\varepsilon \in [0,1]$. We say that X is *ε-almost k-wise uniform* if, for every size-k set $S \subseteq [n]$, the total variation distance between X_S and U_k is at most ε. Here X_S denotes the projection of X to the coordinates in S, and U_k denotes the uniform distribution over $\{0,1\}^k$. If $\varepsilon = 0$, we simply say that X is *k-wise uniform*. An *(ε-almost) k-wise uniform generator* is a function $G\colon \{0,1\}^s \to \{0,1\}^n$ such that $G(U_s)$ is (ε-almost) k-wise uniform. We refer to s as the *seed length* of G.

When $k \geq (\frac{1}{2} + \Omega(1)) \cdot n$ and $\varepsilon = 0$, Karloff and Mansour showed that every k-wise uniform generator has seed length at least $n - O(1)$ [KM97], which might be disappointing. On the bright side, the seed length can be improved if a small positive error ($\varepsilon > 0$) is permitted. Using a connection with "small-bias distributions" [NN93], Alon, Goldreich, Håstad, and Peralta constructed an explicit[1] ε-almost k-wise uniform generator with seed length $k + O(\log(k/\varepsilon) + \log\log n)$ [AGHP92]. Notably, their seed length is meaningful even for large k such as $k = 0.9 \cdot n$.

In this work, we introduce a new variant of almost k-wise uniformity, called *k-wise probable uniformity*, which strengthens Definition 1.1. There are two equivalent definitions, described below.

[1] We consider a generator G to be *explicit* if $G(x)$ can be computed in $\text{poly}(n)$ time, given the parameters (in this case n, k, and ε) and the seed x.

ISSN 1433-8092

Definition 1.2 (k-wise probable uniformity)**.** Let X be a distribution over $\{0,1\}^n$, let $k \in [n]$, and let $\varepsilon \in [0,1]$. We say that X is *k-wise ε-probably uniform* if it satisfies either of the following two equivalent conditions.

1. For every size-k set $S \subseteq [n]$, there exists a distribution E over $\{0,1\}^k$ such that the distribution X_S can be written as the mixture distribution $X_S \equiv (1-\varepsilon) \cdot U_k + \varepsilon \cdot E$. That is, the distribution X_S is identical to the following distribution: With probability $1-\varepsilon$, sample a k-bit string uniformly at random, and with probability ε, sample a string according to E.

2. For every k-junta[2] $f\colon \{0,1\}^n \to \{0,1\}$, we have

$$\mathbb{E}[f(X)] \geq (1-\varepsilon) \cdot \mathbb{E}[f],$$

   where $\mathbb{E}[f]$ is a shorthand for $\mathbb{E}[f(U_n)]$.

(See Section 3 for a proof that the two conditions above are equivalent.) We say that $G\colon \{0,1\}^s \to \{0,1\}^n$ is a *k-wise ε-probably uniform generator* if $G(U_s)$ is k-wise ε-probably uniform.

We find the first condition above to be more conceptually appealing. It is clearly a strengthening of ε-almost k-wise uniformity, and it inspires the terminology "k-wise ε-probably uniform." On the other hand, we find the second condition above to be easier to work with mathematically.

The concept of k-wise probable uniformity is motivated primarily by an application to fooling decision trees, which we will discuss momentarily, but we also consider it to be an interesting concept in its own right. Using a standard nonconstructive argument (see Proposition 4.6), one can show that there exists a non-explicit k-wise ε-probably uniform generator with seed length[3]

$$k + \log k + 2\log(1/\varepsilon) + \log\log(n/k) + O(1). \tag{1}$$

The challenge is to construct an *explicit* generator.

Classic results regarding small-bias generators [NN93; AGHP92] imply that there is an explicit k-wise ε-probably uniform generator with seed length $2k + O(\log k + \log(1/\varepsilon) + \log\log n)$. However, this seed length is unsatisfactory, because it is trivial when $k \geq n/2$. Meanwhile, Bshouty used a different approach (the method of conditional probabilities with pessimistic estimators) to construct a generator $G\colon \{0,1\}^s \to \{0,1\}^n$ such that

$$(1-\varepsilon) \cdot \mathbb{E}[f] \leq \mathbb{E}[f(G(U_s))] \leq (1+\varepsilon) \cdot \mathbb{E}[f]$$

for every Boolean k-junta f [Bsh16], which is even stronger than Definition 1.2. Furthermore, his generator's seed length matches Eq. (1). However, his generator's time complexity is more than $\binom{n}{k} \cdot 2^k$ [Bsh16]. His generator can therefore be considered "explicit" only when $k = O(1)$, whereas we are primarily interested in the case $k = \Theta(n)$.

In this work, we present an explicit k-wise ε-probably uniform generator with seed length $(1+\alpha) \cdot k + O(\log(1/\varepsilon) + \log\log n)$, where α is an arbitrarily small positive constant.

Theorem 1.3 (Explicit k-wise probably uniform generator)**.** *For every $n, k \in \mathbb{N}$ and $\varepsilon \in (0,1)$, there exists an explicit k-wise ε-probably uniform generator $G\colon \{0,1\}^s \to \{0,1\}^n$ with seed length*

$$s = k + O\left(k^{2/3} \cdot \log^{1/3}(k/\varepsilon) + \log(1/\varepsilon) + \log\log n\right).$$

The simpler seed length bound $(1+\alpha) \cdot k + O(\log(1/\varepsilon) + \log\log n)$ follows from Theorem 1.3 by the weighted AM-GM inequality.

[2] A k-junta is a function f that depends on at most k variables.

[3] Throughout this paper, $\log(\cdot)$ denotes the base-two logarithm.

1.2 Fooling decision trees

Instead of modeling the observer as a k-junta, we can consider the more powerful model of *depth-k decision trees*. A decision tree T makes queries to the input x and then produces a Boolean output value $T(x)$. The crucial feature of the decision tree model is that the tree can *adaptively* decide which variable to query next, based on the results of previous queries. (See Definition 2.1 for a precise definition.) Consequently, the output $T(x)$ of a depth-k decision tree T might depend on all n variables even if $k \ll n$. The problem of sampling bits that "appear random" to depth-k decision trees can be formalized using the concept of a *pseudorandom generator*.

Definition 1.4 (Pseudorandom generators)**.** Let X be a distribution over $\{0,1\}^n$, let $f\colon \{0,1\}^n \to \{0,1\}$, and let $\varepsilon \in (0,1)$. We say that X *fools f with error ε* if

$$|\,\mathbb{E}[f(X)] - \mathbb{E}[f]| \leq \varepsilon.$$

We say that $G\colon \{0,1\}^s \to \{0,1\}^n$ is a *pseudorandom generator (PRG) that fools f with error ε* if $G(U_s)$ fools f with error ε. The parameter s is called the *seed length* of the PRG. If $\mathcal{F}$ is a class of functions $f\colon \{0,1\}^n \to \{0,1\}$, we say that X (respectively G) fools $\mathcal{F}$ with error ε if X (respectively G) fools every $f \in \mathcal{F}$ with error ε.

Almost k-wise uniformity is the special case of Definition 1.4 in which we take $\mathcal{F}$ to be the class of all Boolean k-juntas. The aforementioned concept of small-bias distributions is another special case. By definition, a distribution X is *k-wise γ-biased* if it fools all functions of the form $f(x) = \bigoplus_{i \in S} x_i$, where $S \subseteq [n]$ and $|S| \leq k$, with error $\gamma/2$ [NN93].

To fool decision trees, one could try using a generic small-bias generator. This approach works extremely well in the nonadaptive setting, as mentioned previously. In the adaptive setting, the approach still works fairly well, but it turns out that *the parameters are worse*. Specifically, Kushilevitz and Mansour's analysis [KM93] implies that if X is k-wise γ-biased, then X fools depth-k size-m decision trees with error $\gamma \cdot m$. Every depth-k decision tree has size at most 2^k, so we can choose $\gamma = \varepsilon \cdot 2^{-k}$. By combining this reduction with one of Alon, Goldreich, Håstad, and Peralta's k-wise γ-biased generators [AGHP92], one can construct an explicit PRG that fools depth-k decision trees with error ε and seed length $2k + O(\log(k/\varepsilon) + \log\log n)$. This seed length is sufficient for many purposes, but we emphasize that it gives us nothing nontrivial for trees of depth $k \geq n/2$.

In this paper, we show how to improve the leading constant from 2 to $1 + \alpha$ for any constant $\alpha > 0$, as a consequence of our new k-wise ε-probably uniform generator. More generally, we prove the following.

Theorem 1.5 (Fooling near-maximal decision trees)**.** *Let $n, m \in \mathbb{N}$ and $\varepsilon \in (0,1)$. There exists an explicit PRG $G\colon \{0,1\}^s \to \{0,1\}^n$ that fools n-variate decision trees of size m with error ε and seed length*

$$s = \log m + O\left(\log^{2/3} m \cdot \log^{1/3}\left(\frac{\log m}{\varepsilon}\right) + \log(1/\varepsilon) + \log\log n\right).$$

Observe that our PRG is meaningful even for trees of near-maximal size such as $m = 2^{0.9 \cdot n}$. Furthermore, it turns out that Theorem 1.5 extends to the more powerful model of size-m "subcube partitions." See Section 5 for further details.

1.3 Application: Fooling U_2-circuits of size $(3-\alpha)\cdot n$

In general, the motivation behind PRGs is that many algorithms and protocols rely on a large number of random bits, but producing truly random bits can sometimes be difficult or expensive. We think of randomness as a computational resource, similar to time or space. We try to use as little "true randomness" as possible to sample bits that are "random enough" to run randomized algorithms and protocols without distorting their behavior.

The specific problem of fooling near-maximal decision trees is motivated by an application in the area of *circuit complexity*. Arguably, the central challenge of complexity theory is to understand the power of general Boolean circuits. Unfortunately, our current understanding of circuits is extremely meager. Indeed, circuit lower bound proofs are often so weak that they are sensitive to the specific choice of gate basis. In this paper, we focus on the U_2 *basis*, consisting of all functions $\phi\colon \{0,1\}^2 \to \{0,1\}$ other than the XOR function and its complement. A "U_2-circuit" is a circuit in which each gate computes a function from the U_2 basis. Chen and Kabanets used "gate elimination" methods to prove that every U_2-circuit of size $(3-\alpha)\cdot n$ can be computed by a decision tree of size $2^{(1-\Omega(\alpha^2))\cdot n}$, among other results [CK16]. They posed the problem of designing PRGs that fool general Boolean circuits [CK16]. By combining their simulation with our construction, we are able to solve their PRG problem, at least for the case of U_2-circuits of size $(3-\alpha)\cdot n$:

Corollary 1.6 (Fooling circuits over the U_2 basis)**.** *For every $n \in \mathbb{N}$ and $\alpha \in (0,3)$, there exists an explicit PRG $G\colon \{0,1\}^s \to \{0,1\}^n$ that fools n-variate U_2-circuits of size $(3-\alpha)\cdot n$ with error $n \cdot 2^{-\Omega(\alpha^6 n)}$ and seed length $s = (1-\Omega(\alpha^2))\cdot n$.*

Proof of Corollary 1.6, given Theorem 1.5 and Chen and Kabanets' work [CK16]. By Theorem 1.5, we can fool decision trees of size $2^{(1-c\alpha^2)\cdot n}$ with error $2^{-c'\alpha^6 n}\cdot n$ and seed length

$$(1-c\alpha^2)\cdot n + O(n^{2/3}\cdot (c'\alpha^6 n)^{1/3} + c'\alpha^6 n) = n - c\alpha^2 n + O(c'\alpha^2 n).$$

This is $n - \Omega(\alpha^2 n)$ provided we choose c' to be a sufficiently small constant based on c. □

The PRG of Corollary 1.6 is the first of its kind.[4] Note that the challenge of constructing PRGs that fool Boolean circuits is strictly harder than the challenge of proving circuit lower bounds. In more detail, suppose that one could construct a poly(n)-time computable PRG $G\colon \{0,1\}^{\beta n - 1} \to \{0,1\}^n$ that fools U_2-circuits of size cn with error 0.49 for infinitely many n, where $\beta \in [0,1]$ and $c > 1$ are constants. Then by truncating the output of G, one could construct a poly(n)-time computable PRG $G'\colon \{0,1\}^{n-1} \to \{0,1\}^n$ that fools U_2 circuits of size $(c/\beta)\cdot n$ with error 0.49 for infinitely many n. The indicator function for the image of G' would be an example of a function in $h \in \mathsf{NP}$ that cannot be computed by U_2-circuits of size $(c/\beta)\cdot n$. Currently, the best lower bound known on the size of U_2-circuits computing some function in NP is $(5-o(1))\cdot n$ [IM02].

1.4 Overview of our new construction

In this section, we present an informal overview of our new k-wise probably uniform generator (Theorem 1.3). The starting point of the construction is the well-known sampling properties of *pairwise uniform hash functions*. Let $f\colon \{0,1\}^n \to \{0,1\}$ be any nonzero k-junta, or more generally any function such that $\mathbb{E}[f] \geq 2^{-k}$. If we sample a hash function $h\colon \{0,1\}^{k+O(\log(1/\varepsilon))} \to \{0,1\}^n$ from a pairwise uniform family, then with high probability over the choice of h, we have

$$\mathop{\mathbb{E}}_{x}[f(h(x))] \geq (1-\varepsilon)\cdot \mathbb{E}[f].$$

(This follows from Chebyshev's inequality.)

We can think of h as a PRG with an excellent seed length. The only trouble is that sampling h itself is expensive. In general, sampling a hash function $h\colon \{0,1\}^q \to \{0,1\}^\ell$ from a pairwise uniform family costs $\Theta(q+\ell)$ truly random bits, so in our case, the cost is $\Theta(n+\log(1/\varepsilon))$ truly random bits, which is much more than we can afford.

[4] To be fair, we should compare Corollary 1.6 to a different and rather trivial approach that one could use to construct PRGs that fool circuits. In general, if $h\colon \{0,1\}^{n-1} \to \{0,1\}$ is average-case hard for circuits of size cn, then the generator $G(x) = (x, h(x))$ maps $n-1$ bits to n bits and fools circuits of size cn. Similarly, the generator $G'(x,y) = (x,y,h(x),h(y))$ maps $n'-2$ bits to n' bits and fools circuits of size $(c/2)\cdot n'$, where $n' = 2n$. One can similarly try $G''(x,y,z) = (x,y,z,h(x),h(y),h(z))$, etc. One can instantiate this approach with known average-case hardness results for circuits over the U_2 basis or the full binary basis [CK16; GKST18]. However, the PRGs that can be constructed using this approach have seed length $n - O(1)$. The seed length is what makes Corollary 1.6 interesting. If α is constant, then our PRG has linear stretch.

We can slightly decrease the cost of sampling h by composing with a γ-almost k-wise uniform generator, where $\gamma \approx \varepsilon \cdot 2^{-k}$, with seed length $\ell = O(k + \log(1/\varepsilon) + \log\log n)$. Such a generator fools f with error γ, which is negligible. Now the output length of h is decreased from n down to ℓ, hence the cost of sampling h is "only" $O(k + \log(1/\varepsilon) + \log\log n)$. However, this cost is still more than we can afford.

To explain how we bring the cost down to $o(k)$, for simplicity's sake, let us assume that $\varepsilon = 1/\operatorname{poly}(k)$ and let us neglect $\log\log n$ terms. We can assume without loss of generality that f is simply a conjunction of k literals, because every k-junta can be written as a sum of such functions. Our approach is to pseudorandomly partition the n coordinates into $r = \widetilde{\Theta}(k^{1/3})$ buckets: $[n] = B_1 \cup \cdots \cup B_r$. In expectation, each bucket contains k/r of the k relevant variables. With high probability, each bucket has at most k_0 of the variables, where $k_0 = k/r + \widetilde{O}(\sqrt{k/r}) = k/r + \widetilde{O}(k^{1/3})$.

We can write $f(x) = f_1(x) \wedge \cdots \wedge f_r(x)$, where $f_i(x)$ only depends on variables in B_i, so f_i is a k_0-junta. We sample a hash function $h\colon \{0,1\}^{k_0+O(\log k)} \to \{0,1\}^n$ such that with high probability over the choice of h, we have

$$\mathop{\mathbb{E}}_{x}[f_i(h(x))] \geq \left(1 - \frac{1}{\operatorname{poly}(k)}\right) \cdot \mathbb{E}[f_i].$$

For each bucket B_i independently, we sample x at random and put $h(x)$ in B_i. Crucially, *we reuse the same hash function h* for all of the buckets, which is justified by a simple union bound. The cost of sampling h is $O(k_0) = \widetilde{O}(k^{2/3})$ truly random bits, and the cost of sampling the x values is

$$r \cdot (k_0 + O(\log k)) = k + \widetilde{O}(k^{2/3}).$$

A more careful calculation, also taking into account the cost of sampling the partition $[n] = B_1 \cup \cdots \cup B_r$, leads to the seed length bound that appears in Theorem 1.3.

Observe that in this construction, there are some "bad events" that occur with probability roughly ε, namely, we might get a "bad" partition of the variables into buckets or we might get a "bad" hash function h. Let B be the union of these bad events. To analyze the impact of these bad events, let X be the output distribution of our generator and let f be an arbitrary Boolean k-junta. Then

$$\mathbb{E}[f(X)] = \underbrace{\Pr[B] \cdot \mathbb{E}[f(X) \mid B]}_{(*)} + \Pr[\neg B] \cdot \mathbb{E}[f(X) \mid \neg B].$$

The quantity marked $(*)$ is certainly nonnegative, which allows us to prove $\mathbb{E}[f(X)] \geq (1-\varepsilon) \cdot \mathbb{E}[f]$. On the other hand, note that the quantity marked $(*)$ might be much larger than $\mathbb{E}[f]$, and hence we are not able to prove an upper bound of the form $\mathbb{E}[f(X)] \leq (1+\varepsilon) \cdot \mathbb{E}[f]$. Thankfully, such an upper bound is not necessary for our applications.

1.5 Limitations of k-wise γ-biased generators

A great deal of effort has been spent trying to optimize the constant factors in the seed lengths of small-bias generators [NN93; ABNNR92; AGHP92; BT13; Bsh16; Ta-17; BD22]. Researchers have also developed many sophisticated techniques for proving that small-bias generators fool various models of computation; see Hatami and Hoza's survey for a few examples [HH24]. The reader might reasonably wonder whether one could have proven our results by simply improving known constructions or analyses of k-wise γ-biased distributions. We prove that the answer is no. In more detail, in Section 6, we present examples showing that:

- If every k-wise γ-biased distribution is t-wise 0.49-probably uniform, then $k \geq t$ and $\gamma \leq O(2^{-t})$.
- If every k-wise γ-biased distribution fools decision trees of depth $0.76 \cdot n$ with error 0.49, then $k \geq 0.76 \cdot n$ and $\gamma \leq O(2^{-n/2})$.
- If every k-wise γ-biased distribution fools U_2-circuits of size $2n$ with error 0.49, then $k \geq \frac{2}{3} \cdot n$ and $\gamma \leq O(2^{-n/2})$.

Then, we observe that Karloff and Mansour's work [KM97] can be extended to prove the following lower bound on the seed length of k-wise γ-biased generators in the regime $k \geq (\frac{1}{2} + \Omega(1)) \cdot n$.

Theorem 1.7 (Seed length lower bound for k-wise γ-biased generators)**.** *Let $G\colon \{0,1\}^s \to \{0,1\}^n$ be a k-wise γ-biased generator, where $k = \lfloor (1/2 + \alpha) \cdot n \rfloor$ for some $\alpha \in (0, 1/2]$. Then*

$$s \geq \min\{n, 2\log(1/\gamma)\} - \log(1/\alpha) - O(1).$$

Consequently, if one tries using a generic k-wise γ-biased generator to construct a $(0.51 \cdot n)$-wise probably uniform generator, or to fool decision trees of depth $0.76 \cdot n$, or to fool U_2-circuits of size $2n$, then the seed length will inevitably be at least $n - O(1)$. Thus, the concept of k-wise γ-biased distributions is inherently too weak to prove Theorems 1.3 and 1.5 and Corollary 1.6.

For context, a sequence of prior works [Rao47; CGHFRS85; ABI86; AGHP92; Alo09; AAKMRX07; Bsh16] has shown that every k-wise γ-biased generator $G\colon \{0,1\}^s \to \{0,1\}^n$ has seed length at least

$$\min\left\{\log\left(\binom{n}{\leq k/2}\right),\quad 2\log(1/\gamma) + \log\log\left(\binom{n}{\leq k/2}\right) - \log\log(1/\gamma)\right\} - O(1). \tag{2}$$

Eq. (2) and Theorem 1.7 are incomparable in general, but our new Theorem 1.7 is superior in the parameter regime in which we are interested. In particular, if $\gamma = O(2^{-n/2})$ and $k = cn$ for a constant $1/2 < c < 1$, then the prior bound Eq. (2) is $(1 - \Omega(1)) \cdot n$, whereas our new Theorem 1.7 gives a bound of $n - O(1)$.

1.6 Related work

1.6.1 Approximate forms of k-wise uniformity

Prior researchers have studied several different ways of quantifying what it means for a distribution X over $\{0,1\}^n$ to be "approximately" k-wise uniform.

- We could require that the total variation distance between X_S and U_k is at most ε for every size-k set $S \subseteq [n]$. This is the definition of an ε-almost k-wise uniform distribution (Definition 1.1). See, for example, work by Naor and Naor [NN93] and work by Alon, Goldreich, Håstad, and Peralta [AGHP92].
- We could require that $|\Pr[\bigoplus_{i\in S} X_i = 1] - \Pr[\bigoplus_{i\in S} X_i = 0]| \leq \varepsilon$ for every nonempty set $S \subseteq [n]$ of size at most k [NN93]. This is the definition of a k-wise ε-biased distribution. See, for example, the works mentioned above [NN93; AGHP92].
- We could require that the ℓ_∞ distance between X_S and U_k is at most ε for every size-k set $S \subseteq [n]$. See, for example, work by Alon, Goldreich, Håstad, and Peralta [AGHP92] and work by Bshouty [Bsh16].
- We could require that X is ε-close in total variation distance to some exactly k-wise uniform distribution X'. See, for example, work by Alon, Goldreich, and Mansour [AGM03]; work by Alon, Andoni, Kaufman, Matulef, Rubinfeld, and Xie [AAKMRX07]; and work by O'Donnell and Zhao [OZ18].

Despite the attention paid to all of the above variations, we seem to be the first to study the concept of k-wise probable uniformity.

1.6.2 Universal sets

A set $H \subseteq \{0,1\}^n$ is called *k-universal* if, for every nonzero k-junta $f\colon \{0,1\}^n \to \{0,1\}$, there exists $x \in H$ such that $f(x) = 1$. The concept of k-universal sets has been studied in many prior works going back more than half a century [KS73; CKMZ83; TW83; Alo86; SB88; BS88; ABNNR92; NN93; NSS95; Bsh14]. The best explicit construction, due to Naor, Schulman, and Srinivasan [NSS95], has cardinality $2^{k+O(\log^2 k)} \cdot \log n$. Our k-wise probably uniform generator was inspired by Naor, Schulman, and Srinivasan's universal set construction [NSS95]. The notion of k-wise probable uniformity can be considered a strengthening of

k-universality, because if X is k-wise probably uniform, then the support of X is k-universal. Consequently, Theorem 1.3 implies the existence of an explicit k-universal set with cardinality $2^{k+\widetilde{O}(k^{2/3})} \cdot \operatorname{polylog} n$, but this is inferior to Naor, Schulman, and Srinivasan's construction [NSS95].[5] Our construction also has similarities with a recent construction of a "biased" variant of universal sets by Harel, Hoza, Vardi, Evron, Srebro, and Soudry [HHVESS24].

1.6.3 PRGs based on pseudorandom partitions of the variables

The trick of pseudorandomly partitioning the variables into buckets is not new; similar tricks have been used in many prior PRG constructions. For a few examples that are especially similar to our work, see work by Meka and Zuckerman [MZ13], work by Lovett, Reingold, Trevisan, and Vadhan [LRTV09], and work by Gopalan, Kane, and Meka [GKM18].

1.6.4 Correlation bounds and #SAT algorithms for general circuit models

In general, PRGs are intimately related to *correlation bounds*, aka average-case hardness. Loosely speaking, correlation bounds are a prerequisite to designing PRGs. See, e.g., Hatami and Hoza's survey [HH24, Chapter 4] for further discussion. Chen and Kabanets proved the first correlation bounds for general, unbounded-depth circuit models [CK16], and our PRG for U_2-circuits uses their work, as mentioned previously. Golovnev, Kulikov, Smal, and Tamaki subsequently proved better correlation bounds [GKST18]. Both of these papers also designed #SAT algorithms, i.e., algorithms that count the number of satisfying assignments to a given circuit [CK16; GKST18] (see also work by Nurk [Nur09]).

1.7 Organization

After some preliminaries, in Section 3, we record some straightforward characterizations of k-wise probable uniformity. Then, in Section 4, we present the details of our k-wise probably uniform generator, following the outline in Section 1.4. In Section 5, we explain why k-wise probable uniformity is sufficient for fooling decision trees and the more general subcube partition model. In Section 6, we prove that k-wise γ-biased generators are too weak to prove our main results. Finally, we conclude in Section 7 with some suggested open problems.

2 Preliminaries

2.1 The decision tree model

Below we record the standard definition of a decision tree.

Definition 2.1 (Decision trees)**.** An n-variate *decision tree* is a rooted tree T in which each internal node is labeled with a variable from among $x_1, \ldots, x_n$; each internal node has two outgoing edges labeled 0 and 1; and each leaf is labeled either 0 or 1. The tree T computes a Boolean function $T\colon \{0,1\}^n \to \{0,1\}$ defined inductively as follows. If T consists of a single leaf labeled $b \in \{0,1\}$, then we define $T(x) \equiv b$. Otherwise, let x_i be the variable labeling the root node. Given an input $x \in \{0,1\}^n$, we start at the root node and traverse the outgoing edge labeled with the value x_i. This leads to a vertex u, which is the root of a subtree T'. Then we set $T(x) = T'(x)$. The *depth* of the tree is the length of the longest path from the root to a leaf. The *size* of the tree is the total number of leaves.

[5] A k-universal set H is typically considered "explicit" if the entire set can be computed in $\operatorname{poly}(|H|)$ time. Our set has stronger explicitness guarantees, which might possibly be of value, but note that Naor, Schulman, and Srinivasan already constructed a k-universal set of cardinality $2^{k+o(k)} \cdot \log n$ with similar explicitness guarantees [NSS95].

2.2 Pairwise uniform hashing

We rely on the standard notion of a *pairwise uniform hashing*, aka "strongly universal hashing," introduced in Carter and Wegman's seminal papers [CW79; WC81].

Definition 2.2 (Pairwise uniform families of hash functions)**.** A family $\mathcal{H}$ of hash functions $h\colon \{0,1\}^q \to \{0,1\}^\ell$ is called *pairwise uniform* if, for every two distinct $x, x' \in \{0,1\}^q$, if we sample $h \sim \mathcal{H}$, then $(h(x), h(x'))$ is distributed uniformly at random over $\{0,1\}^{2\ell}$.

Theorem 2.3 (Explicit pairwise uniform families of hash functions)**.** *For every $q, \ell \in \mathbb{N}$, there exists an explicit[6] pairwise uniform family $\mathcal{H}$ of hash functions $h\colon \{0,1\}^q \to \{0,1\}^\ell$ such that $h \in \mathcal{H}$ can be sampled using a seed of length $O(q+\ell)$.*

For example, if we define $h_{a,b}(x) = a * x + b$, where $*$ is convolution mod 2 and $+$ is bitwise XOR, then $\{h_{a,b} : a \in \{0,1\}^{q+\ell-1}$ and $b \in \{0,1\}^\ell\}$ is a pairwise uniform family [MNT93].

2.3 Small-bias distributions

We also rely on asymptotically optimal constructions of k-wise γ-biased generators, which were defined in Section 1.2.

Theorem 2.4 (Explicit k-wise γ-biased generators [NN93])**.** *For every $n, k \in \mathbb{N}$ and every $\gamma \in (0,1)$, there exists an explicit k-wise γ-biased generator $G\colon \{0,1\}^s \to \{0,1\}^n$ with seed length $O(\log(k/\gamma) + \log\log n)$.*

The reason k-wise γ-biased generators are useful for us is that they satisfy the following two properties.

Lemma 2.5 (Small-bias generators fool juntas and conjunctions of literals [NN93; AGHP92])**.** *Let X be a k-wise γ-biased distribution over $\{0,1\}^n$. Then X is ε-almost k-wise uniform, where $\varepsilon = \gamma \cdot 2^{k/2}$. Furthermore, X fools every conjunction of at most k literals with error γ.*

2.4 Parity circuits

To construct examples showing the weakness of k-wise γ-biased generators, we will rely on circuits computing the parity function.

Proposition 2.6 (Parity circuits)**.** *For any integer $n \geq 2$, the function $f(x_1, \ldots, x_n) = x_1 \oplus x_2 \oplus \cdots \oplus x_n$ can be computed by a U_2-circuit of size $3n - 3$.*

Proof. When $n = 2$, we have $x_1 \oplus x_2 = (x_1 \wedge \overline{x}_2) \vee (\overline{x}_1 \wedge x_2)$. When $n > 2$, we perform a tree of binary $\oplus$ operations, each of which can be computed using three gates. □

2.5 Fourier analysis of Boolean functions

Our seed length lower bound for fooling decision trees using small-bias distributions uses Fourier analysis. For a set $S \subseteq [n]$, we use the notation $\chi_S\colon \{0,1\}^n \to \mathbb{R}$ to denote the function $\chi_S(x) = \prod_{i \in S} (-1)^{x_i}$. For a function $f\colon \{0,1\}^n \to \mathbb{R}$, we use the notation $\widehat{f}(S)$ to denote the Fourier coefficient of f at S:

$$\widehat{f}(S) = \mathop{\mathbb{E}}_{x \in \{0,1\}^n} [f(x) \cdot \chi_S(x)].$$

Parseval's theorem states that

$$\mathop{\mathbb{E}}_{x \in \{0,1\}^n} [f(x)^2] = \sum_{S \subseteq [n]} \widehat{f}(S)^2.$$

[6] That is, given a seed $x \in \{0,1\}^{O(q+\ell)}$ and an input $y \in \{0,1\}^q$, the value $h_x(q)$ can be computed in $\mathrm{poly}(q, \ell)$ time, where h_x is the hash function corresponding to the seed x.

3 Characterizing k-wise probable uniformity

The following proposition shows the equivalence of three ways of defining k-wise probably uniform distributions.

Proposition 3.1 (Equivalence of three definitions of k-wise probable uniformity)**.** *Let X be a distribution over $\{0,1\}^n$, let $k \in [n]$, and let $\varepsilon \in [0,1]$. Then the following are equivalent.*

1. *For every k-junta $f\colon \{0,1\}^n \to \{0,1\}$, we have $\mathbb{E}[f(X)] \geq (1-\varepsilon)\cdot \mathbb{E}[f]$.*
2. *For every size-k set $S \subseteq [n]$ and every $z \in \{0,1\}^k$, we have $\Pr[X_S = z] \geq (1-\varepsilon)\cdot 2^{-k}$.*
3. *For every size-k set $S \subseteq [n]$, there exists a distribution E over $\{0,1\}^k$ such that one can sample from X_S by sampling from U_k with probability $1-\varepsilon$ and sampling from E with probability ε.*

Proof.

- (1 $\implies$ 2) Consider the function $f(x) = 1 \iff x_S = z$.
- (2 $\implies$ 3) If $\varepsilon = 0$, then for every $x \in \{0,1\}^k$, we have $\Pr[X_S = x] \geq 2^{-k}$, which implies that X_S is exactly uniform over $\{0,1\}^k$. If $\varepsilon > 0$, define $p\colon \{0,1\}^k \to \mathbb{R}$ by the formula

$$p(x) = \frac{\Pr[X_S = x] - (1-\varepsilon)\cdot 2^{-k}}{\varepsilon}.$$

  Then $p(x)$ is a probability mass function: it is nonnegative because $\Pr[X_S = x] \geq (1-\varepsilon)\cdot 2^{-k}$, and it sums to 1 because X_S is a probability distribution. Let E be corresponding probability distribution.

- (3 $\implies$ 1) If f is a k-junta, then there is some set $S \subseteq [n]$ of size k and some function $g\colon \{0,1\}^k \to \{0,1\}$ such that $f(x) = g(x_S)$ for all $x \in \{0,1\}^n$. Therefore,

$$\mathbb{E}[f(X)] = \mathbb{E}[g(X_S)] = (1-\varepsilon)\cdot \mathbb{E}[g(U_k)] + \varepsilon \cdot \mathbb{E}[g(E_S)] \geq (1-\varepsilon)\cdot \mathbb{E}[f]. \qquad \square$$

By definition, if X satisfies any of the three equivalent conditions in Proposition 3.1, then X is k-wise ε-probably uniform. The third condition in Proposition 3.1 motivates the name "k-wise probably uniform," but we find it more mathematically convenient to work with the first two conditions.

4 Constructing k-wise probably uniform generators

In this section, we present our new k-wise probably uniform generator, thereby proving Theorem 1.3. At the end of this section, for completeness' sake, we record the standard nonconstructive proof of the existence of nonexplicit k-wise probably uniform generators with excellent seed lengths.

4.1 A small family of generators, each with a good seed length

As a first step, we begin by constructing a family of generator $\mathcal{G}$, such that for any k_0-junta f, most generators $g \in \mathcal{G}$ satisfy $\mathbb{E}_x[f(g(x))] \geq (1-\zeta)\cdot \mathbb{E}[f]$. This construction is based on a combination of pairwise uniform hash functions and k-wise γ-biased generators.

Lemma 4.1 (Family of generators)**.** *For every $n, k_0 \in \mathbb{N}$ and $\zeta \in (0,1)$, there exists an explicit family $\mathcal{G}$ of PRGs $g\colon \{0,1\}^q \to \{0,1\}^n$ satisfying the following.*

1. *A generator $g \sim \mathcal{G}$ can be sampled using $O(k_0 + \log(1/\zeta) + \log\log n)$ truly random bits.*
2. *Each generator g in $\mathcal{G}$ has seed length $q = k_0 + O(\log(1/\zeta))$.*

3. If $f\colon \{0,1\}^n \to \{0,1\}$ is a k_0-junta, then

$$\Pr_{g\sim\mathcal{G}}\left[\mathop{\mathbb{E}}_{x\in\{0,1\}^q}[f(g(x))] \geq (1-\zeta)\cdot\mathbb{E}[f]\right] \geq 1-\zeta.$$

Proof. Let $G_{\mathrm{sb}}\colon \{0,1\}^\ell \to \{0,1\}^n$ be a k-wise γ-biased generator where $\gamma = (\zeta/2)\cdot 2^{-3k_0/2}$ and

$$\ell = O(k_0 + \log(1/\zeta) + \log\log n).$$

Let $\mathcal{H}$ be a pairwise uniform family of hash functions $h\colon \{0,1\}^q \to \{0,1\}^\ell$. For each hash function h in $\mathcal{H}$, we define a generator $g(x) = G_{\mathrm{sb}}(h(x))$. By Theorems 2.3 and 2.4, this family is explicit and $\mathcal{G}$ can be sampled using $O(k_0 + \log(1/\zeta) + \log\log n)$ truly random bits.

For the correctness proof, let $\mu = \mathbb{E}_{y\in\{0,1\}^\ell}[f(G_{\mathrm{sb}}(y))]$. The generator G_{sb} fools f with error $\gamma\cdot 2^{k_0/2}$ (see Lemma 2.5), and $\mathbb{E}[f] \geq 2^{-k_0}$ unless $f\equiv 0$, so $\mu \geq (1-\zeta/2)\cdot\mathbb{E}[f]$. Now pick $h\sim\mathcal{H}$, and let $E_x = f(G_{\mathrm{sb}}(h(x))$ for every $x\in\{0,1\}^q$, so E_x is a random variable based on the choice of h. Then for each fixed x, we have $\mathbb{E}[E_x] = \mu$, so $\mathbb{E}[\sum_x E_x] = 2^q\cdot\mu$. Furthermore, $\mathrm{Var}[E_x] = \mu\cdot(1-\mu) \leq \mu$, and the variables E_x are pairwise independent, so $\mathrm{Var}[\sum_x E_x] \leq 2^q\cdot\mu$. Therefore, by Chebyshev's inequality, we have

$$\Pr\left[\left|\sum_x E_x - 2^q\cdot\mu\right| \geq (\zeta/2)\cdot 2^q\cdot\mu\right] \leq \frac{2^q\cdot\mu}{(1/4)\cdot\zeta^2\cdot 2^{2q}\cdot\mu^2} = \frac{4}{\zeta^2\cdot\mu\cdot 2^q} \leq \frac{8\cdot 2^{k_0}}{\zeta^2\cdot 2^q} \leq \zeta,$$

provided we choose a suitable value $q = k_0 + O(\log(1/\zeta))$. Now fix an h such that the bad event above does not occur. Then with respect to the choice of $x\in\{0,1\}^q$, we have

$$\begin{aligned}\mathop{\mathbb{E}}_x[f(G_{\mathrm{sb}}(h(x)))] = \mathop{\mathbb{E}}_x[E_x] = 2^{-q}\cdot\sum_x E_x &\geq 2^{-q}\cdot(1-\zeta/2)\cdot 2^q\cdot\mu \\ &\geq (1-\zeta/2)\cdot(1-\zeta/2)\cdot\mathbb{E}[f] \\ &> (1-\zeta)\cdot\mathbb{E}[f].\end{aligned}$$

□

4.2 Pseudorandomly partitioning the coordinates into buckets

In this subsection, we explain how to pseudorandomly partition the coordinates into buckets, $[n] = B_1\cup\cdots\cup B_r$, such that no single bucket gets too many of the k coordinates we care about. To be more precise, we construct a *balanced partition generator*, defined as follows.

Definition 4.2 (Balanced partition generator [MZ13])**.** A (k,k_0,δ)-*balanced partition generator* is a function $G_{\mathrm{vars}}\colon \{0,1\}^a \to [r]^n$ such that for every set $S\subseteq[n]$ with $|S|\leq k$, with probability at least $1-\delta$ over a uniform random choice of seed $x\in\{0,1\}^a$, for every bucket $j\in[r]$, we have $|\{i\in S : G_{\mathrm{vars}}(x)_i = j\}| \leq k_0$.

Definition 4.2 is due to Meka and Zuckerman, who used the term "balanced hash family" [MZ13, Definition 4.9]. We use the term "balanced partition generator" to avoid confusion with the hash functions that appear in the proof of Lemma 4.1. Our balanced partition generator will essentially consist of a d-wise γ-biased generator for appropriate values d and γ. The analysis will be based on the following bound on the moments of a sum of independent Bernoulli random variables [SSS95].[7]

Theorem 4.3 (Moment bound for a sum of independent Bernoulli random variables [SSS95])**.** *Let $X_1,\ldots,X_k$ be independent $\{0,1\}$-valued random variables. Let $X = \sum_{i=1}^k X_i$, let $\mu_i = \mathbb{E}[X_i]$, and let $\mu = \sum_{i=1}^k \mu_i$. Then for every even positive integer t, we have*

$$\mathbb{E}[(X-\mu)^t] \leq \max\{t^t, (t\mu)^{t/2}\}.$$

[7] The exact statement of Theorem 4.3 does not appear in Schmidt, Siegel, and Srinivasan's work [SSS95], but it follows from the proof of item "(III)" in their "Theorem 4."

Theorem 4.3 can be improved in some parameter regimes [Sko22], but the simple bound in Theorem 4.3 suffices for our purposes. Using Theorem 4.3, we now present a tail bound for sums of random variables that satisfy a certain "near t-wise independence" condition. Similar bounds were proven in several previous papers [LRTV09; CRSW13; SVW17], and our proof is almost identical to their proofs.

Corollary 4.4 (Tail bound for sums of nearly t-wise independent random variables)**.** *Let $X_1, \dots, X_k$ be $\{0,1\}$-valued random variables and let $\mu_1, \dots, \mu_k \in [0,1]$. Let $X = \sum_{i=1}^k X_i$ and $\mu = \sum_{i=1}^k \mu_i$. Let t be an even positive integer, let $\gamma \in (0,1)$, and assume that for every set $S \subseteq [k]$ with $|S| \le t$, we have*

$$\left| \mathbb{E}\left[\prod_{i \in S} X_i\right] - \prod_{i \in S} \mu_i \right| \le \gamma.$$

Then for every $\Delta > 0$, we have

$$\Pr[|X - \mu| \ge \Delta] \le \left(\frac{t}{\Delta}\right)^t + \left(\frac{\sqrt{\mu t}}{\Delta}\right)^t + \gamma \cdot \left(\frac{2k}{\Delta}\right)^t.$$

Proof. Sample $X'_1, \dots, X'_k \in \{0,1\}$ independently, where $\mathbb{E}[X'_i] = \mu_i$, and let $X' = \sum_{i=1}^k X'_i$. Then

$$\begin{aligned}
\Pr[|X - \mu| \ge \Delta] &= \Pr[(X-\mu)^t \ge \Delta^t] \\
&\le \Delta^{-t} \cdot \mathbb{E}[(X-\mu)^t] && \text{(Markov's inequality)} \\
&= \Delta^{-t} \cdot \sum_{i=0}^{t} \binom{t}{i} (-\mu)^{t-i} \cdot \mathbb{E}[X^i] && \text{(Binomial theorem)} \\
&= \Delta^{-t} \cdot \sum_{i=0}^{t} \binom{t}{i} (-\mu)^{t-i} \cdot \sum_{j_1, \dots, j_i \in [k]} \mathbb{E}[X_{j_1} X_{j_2} \cdots X_{j_i}] \\
&\le \Delta^{-t} \cdot \sum_{i=0}^{t} \binom{t}{i} \cdot \left((-\mu)^{t-i} \cdot \sum_{j_1, \dots, j_i \in [k]} \mu_{j_1} \cdots \mu_{j_i} + \mu^{t-i} \cdot k^i \cdot \gamma \right) \\
&= \Delta^{-t} \cdot \left(\mathbb{E}[(X' - \mu)^t] + \gamma \cdot \sum_{i=0}^{t} \binom{t}{i} \mu^{t-i} \cdot k^i \right) \\
&= \Delta^{-t} \cdot \left(\mathbb{E}[(X' - \mu)^t] + \gamma \cdot (\mu + k)^t \right) && \text{(Binomial theorem)} \\
&\le \left(\frac{t}{\Delta}\right)^t + \left(\frac{\sqrt{\mu t}}{\Delta}\right)^t + \gamma \cdot \left(\frac{2k}{\Delta}\right)^t && \text{(Theorem 4.3.)}
\end{aligned}$$

□

Given Corollary 4.4, we are ready to construct our balanced partition generator.

Lemma 4.5 (Balanced partition generator)**.** *Let $n, k, r \in \mathbb{N}$ and $\delta \in (0,1)$. Assume r is a power of two and $r \le k \le n$. There exists an explicit (k, k_0, δ)-balanced partition generator $G_{\text{vars}} \colon \{0,1\}^a \to [r]^n$, where*

$$k_0 = k/r + O\left(\sqrt{k/r \cdot \log(r/\delta)} + \log(r/\delta)\right),$$

with seed length

$$a = O\left(\log(r/\delta) \cdot \log\left(2 \cdot \left\lceil \frac{rk}{\log(r/\delta)} \right\rceil\right) + \log\log n\right).$$

Proof. Identify $[r]^n$ with $\{0,1\}^{n \log r}$. We let G_{vars} be a $(t \log r)$-wise γ-biased generator for appropriate values

$$t = \log(3r/\delta)$$
$$\gamma = \frac{\delta}{3r} \cdot \left(\frac{t}{rk}\right)^{t/2}.$$

The seed length bound follows from Theorem 2.4. For the correctness proof, assume without loss of generality that $|S| = k$. Sample $Z \in [r]^n$ using the generator. Fix any bucket $j \in [r]$. For each $i \in S$, let X_i indicate whether $Z_i = j$. Then for any set $T \subseteq S$ with $|T| \leq t$, the value $\prod_{i \in T} X_i$ can be expressed in terms of the underlying bits of Z as a conjunction of at most $t \log r$ literals. Therefore, by Lemma 2.5, we have $|\mathbb{E}[\prod_{i \in T} X_i] - r^{-|T|}| \leq \gamma$. Therefore, by Corollary 4.4, for every $\Delta > 0$, we have

$$\Pr\left[\sum_{i \in S} X_i \geq k/r + \Delta\right] \leq \left(\frac{t}{\Delta}\right)^t + \left(\frac{\sqrt{kt/r}}{\Delta}\right)^t + \gamma \cdot \left(\frac{2k}{\Delta}\right)^t.$$

We choose $\Delta = \max\left\{2t, 2\sqrt{kt/r}\right\}$. Then we get

$$\begin{aligned}\Pr\left[\sum_{i \in S} X_i \geq k/r + \Delta\right] &\leq 2^{-t} + 2^{-t} + \gamma \cdot \left(\sqrt{\frac{rk}{t}}\right)^t \\ &\leq \frac{\delta}{3r} + \frac{\delta}{3r} + \frac{\delta}{3r}\end{aligned}$$

due to our choices of t and γ. The union bound over r buckets completes the proof. □

For comparison, Lovett, Reingold, Trevisan, and Vadhan constructed an explicit (k, k_0, δ)-balanced partition generator for the special case $k = \Theta(r \cdot \log(1/\delta))$, with $k_0 = O(k/r)$ and seed length $a = O(\log n + \log(r/\delta) \cdot \log(r \cdot \log(1/\delta)))$ [LRTV09]. For any k, one can also use Gopalan, Kane, and Meka's PRG for Fourier shapes [GKM18] to construct a (k, k_0, δ)-balanced partition generator with the same value of k_0 as in Lemma 4.5 and with seed length $a = \widetilde{O}(\log(n/\delta))$.

4.3 The full k-wise probably uniform generator

Proof of Theorem 1.3. Let $G_{\text{vars}} \colon \{0,1\}^a \to [r]^n$ be the (k, k_0, δ)-balanced partition generator from Lemma 4.5 with parameters $\delta = \varepsilon/3$ and $r = (k/\log(k/\varepsilon))^{1/3}$, or to be more precise, r is the largest power of two that is at most $(k/\log(k/\varepsilon))^{1/3}$. Let $\mathcal{G}$ be the family of generators $g \colon \{0,1\}^q \to \{0,1\}^n$ from Lemma 4.1, using $\zeta = \varepsilon/(3r)$ and using the value k_0 from G_{vars}. The final generator G is defined as follows.

1. Sample a partition $Z = (Z_1, \ldots, Z_n) \in [r]^n$ using G_{vars}.
2. Sample a generator $g \sim \mathcal{G}$.
3. Sample seeds $X^{(1)}, \ldots, X^{(r)} \in \{0,1\}^q$ independently and uniformly at random.
4. Output $Y \in \{0,1\}^n$, where
$$Y_i = g(X^{(Z_i)})_i$$
for every $i \in [n]$.

To prove that this works, let $f \colon \{0,1\}^n \to \{0,1\}$ be a conjunction of k literals, say

$$f(x) = \bigwedge_{i \in S} (x_i \oplus b_i)$$

where $S \subseteq [n]$, $|S| = k$, and $b_i \in \{0,1\}$ for every $i \in S$. We will prove that $\mathbb{E}[f(X)] \geq (1 - \varepsilon) \cdot 2^{-k}$, which is sufficient by Proposition 3.1.

For each bucket $j \in [r]$, let $B_j = Z^{-1}(j)$. The definition of a balanced partition generator ensures that except with probability $\varepsilon/3$ over the choice of Z, we have $|S \cap B_j| \leq k_0$ for every $j \in [r]$. Let E_1 be this "good" event. Fix any choice of Z such that E_1 occurs.

For each $j \in [r]$, define $f_j \colon \{0,1\}^n \to \{0,1\}$ by

$$f_j(x) = \bigwedge_{i \in S \cap B_j} (x_i \oplus b_i),$$

so $f(x) = f_1(x) \wedge \cdots \wedge f_r(x)$. By Lemma 4.1 and the union bound over the r buckets, except with probability $\varepsilon/3$ over the choice of $g \sim \mathcal{G}$, we have

$$\mathop{\mathbb{E}}_{x \in \{0,1\}^q}[f_j(g(x))] \geq \left(1 - \frac{\varepsilon}{3r}\right) \cdot \mathbb{E}[f_j]$$

for every $j \in [r]$. Let E_2 be this "good" event. Fix any choice of g such that E_2 occurs.

For any such fixing of Z and g, with respect to the choice of $X^{(1)}, \ldots, X^{(r)}$ alone, we have

$$\mathop{\mathbb{E}}_{X^{(1)},\ldots,X^{(r)}}[f(Y)] = \prod_{j=1}^{r} \mathop{\mathbb{E}}_{X^{(j)}}[f_j(g(X^{(j)}))] \geq \prod_{j=1}^{r} \left(1 - \frac{\varepsilon}{3r}\right) \cdot \mathbb{E}[f_j] = \left(1 - \frac{\varepsilon}{3r}\right)^r \cdot 2^{-k} \geq (1 - \varepsilon/3) \cdot 2^{-k}$$

by Bernoulli's inequality. Therefore, with respect to all the randomness, we have

$$\begin{aligned}
\mathbb{E}[f(Y)] \geq \Pr[f(Y) = 1 \text{ and } E_1 \text{ and } E_2] &= \Pr[E_1] \cdot \Pr[E_2 \mid E_1] \cdot \Pr[f(Y) = 1 \mid E_1, E_2] \\
&\geq (1 - \varepsilon/3) \cdot (1 - \varepsilon/3) \cdot (1 - \varepsilon/3) \cdot 2^{-k} \\
&\geq (1 - \varepsilon) \cdot 2^{-k}
\end{aligned}$$

by another application of Bernoulli's inequality.

Now let us bound the seed length. By Lemma 4.5, the cost of sampling Z is

$$\begin{aligned}
O\left(\log(r/\varepsilon) \cdot \log\left(2 \cdot \left\lceil \frac{rk}{\log(r/\varepsilon)} \right\rceil\right) + \log\log n\right) &\leq O\left(\log(k/\varepsilon) \cdot \log\left(2 \cdot \left\lceil \frac{k}{\log(k/\varepsilon)} \right\rceil\right) + \log\log n\right) \\
&\leq O\left(\log(k/\varepsilon) \cdot \left(\frac{k}{\log(k/\varepsilon)}\right)^{2/3} + \log(k/\varepsilon) + \log\log n\right) \\
&= O(k^{2/3} \cdot \log^{1/3}(k/\varepsilon) + \log(k/\varepsilon) + \log\log n).
\end{aligned}$$

Furthermore, the parameter k_0 is given by

$$k_0 = k/r + O\left(\sqrt{k/r \cdot \log(r/\varepsilon)} + \log(r/\varepsilon)\right) \leq k/r + O\left(\sqrt{k/r \cdot \log(k/\varepsilon)} + \log(k/\varepsilon)\right).$$

Therefore, by Lemma 4.1, the cost of sampling $g \sim \mathcal{G}$ is

$$\begin{aligned}
O\left(k_0 + \log(k/\varepsilon) + \log\log n\right) &= O(k^{2/3} \cdot \log^{1/3}(k/\varepsilon) + k^{1/3} \cdot \log^{2/3}(k/\varepsilon) + \log(k/\varepsilon) + \log\log n) \\
&= O(k^{2/3} \cdot \log^{1/3}(k/\varepsilon) + \log(k/\varepsilon) + \log\log n).
\end{aligned}$$

Finally, the cost of sampling $X^{(1)}, \ldots, X^{(r)}$ is

$$\begin{aligned}
r \cdot q &= r \cdot k_0 + O(r \cdot \log(k/\varepsilon)) \\
&= k + O(k^{2/3} \cdot \log^{1/3}(k/\varepsilon) + k^{1/3} \log^{2/3}(k/\varepsilon) + \log(k/\varepsilon)) \\
&= k + O(k^{2/3} \cdot \log^{1/3}(k/\varepsilon) + \log(k/\varepsilon)).
\end{aligned}$$

□

4.4 Nonexplicit k-wise probably uniform generators

At this point, we have completed our explicit k-wise uniform generator construction. We now use a standard probabilistic argument to show the existence of nonexplicit k-wise probably uniform generators with a very good seed length.

Proposition 4.6 (Nonexplicit k-wise probably uniform generator)**.** *For every $n, k \in \mathbb{N}$ and every $\varepsilon \in (0,1)$, there exists a k-wise ε-probably uniform generator $G\colon \{0,1\}^s \to \{0,1\}^n$ with seed length*

$$s = k + \log k + 2\log(1/\varepsilon) + \log\log(n/k) + O(1).$$

Proof. Pick G uniformly at random. For every function f that is a conjunction of k literals, let $Z_f = \sum_{x\in\{0,1\}^s} f(G(x))$. Then Z_f is a sum of 2^s independent $\{0,1\}$-valued random variables with mean $\mu := \mathbb{E}[Z_f] = 2^{s-k}$. Therefore, by the Chernoff bound,

$$\Pr[Z_f < (1-\varepsilon)\cdot\mu] \le \exp(-\varepsilon^2\mu/2).$$

By the union bound, it follows that

$$\Pr[\text{there exists } f \text{ such that } Z_f < (1-\varepsilon)\cdot\mu] \le \binom{n}{k}\cdot 2^k \cdot \exp(-\varepsilon^2\mu/2) \le (2en/k)^k \cdot \exp(-\varepsilon^2 2^{s-k}/2).$$

This probability is less than 1 if we choose a suitable value $s = k+\log k+2\log(1/\varepsilon)+\log\log(n/k)+O(1)$. Now suppose G is such that $Z_f \ge (1-\varepsilon)\cdot\mu$ for every f that is a conjunction of k literals. Let $g\colon \{0,1\}^n \to \{0,1\}$ be a k-junta. Then we can write $g = \sum_{i=1}^m f_i$ where each f_i is a conjunction of k literals, hence

$$\mathop{\mathbb{E}}_x[g(G(x))] = \sum_{i=1}^m 2^{-s}\cdot Z_{f_i} \ge \sum_{i=1}^m 2^{-s}\cdot(1-\varepsilon)\cdot 2^{s-k} = (1-\varepsilon)\cdot m\cdot 2^{-k} = (1-\varepsilon)\cdot\mathbb{E}[g]. \qquad \square$$

5 Implications of k-wise probable uniformity

In this section, we will show that every k-wise probably uniform distribution fools decision trees. In fact, we will show that such distributions fool a more general model, called the *subcube partition model*.

Definition 5.1 (The subcube partition model)**.** A *subcube partition* f is a collection of *terms* $f_1,\dots,f_m$ and *values* $b_1,\dots,b_m \in \{0,1\}$. Each term $f_i\colon \{0,1\}^n \to \{0,1\}$ is a conjunction of literals, and the sets $f_1^{-1}(1),\dots,f_m^{-1}(1)$ must partition the domain $\{0,1\}^n$. That is, for every $x \in \{0,1\}^n$, we have $\sum_{i=1}^m f_i(x) = 1$. The subcube partition computes the function $f\colon \{0,1\}^n \to \{0,1\}$ defined by

$$f(x) = \sum_{i=1}^m f_i(x)\cdot b_i.$$

The *width* of a term f_i is the number of literals in the term. The width of the subcube partition is the maximum width of any term. The *size* of the subcube partition is the number of terms (m).

Every width-k subcube partition has size at most 2^k, because $1 = \sum_{i=1}^m \mathbb{E}[f_i] \ge m\cdot 2^{-k}$. A decision tree of depth k and size m can be simulated by a subcube partition of width k and size m: for each leaf u, we construct a term f_u that indicates whether the tree reaches the leaf u on a given input. The converse does not hold. In fact, there exist subcube partitions of width k that cannot be simulated by decision trees of depth $k^{2-\Omega(1)}$ [Sav02; KRDS15; GPW18; AKK16]. We now explain why k-wise probably uniform generators fool subcube partitions.

Lemma 5.2 (k-wise probable uniformity fools subcube partitions)**.** *Let X be a distribution over $\{0,1\}^n$ that is k-wise ε-probably uniform. Then:*

- *X fools width-k subcube partitions (hence also depth-k decision trees) with error ε.*
- *X fools size-m subcube partitions (hence also size-m decision trees) with error $\varepsilon + m\cdot 2^{-(k+1)}$.*

Proof. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function computed by a subcube partition with terms $f_1, \ldots, f_m$ and values $b_1, \ldots, b_m$. Let $S \subseteq [m]$ be the set of terms of width at most k. We will show that X fools f with error $\varepsilon + \sum_{i \notin S} \mathbb{E}[f_i]$. To prove it, sample $R \in \{0,1\}^n$ uniformly at random. Then

$$\begin{aligned}
\mathbb{E}[f(X)] = \sum_{i=1}^{m} b_i \cdot \mathbb{E}[f_i(X)] \geq \sum_{i \in S} b_i \cdot \mathbb{E}[f_i(X)] \geq \sum_{i \in S} b_i \cdot (1-\varepsilon) \cdot \mathbb{E}[f_i] &= (1-\varepsilon) \cdot \mathbb{E}\left[\sum_{i \in S} b_i \cdot f_i(R)\right] \\
&\geq \mathbb{E}\left[\sum_{i \in S} b_i \cdot f_i(R)\right] - \varepsilon \\
&= \mathbb{E}\left[f(R) - \sum_{i \notin S} b_i \cdot f_i(R)\right] - \varepsilon \\
&\geq \mathbb{E}[f] - \sum_{i \notin S} \mathbb{E}[f_i] - \varepsilon.
\end{aligned}$$

Now we bound the expectation from above. Let $\overline{f} = 1 - f$. Since $\overline{f}$ can also be computed by a subcube partition with the same terms $f_1, \ldots, f_m$, we have

$$\mathbb{E}[f(X)] = 1 - \mathbb{E}\left[\,\overline{f}(X)\,\right] \leq 1 - \mathbb{E}\left[\,\overline{f}\,\right] + \varepsilon + \sum_{i \notin S} \mathbb{E}[f_i] = \mathbb{E}[f] + \varepsilon + \sum_{i \notin S} \mathbb{E}[f_i].$$

The lemma follows, because $\mathbb{E}[f_i] \leq 2^{-(k+1)}$ whenever $i \notin S$. □

By combining Theorem 1.3 (our k-wise probably uniform generator) with Lemma 5.2, we now prove the following theorem, which generalizes Theorem 1.5.

Theorem 5.3 (Fooling near-maximal subcube partitions)**.** *Let $n, m \in \mathbb{N}$ and $\varepsilon \in (0,1)$. There exists an explicit PRG $G\colon \{0,1\}^s \to \{0,1\}^n$ that fools n-variate subcube partitions of size m with error ε and seed length*

$$s = \log m + O\left(\log^{2/3} m \cdot \log^{1/3}\left(\frac{\log m}{\varepsilon}\right) + \log(1/\varepsilon) + \log\log n\right). \tag{3}$$

Proof. We use our k-wise $(\varepsilon/2)$-probably uniform generator, where $k = \log m + \log(2/\varepsilon)$. By Lemma 5.2, the generator fools size-m subcube partitions with error $\varepsilon/2 + m \cdot 2^{-k} = \varepsilon$. By Theorem 1.3, the seed length is

$$k + O(k^{2/3} \cdot \log^{1/3}(k/\varepsilon) + \log(1/\varepsilon) + \log\log n),$$

which, after substituting the choice of k, simplifies to Eq. (3). □

6 Limitations of k-wise γ-biased generators

In this section, we prove that our main results (Theorems 1.3 and 1.5 and Corollary 1.6) cannot be proven by simply developing a better construction and/or analysis of k-wise γ-biased generators.

First, in Section 6.1, we present examples showing that if one wishes to use a generic k-wise γ-biased generator to sample a t-wise probably uniform distribution, or to fool near-maximal decision trees, or to fool fool U_2-circuits of size $2n$, then one is forced to use a very large k and a very small γ. Then, in Section 6.2, we extend Karloff and Mansour's work [KM97] to show that when k is very large and γ is very small, every k-wise γ-biased generator has a very large seed length.

6.1 Counterexamples showing that k must be large and γ must be small

We begin by analyzing the parameter k. The argument is fairly trivial.

Proposition 6.1. *For every $n \in \mathbb{N}$ and $k \in [n-1]$, there exists a k-wise uniform distribution X over $\{0,1\}^n$ such that:*

1. *X is not 0.49-almost $(k+1)$-wise uniform.*
2. *X does not 0.49-fool U_2-circuits of size $3k$.*

Proof. Let X be the uniform distribution over the set $\{x \in \{0,1\}^n : x_1 \oplus x_2 \oplus \cdots \oplus x_{k+1} = 0\}$. Let $f(x) = x_1 \oplus \cdots \oplus x_{k+1}$. Then f can be computed by a circuit of size $3k$ over the U_2 basis (Proposition 2.6), and $f(x)$ depends on only $k+1$ bits of x, and $|\mathbb{E}[f] - \mathbb{E}[f(X)]| = 1/2$. □

Now we move on to the bias parameter, γ. We begin by showing that a very small bias would be required to achieve k-wise probable uniformity.

Proposition 6.2. *For every $n \in \mathbb{N}$, every $k \in [n]$, and every $\varepsilon \in (0,1)$, there exists a distribution X over $\{0,1\}^n$ such that X is n-wise $O(\varepsilon \cdot 2^{-k})$-biased, but X is not k-wise ε-probably uniform.*

Proof. With probability $1 - 2\varepsilon$, we sample X uniformly at random. With probability 2ε, we sample X uniformly at random from the set $\{x \in \{0,1\}^n : (x_1, \ldots, x_k) \neq 0^k\}$. To show that this distribution is n-wise $(2\varepsilon \cdot 2^{-k})$-biased, let $S \subseteq [n]$ be any nonempty set of size at most k. If $S \nsubseteq [k]$, then $\mathbb{E}[\chi_S(X)] = 0$, because $(X_{k+1}, \ldots, X_n)$ is uniform over $\{0,1\}^{n-k}$ and independent of $(X_1, \ldots, X_k)$. If $S \subseteq [k]$, then

$$\begin{aligned}
|\mathbb{E}[\chi_S(X)]| = \left|(1-2\varepsilon)\cdot 0 + 2\varepsilon \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^k \setminus \{0^k\}}[\chi_S(x)]\right| &= \frac{2\varepsilon}{2^k - 1} \cdot \left|\sum_{x \in \{0,1\}^k \setminus \{0^k\}} \chi_S(x)\right| \\
&= \frac{2\varepsilon}{2^k - 1} \cdot \left|\left(\sum_{x \in \{0,1\}^k} \chi_S(x)\right) - \chi_S(0^k)\right| \\
&= \frac{2\varepsilon}{2^k - 1} \\
&\leq \frac{4\varepsilon}{2^k}.
\end{aligned}$$

On the other hand, X is not k-wise ε-probably uniform, because

$$\Pr[(X_1, \ldots, X_k) = 0^k] = (1-2\varepsilon) \cdot 2^{-k} < (1-\varepsilon) \cdot 2^{-k}.$$ □

Next, we show that a very small bias would be required to fool decision trees of depth $0.76 \cdot n$, or to fool circuits of size $2n$ over the U_2 basis. The proof is based on the "inner product mod 2" function. For each even positive integer n, we define $\mathsf{IP}_n \colon \{0,1\}^n \to \{0,1\}$ by the formula

$$\mathsf{IP}_n(x,y) = \bigoplus_{i=1}^{n/2} x_i y_i.$$

Proposition 6.3. *Let n be an even positive integer and let X be the uniform distribution over $\mathsf{IP}_n^{-1}(0)$. Then:*

1. *The distribution X is n-wise $(2^{-n/2})$-biased.*[8]
2. *The distribution X does not fool U_2-circuits of size $2n-3$ with error 0.49, assuming n is sufficiently large.*

[8] For context, Bogdanov and Viola previously showed that X is n-wise $(2^{-\Omega(n)})$-biased, and they also showed a generalization of this statement to larger fields [BV10].

3. *There is a value $k = \frac{3}{4} \cdot n + O(\sqrt{n})$ such that X does not fool depth-k decision trees with error 0.49, assuming n is sufficiently large.*

Proof. Let $f(x, y) = (-1)^{\mathsf{IP}_n(x,y)}$. Let χ_S be any nontrivial character function. Sample $R \in \{0,1\}^n$ uniformly at random, and sample Y uniformly from $\mathsf{IP}^{-1}(1)$. For each $b \in \{0,1\}$, let $p_b = \Pr[\mathsf{IP}(R) = b]$. Note that $p_0 > 1/2$. Therefore,

$$\begin{aligned}
|\mathbb{E}[\chi_S(X)]| &< 2p_0 \cdot |\mathbb{E}[\chi_S(X)]| \\
&= |p_0 \cdot \mathbb{E}[\chi_S(X)] + p_1 \cdot \mathbb{E}[\chi_S(Y)] + p_0 \cdot \mathbb{E}[\chi_S(X)] - p_1 \cdot \mathbb{E}[\chi_S(Y)]| \\
&= |\mathbb{E}[\chi_S(R)] + \mathbb{E}[\chi_S(R) \cdot f(R)]| \\
&= |\widehat{f}(S)|.
\end{aligned}$$

(The second-to-last equation is an application of the law of total expectation.) It follows that X is n-wise $(2^{-n/2})$-biased, because the inner product mod 2 function is famously "bent," meaning that $|\widehat{f}(S)| = 2^{-n/2}$ for every S. For completeness, we include the calculation showing that $|\widehat{f}(S)| = 2^{-n/2}$ below:

$$\begin{aligned}
\widehat{f}(S) &= \mathop{\mathbb{E}}_{x,y}[f(x,y) \cdot \chi_S(x,y)] \\
&= \mathop{\mathbb{E}}_{x,y}\left[\left(\prod_{i=1}^{n/2}(-1)^{x_i y_i}\right) \cdot \left(\prod_{i=1}^{n/2}(-1)^{x_i u_i}\right) \cdot \left(\prod_{i=1}^{n/2}(-1)^{y_i v_i}\right)\right] \qquad \text{for some } u, v \in \{0,1\}^{n/2} \\
&= \prod_{i=1}^{n/2} \mathop{\mathbb{E}}_{a,b \in \{0,1\}}[(-1)^{ab + a u_i + b v_i}] \\
&= \prod_{i=1}^{n/2} \frac{1 + (-1)^{v_i} + (-1)^{u_i} + (-1)^{1+u_i+v_i}}{4} \\
&= \prod_{i=1}^{n/2} \left(\pm \frac{1}{2}\right) \\
&= \pm 2^{-n/2}.
\end{aligned}$$

The distribution X does not 0.49-fool circuits of size $2n - 3$ over the U_2 basis, because $\mathbb{E}[\mathsf{IP}_n] = 1/2 - o(1)$, and IP_n can be computed by a circuit of size $2n - 3$:

- We use $n/2$ "AND" gates to compute the bits $x_1 y_1, \ldots, x_{n/2} y_{n/2}$.
- Then we use $3(n/2) - 3$ gates to compute the parity of those $n/2$ bits (Proposition 2.6).

Finally, we will show that X does not 0.49-fool decision trees of depth $\frac{3}{4} \cdot n + O(\sqrt{n})$. Define

$$T(x,y) = \begin{cases} \mathsf{IP}(x,y) & \text{if } |x| \le n/4 + 2\sqrt{n} \\ 0 & \text{if } |x| > n/4 + 2\sqrt{n}, \end{cases}$$

where $|x|$ denotes the Hamming weight of x and c is an appropriate constant. Then $T(x,y)$ can be computed by a decision tree of depth $\frac{3}{4} \cdot n + O(\sqrt{n})$, and $T \le \mathsf{IP}_n$, so $\mathbb{E}[T(X)] = 0$. On the other hand, if we pick x and y uniformly at random:

- There is a $2^{-n/2}$ chance that $x = 0^{n/2}$.
- There is at most an $\exp(-16)$ chance that x has Hamming weight more than $n/4 + 2\sqrt{n}$, by Hoeffding's inequality.
- For any fixing of x such that neither of the two events above occur, we have $\mathbb{E}_y[T(x,y)] = 1/2$.

Therefore, $\mathbb{E}[T] \ge \frac{1}{2} - 2^{-n/2} - \exp(-16) \ge 0.49$. □

6.2 Seed length lower bound for k-wise γ-biased generators

In this section, we prove our seed length lower bound for k-wise γ-biased generators (Theorem 1.7). The proof is a straightforward extension of Karloff and Mansour's argument [KM97], which covers the case $\gamma = 0$. The approach is to bound the *collision probability* of a k-wise γ-biased distribution.

Definition 6.4 (Collision probability)**.** Let X be a probability distribution over the space $\mathcal{X}$. The *collision probability* $\mathsf{CP}(X)$ is defined by

$$\mathsf{CP}(X) = \Pr_{\substack{x \sim X \\ x' \sim X}}[x = x'],$$

where x and x' are sampled independently from X. Equivalently, $\mathsf{CP}(X) = \sum_{x \in \mathcal{X}} \Pr[X = x]^2$.

Theorem 6.5 (Collision probability of k-wise γ-biased distributions)**.** *Let $n \in \mathbb{N}$, let $\gamma \in (0, 1)$, let $\alpha \in (0, 1/2]$, let $k = \lfloor (\frac{1}{2} + \alpha) \cdot n \rfloor$, and let X be a distribution that is k-wise γ-biased. Then*

$$\mathsf{CP}(X) \le \left(1 + \frac{1}{2\alpha}\right) \cdot \left(2^{-n} + \gamma^2\right).$$

Proof. Let $p \colon \{0,1\}^n \to [0,1]$ be the probability mass function of X, i.e., $p(x) = \Pr[X = x]$. Since X is a probability distribution, we have $\widehat{p}(\varnothing) = 2^{-n}$. Furthermore, since X is k-wise γ-biased, we have $|\widehat{p}(S)| \le \gamma \cdot 2^{-n}$ whenever $1 \le |S| \le k$. Therefore, we can bound the collision probability of X as follows.

$$\begin{aligned}
\mathsf{CP}(X) = \sum_{x \in \{0,1\}^n} p(x)^2 &= 2^n \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^n}[p(x)^2] \\
&= 2^n \cdot \sum_{S \subseteq [n]} \widehat{p}(S)^2 && \text{(Parseval's theorem)} \\
&\le 2^n \cdot \left(\frac{1}{2^{2n}} + \binom{n}{\le k} \cdot \frac{\gamma^2}{2^{2n}} + \sum_{\substack{S \subseteq [n] \\ |S| > k}} \widehat{p}(S)^2 \right) \\
&\le 2^{-n} + \gamma^2 + 2^n \cdot \sum_{\substack{S \subseteq [n] \\ |S| > k}} \widehat{p}(S)^2.
\end{aligned}$$

To bound the high-degree Fourier weight, let $x^{\oplus i}$ denote x with the i-th bit flipped. Identify a set $T \subseteq [n]$

with its indicator function $T\colon [n] \to \{0,1\}$. Then

$$\begin{aligned}
0 \le \sum_{i=1}^{n} \sum_{x\in\{0,1\}^n} p(x)\cdot p(x^{\oplus i}) &= \sum_{S\subseteq[n]} \sum_{T\subseteq[n]} \widehat{p}(S)\cdot\widehat{p}(T)\cdot \sum_{i=1}^{n} \sum_{x\in\{0,1\}^n} \chi_S(x)\cdot\chi_T(x^{\oplus i}) \\
&= \sum_{S\subseteq[n]} \sum_{T\subseteq[n]} \widehat{p}(S)\cdot\widehat{p}(T)\cdot \sum_{i=1}^{n} (-1)^{T(i)} \cdot \sum_{x\in\{0,1\}^n} \chi_S(x)\cdot\chi_T(x) \\
&= 2^n \cdot \sum_{S\subseteq[n]} \widehat{p}(S)^2 \cdot \sum_{i=1}^{n} (-1)^{S(i)} \\
&= 2^n \cdot \sum_{d=0}^{n} \sum_{|S|=d} \widehat{p}(S)^2 \cdot (n-2d) \\
&\le 2^n \cdot \left(\left(n \cdot \sum_{\substack{S\subseteq[n] \\ |S|\le k}} \widehat{p}(S)^2 \right) - (2k+2-n) \cdot \sum_{\substack{S\subseteq[n] \\ |S|>k}} \widehat{p}(S)^2 \right) \\
&\le n\cdot(2^{-n}+\gamma^2) - 2^n\cdot(2k+2-n)\cdot \sum_{d=k+1}^{n} \sum_{|S|=d} \widehat{p}(S)^2.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathsf{CP}(X) \le 2^{-n} + \gamma^2 + \frac{n\cdot(2^{-n}+\gamma^2)}{2k+2-n} = \frac{2k+2}{2k+2-n}\cdot(2^{-n}+\gamma^2) &= \left(1+\frac{n}{2k+2-n}\right)\cdot(2^{-n}+\gamma^2) \\
&\le \left(1+\frac{n}{(1+2\alpha)n-n}\right)\cdot(2^{-n}+\gamma^2) \\
&= \left(1+\frac{1}{2\alpha}\right)\cdot(2^{-n}+\gamma^2). \qquad \square
\end{aligned}$$

Proof of Theorem 1.7. The output of G has collision probability at least 2^{-s}, since this is the chance of getting the same seed twice in a row. Therefore,

$$2^{-s} \le \left(1+\frac{1}{2\alpha}\right)\cdot(2^{-n}+\gamma^2) \le \frac{2}{\alpha}\cdot\max\{2^{-n},\gamma^2\},$$

and consequently

$$s \ge \min\{n, 2\log(1/\gamma)\} - \log(2/\alpha). \qquad \square$$

By combining the results of this subsection with the counterexamples from the previous subsection, we get the following conclusion.

Corollary 6.6. *Let $n,k\in\mathbb{N}$ and $\gamma\in(0,1)$. Suppose that at least one of the following holds.*

1. *Every k-wise γ-biased distribution over $\{0,1\}^n$ is $(0.51\cdot n)$-wise 0.49-probably uniform.*
2. *Every k-wise γ-biased distribution over $\{0,1\}^n$ fools decision trees of depth $0.76\cdot n$ with error 0.49.*
3. *Every k-wise γ-biased distribution over $\{0,1\}^n$ fools circuits of size $2n$ over the U_2 basis with error 0.49.*

Then every k-wise γ-biased generator has seed length $n-O(1)$.

Proof. First, we show that $k \geq \frac{1}{2} + \Omega(1)$. Case (1) implies that every k-wise uniform distribution is 0.49-almost $(0.51 \cdot n)$-wise uniform, hence $k \geq \lfloor 0.51 \cdot n \rfloor$ by Proposition 6.1. Similarly, case (2) implies that every k-wise uniform distribution is 0.49-almost $(0.76 \cdot n)$-wise uniform, hence $k \geq \lfloor 0.76 \cdot n \rfloor$. Finally, case (3) implies $k > 2n/3$ by Proposition 6.1.

Next, we show that $\gamma \leq O(2^{-n/2})$. In case (1), this follows immediately from Proposition 6.2. Now suppose we are in case (2) or (3). Let Z be the distribution over $\{0,1\}^{n'}$ from Proposition 6.3, where $n' \in \{n, n-1\}$ and n' is even. By appending a uniform random bit to Z if necessary, we get a distribution Z' over $\{0,1\}^n$ such that (a) Z' is n-wise $(2^{-(n-1)/2})$-biased, but (b) Z' does not fool decision trees of depth $0.76n$ with error 0.49, nor does it fool circuits of size $2n$ over the U_2 basis with error 0.49. Therefore, $\gamma < 2^{-(n-1)/2}$.

Finally, because the parameters k and γ have such extreme values, Theorem 1.7 tells us that every k-wise γ-biased generator has seed length at least $\min\{n, 2\log(1/\gamma)\} - O(1) = n - O(1)$. □

7 Open problems

- Find more applications of k-wise probably uniform generators.
- Improve the seed lengths of our constructions.
- Design an explicit PRG, with a seed length similar to that of our k-wise probably uniform generator, that samples a distribution X such that

  $$(1-\varepsilon) \cdot \mathbb{E}[f] \leq \mathbb{E}[f(X)] \leq (1+\varepsilon) \cdot \mathbb{E}[f]$$

  for every k-junta f. This is equivalent to saying that every k coordinates of X are uniform to within ℓ_∞ error $\varepsilon \cdot 2^{-k}$. Such a PRG could be used to fool near-maximal unambiguous DNF formulas.
- Design PRGs that fool near-maximal parity decision trees. Such PRGs would fool circuits of size $2.49n$ over the full binary basis, due to another simulation by Chen and Kabanets [CK16]. Currently, no nontrivial PRGs are known that fool circuits over the full binary basis.
- Improve the seed length in Lemma 4.5 (the balanced partition generator) to $O(\log(k/\delta) + \log\log n)$. This would not have any effect on our main theorems, but it is a natural problem in its own right.
- Prove tight bounds on the optimal nonexplicit seed length of PRGs fooling depth-k decision trees with error ε when k and $\log(1/\varepsilon)$ are both large. For example, does there exist a PRG that fools decision trees of depth $k = 0.9 \cdot n$ with error $\varepsilon = 2^{-0.4n}$ and seed length $(1 - \Omega(1)) \cdot n$?
- Prove matching upper and lower bounds on the power of small-bias distributions to fool decision trees. For example, does there exist a constant $c < 1/2$ such that every n-wise (2^{-cn})-biased distribution fools decision trees of depth $n/2$ with error 0.1?

8 Acknowledgments

I thank Avishay Tal for valuable comments on a draft of this paper and for a discussion about the Fourier spectra of decision trees. I thank Alicia Torres Hoza for helpful comments on drafts of this paper.

References

[AAKMRX07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. "Testing k-wise and almost k-wise independence". In: *Proceedings of the 39th Annual Symposium on Theory of Computing (STOC)*. 2007, pp. 496–505. DOI: 10.1145/1250790.1250863.

[ABI86] Noga Alon, László Babai, and Alon Itai. "A fast and simple randomized parallel algorithm for the maximal independent set problem". In: *J. Algorithms* 7.4 (1986), pp. 567–583. ISSN: 0196-6774. DOI: 10.1016/0196-6774(86)90019-2.

[ABNNR92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. "Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs". In: *IEEE Transactions on Information Theory* 38.2 (1992), pp. 509–516. DOI: 10.1109/18.119713.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. "Simple constructions of almost k-wise independent random variables". In: *Random Structures Algorithms* 3.3 (1992), pp. 289–304. ISSN: 1042-9832. DOI: 10.1002/rsa.3240030308.

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. "Almost k-wise independence versus k-wise independence". In: *Inform. Process. Lett.* 88.3 (2003), pp. 107–110. ISSN: 0020-0190. DOI: 10.1016/S0020-0190(03)00359-4.

[AKK16] Andris Ambainis, Martins Kokainis, and Robin Kothari. "Nearly Optimal Separations Between Communication (or Query) Complexity and Partitions". In: *Proceedings of the 31st Conference on Computational Complexity (CCC)*. 2016, 4:1–4:14. DOI: 10.4230/LIPIcs.CCC.2016.4.

[Alo09] Noga Alon. "Perturbed identity matrices have high rank: proof and applications". In: *Combin. Probab. Comput.* 18.1-2 (2009), pp. 3–15. ISSN: 0963-5483. DOI: 10.1017/S0963548307008917.

[Alo86] N. Alon. "Explicit construction of exponential sized families of k-independent sets". In: *Discrete Math.* 58.2 (1986), pp. 191–193. ISSN: 0012-365X. DOI: 10.1016/0012-365X(86)90161-5.

[BD22] Guy Blanc and Dean Doron. "New Near-Linear Time Decodable Codes Closer to the GV Bound". In: *Proceedings of the 37th Annual Computational Complexity Conference (CCC)*. 2022, 10:1–10:40. DOI: 10.4230/LIPIcs.CCC.2022.10.

[BS88] Bernd Becker and Hans-Ulrich Simon. "How robust is the n-cube?" In: *Inform. and Comput.* 77.2 (1988), pp. 162–178. ISSN: 0890-5401. DOI: 10.1016/0890-5401(88)90056-9.

[Bsh14] Nader H. Bshouty. "Testers and their applications [extended abstract]". In: *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science (ITCS)*. ACM, New York, 2014, pp. 327–351. DOI: 10.1145/2554797.2554828.

[Bsh16] Nader H. Bshouty. *Derandomizing Chernoff Bound with Union Bound with an Application to k-wise Independent Sets*. 2016. arXiv: 1608.01568 [cs.DM].

[BT13] Avraham Ben-Aroya and Amnon Ta-Shma. "Constructing small-bias sets from algebraic-geometric codes". In: *Theory Comput.* 9 (2013), pp. 253–272. DOI: 10.4086/toc.2013.v009a005.

[BV10] Andrej Bogdanov and Emanuele Viola. "Pseudorandom bits for polynomials". In: *SIAM J. Comput.* 39.6 (2010), pp. 2464–2486. ISSN: 0097-5397. DOI: 10.1137/070712109.

[CGHFRS85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Freidmann, Steven Rudich, and Roman Smolensky. "The bit extraction problem or t-resilient functions". In: *Proceedings of the 26th Annual Symposium on Foundations of Computer Science (FOCS)*. 1985, pp. 396–407. DOI: 10.1109/SFCS.1985.55.

[CK16] Ruiwen Chen and Valentine Kabanets. "Correlation bounds and #SAT algorithms for small linear-size circuits". In: *Theoret. Comput. Sci.* 654 (2016), pp. 2–10. ISSN: 0304-3975. DOI: 10.1016/j.tcs.2016.05.005.

[CKMZ83] Ashok K. Chandra, Lawrence T. Kou, George Markowsky, and Shmuel Zaks. "On sets of Boolean n-vectors with all k-projections surjective". In: *Acta Inform.* 20.1 (1983), pp. 103–111. ISSN: 0001-5903. DOI: 10.1007/BF00264296.

[CRSW13] L. Elisa Celis, Omer Reingold, Gil Segev, and Udi Wieder. "Balls and bins: smaller hash families and faster evaluation". In: *SIAM J. Comput.* 42.3 (2013), pp. 1030–1050. ISSN: 0097-5397. DOI: 10.1137/120871626.

[CW79] J. Lawrence Carter and Mark N. Wegman. "Universal classes of hash functions". In: *J. Comput. System Sci.* 18.2 (1979), pp. 143–154. ISSN: 0022-0000. DOI: 10.1016/0022-0000(79)90044-8.

[GKM18] Parikshit Gopalan, Daniel M. Kane, and Raghu Meka. "Pseudorandomness via the discrete Fourier transform". In: *SIAM J. Comput.* 47.6 (2018), pp. 2451–2487. ISSN: 0097-5397. DOI: 10.1137/16M1062132.

[GKST18] Alexander Golovnev, Alexander S. Kulikov, Alexander V. Smal, and Suguru Tamaki. "Gate elimination: circuit size lower bounds and #SAT upper bounds". In: *Theoret. Comput. Sci.* 719 (2018), pp. 46–63. ISSN: 0304-3975. DOI: 10.1016/j.tcs.2017.11.008.

[GPW18] Mika Göös, Toniann Pitassi, and Thomas Watson. "Deterministic communication vs. partition number". In: *SIAM J. Comput.* 47.6 (2018), pp. 2435–2450. ISSN: 0097-5397. DOI: 10.1137/16M1059369.

[HH24] Pooya Hatami and William Hoza. "Paradigms for unconditional pseudorandom generators". In: *Found. Trends Theor. Comput. Sci.* 16.1-2 (2024), pp. 1–210. ISSN: 1551-305X. DOI: 10.1561/0400000109.

[HHVESS24] Itamar Harel, William M. Hoza, Gal Vardi, Itay Evron, Nathan Srebro, and Daniel Soudry. *Provable Tempered Overfitting of Minimal Nets and Typical Nets*. 2024. arXiv: 2410.19092 [cs.LG].

[IM02] Kazuo Iwama and Hiroki Morizumi. "An explicit lower bound of $5n - o(n)$ for Boolean circuits". In: *Proceedings of the 27th International Symposium on Mathematical Foundations of Computer Science (MFCS)*. Vol. 2420. Lecture Notes in Comput. Sci. Springer, Berlin, 2002, pp. 353–364. DOI: 10.1007/3-540-45687-2_29.

[KM93] Eyal Kushilevitz and Yishay Mansour. "Learning decision trees using the Fourier spectrum". In: *SIAM J. Comput.* 22.6 (1993), pp. 1331–1348. ISSN: 0097-5397. DOI: 10.1137/0222080.

[KM97] Howard Karloff and Yishay Mansour. "On construction of k-wise independent random variables". In: *Combinatorica* 17.1 (1997), pp. 91–107. ISSN: 0209-9683. DOI: 10.1007/BF01196134.

[KRDS15] Robin Kothari, David Racicot-Desloges, and Miklos Santha. "Separating decision tree complexity from subcube partition complexity". In: *Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM)*. 2015, pp. 915–930. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2015.915.

[KS73] Daniel J. Kleitman and Joel Spencer. "Families of k-independent sets". In: *Discrete Math.* 6 (1973), pp. 255–262. ISSN: 0012-365X. DOI: 10.1016/0012-365X(73)90098-8.

[LRTV09] Shachar Lovett, Omer Reingold, Luca Trevisan, and Salil Vadhan. "Pseudorandom bit generators that fool modular sums". In: *Proceedings of the 13th International Workshop on Randomization and Computation (RANDOM)*. 2009, pp. 615–630. DOI: 10.1007/978-3-642-03685-9_46.

[MNT93] Yishay Mansour, Noam Nisan, and Prasoon Tiwari. "The computational complexity of universal hashing". In: *Theoretical Computer Science* 107.1 (1993), pp. 121–133. DOI: 10.1016/0304-3975(93)90257-T.

[MZ13] Raghu Meka and David Zuckerman. "Pseudorandom generators for polynomial threshold functions". In: *SIAM J. Comput.* 42.3 (2013), pp. 1275–1301. ISSN: 0097-5397. DOI: 10.1137/100811623.

[NN93] Joseph Naor and Moni Naor. "Small-bias probability spaces: efficient constructions and applications". In: *SIAM J. Comput.* 22.4 (1993), pp. 838–856. ISSN: 0097-5397. DOI: 10.1137/0222053.

[NSS95] Moni Naor, Leonard J. Schulman, and Aravind Srinivasan. "Splitters and near-optimal derandomization". In: *Proceedings of 36th Annual Conference on Foundations of Computer Science (FOCS)*. 1995, pp. 182–191. DOI: 10.1109/SFCS.1995.492475.

[Nur09] Sergey Nurk. *An $O(2^{0.4058m})$ upper bound for circuit SAT*. PDMI technical report. 2009. URL: http://www.pdmi.ras.ru/preprint/2009/09-10.html.

[OZ18] Ryan O'Donnell and Yu Zhao. "On Closeness to k-Wise Uniformity". In: *Proceedings of the 22nd International Conference on Randomization and Computation (RANDOM)*. 2018, 54:1–54:19. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2018.54.

[Rao47] C. Radhakrishna Rao. "Factorial experiments derivable from combinatorial arrangements of arrays". In: *Suppl. J. Roy. Statist. Soc.* 9 (1947), pp. 128–139. ISSN: 1466-6162. DOI: 10.2307/2983576.

[Sav02] Petr Savický. *On determinism versus unambiquous nondeterminism for decision trees*. ECCC preprint TR02-009. 2002. URL: https://eccc.weizmann.ac.il/report/2002/009/.

[SB88] Gadiel Seroussi and Nader H. Bshouty. "Vector sets for exhaustive testing of logic circuits". In: *IEEE Trans. Inform. Theory* 34.3 (1988), pp. 513–522. ISSN: 0018-9448. DOI: 10.1109/18.6031.

[Sko22] Maciej Skorski. "Tight Chernoff-Like Bounds Under Limited Independence". In: *Proceedings of the 26th International Conference on Randomization and Computation (RANDOM)*. 2022, 15:1–15:14. DOI: 10.4230/LIPIcs.APPROX/RANDOM.2022.15.

[SSS95] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. "Chernoff-Hoeffding bounds for applications with limited independence". In: *SIAM J. Discrete Math.* 8.2 (1995), pp. 223–250. ISSN: 0895-4801. DOI: 10.1137/S089548019223872X.

[SVW17] Thomas Steinke, Salil Vadhan, and Andrew Wan. "Pseudorandomness and Fourier-growth bounds for width-3 branching programs". In: *Theory Comput.* 13 (2017), Paper No. 12, 50. DOI: 10.4086/toc.2017.v013a012.

[Ta-17] Amnon Ta-Shma. "Explicit, almost optimal, epsilon-balanced codes". In: *Proceedings of the 49th Annual Symposium on Theory of Computing (STOC)*. 2017, pp. 238–251. DOI: 10.1145/3055399.3055408.

[TW83] Donald T. Tang and Lin S. Woo. "Exhaustive Test Pattern Generation with Constant Weight Vectors". In: *IEEE Transactions on Computers* C-32.12 (1983), pp. 1145–1150. DOI: 10.1109/TC.1983.1676175.

[WC81] Mark N. Wegman and J. Lawrence Carter. "New hash functions and their use in authentication and set equality". In: *J. Comput. System Sci.* 22.3 (1981). Special issue dedicated to Michael Machtey, pp. 265–279. ISSN: 0022-0000. DOI: 10.1016/0022-0000(81)90033-7.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il