# Corrigendum: Explicit Construction of a Small $\varepsilon$-Net for Linear Threshold Functions

Yuval Rabani* Amir Shpilka†

**Abstract**

We give explicit constructions of $\varepsilon$-nets for linear threshold functions on the binary cube and on the unit sphere. The size of the constructed nets is polynomial in the dimension $n$ and in $\frac{1}{\varepsilon}$. To the best of our knowledge no such constructions were previously known. Our results match, up to the exponent of the polynomial, the bounds that are achieved by probabilistic arguments. As a corollary we also construct subsets of the binary cube that have size polynomial in $n$ and a covering radius of $\frac{n}{2} - c\sqrt{n \log n}$ for any constant $c$. This improves upon the well-known construction of dual BCH codes that guarantee only a covering radius of $\frac{n}{2} - c\sqrt{n}$.

## 1 Introduction

Influenced by the discovery of unexpected connections linking fundamental questions in geometric functional analysis to problems in theoretical computer science, there has been recent interest in explicit or algorithmic construction of certain geometric objects that are known to exist via probabilistic arguments. For example, the celebrated dimension reduction lemma of Johnson and Lindenstrauss [JL84] has been derandomized using the method of conditional expectations [EIO02, Siv02]. Another example that is still mostly open is the construction of high dimensional nearly-Euclidean linear subspaces of $\ell_1^n$ [Ind07, GLR08, GLW08]. This problem is related to the question of constructing compressed sensing schemes [Don06]; other probabilistic compressed sensing schemes, using the restricted isometry property [CT06], also exhibit a geometric flavor. All these geometric objects have numerous applications in areas such as coding theory and data compression, communication complexity, nearest neighbor search, learning theory, and computational linear algebra (see, e.g., the introduction of [GLR08]), hence the desire to discover explicit constructions.

In this paper we study what is perhaps the simplest such question. We construct $\varepsilon$-nets for linear threshold functions on the binary cube $\mathcal{B}_n = \{-1, +1\}^n$ as well as on the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$. A function $f : \mathbb{R}^n \to \{-1, 1\}$ is called a linear threshold function (LTF) iff for some $v \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$ we have that $f(x) = 1$ iff $\langle v, x \rangle \geq \theta$. Notice that when restricted to $\mathbb{S}^{n-1}$, an LTF is simply the indicator function of a closed spherical cap of $\mathbb{S}^{n-1}$. Given a measurable set $\Omega \subset \mathbb{R}^n$ endowed with a measure $\mu$ and a family $\mathcal{F}$ of measurable subsets of $\Omega$, an $\varepsilon$-net for $\mathcal{F}$ is a set $S \subset \Omega$ such

*The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel (yrabani@cs.huji.ac.il). This author's research was supported by Israel Science Foundation grant 1109/07 and by US-Israel Binational Science Foundation grant 2008059.

†Computer Science Department, Technion, Haifa 32000, Israel (shpilka@cs.technion.ac.il). This author's research was supported by the Israel Science Foundation (grant 439/06).

that for every $F \in \mathcal{F}$ with $\mu(F) > \varepsilon$, we have that $|S \cap F| > 0$.[1] Constructing $\varepsilon$-nets for natural set systems $(\Omega, \mu, \mathcal{F})$ has been studied extensively in some cases. For example, the case where $\Omega$ is the convex hull of a $d$-points set $P$ and $\mathcal{F}$ is the family of all convex hulls of subsets of $P$ received a lot of attention (see, e.g., [Cha94, AKN+08]). The case where $\Omega = [m]^d$ and the set $\mathcal{F}$ is the set of all combinatorial rectangles also received a lot of attention [EGL+92, LLSZ97]. Finally, $\varepsilon$-nets were extensively studied for *fixed* dimensions; see, e.g., [CM96]. To the best of our knowledge, the case of LTFs (in high dimensions) has not been previously considered in this context.

We consider $\Omega$, which is either the binary cube or the unit sphere (endowed with the uniform measure), and the family $\mathcal{F}$ includes the subsets $A_f = \{x \in \Omega : f(x) = 1\}$ for all LTFs $f$. We construct $S \subset \Omega$ of cardinality $\mathsf{poly}(n, 1/\varepsilon)$ that includes a point from $A_f$ for every LTF $f$ that satisfies $\mu(A_f) \geq \varepsilon$, where $\mu$ is the uniform measure on $\Omega$. A random sample of $O(n/\varepsilon)$ points is an $\varepsilon$-net with high probability,[2] and our goal is to construct such a set explicitly. We prove the following theorem.

**Theorem 1.1.** *There exist two universal constants $a, b > 0$ such that for every $\varepsilon > 0$ there is an explicit construction[3] of an $\varepsilon$-net, $N_\varepsilon \subset \mathcal{B}_n$, for LTFs of size*

$$|N_\varepsilon| = O(\varepsilon^{-b} \cdot n^a).$$

Note that when $\varepsilon = 1/\mathsf{poly}(n)$ the construction above yields a polynomial sized set. As a corollary of our construction, we get a similar construction for the unit sphere.

**Theorem 1.2.** *There exist two universal constants $a, b > 0$ such that for every $\varepsilon = \exp(-O(\sqrt{n}))$ there is an explicit construction of an $\varepsilon$-net, $S_\varepsilon \subset \mathrm{S}^{n-1}$, for spherical caps of size $|S_\varepsilon| = O(\varepsilon^{-b} \cdot n^a)$.*

As another corollary of our construction we also construct a $\mathsf{poly}(n)$ size subset of $\mathcal{B}_n$ with covering radius of $\frac{n}{2} - \Omega(\sqrt{n \log n})$. The covering radius $r$ of a set of points $S \subset \mathcal{B}_n$ is the smallest $\rho$ such that for every $x \in \mathcal{B}_n$ there is some $s \in S$ with $\mathcal{H}(x, s) \leq \rho$, where $\mathcal{H}$ denotes Hamming distance. We note that this construction improves upon the one guaranteed by dual BCH codes. This result was independently obtained by Alon [Alo08].

**Corollary 1.3.** *There exists $a > 0$ such that for every $c > 0$ there is an explicit construction of a set $C \subset \mathcal{B}_n$ of size $|C| = n^2 \cdot (n^c)^a$ such that for every $z \in \mathcal{B}_n$ there is some $x \in C$ with $\mathcal{H}(z, x) \leq \frac{n}{2} - \sqrt{cn \log n}$.*

We note that LTFs play an important role in both theory and practice. For example, bounded depth $\mathrm{TC}^0$ circuits, composed of a constant number of layers of threshold functions, received considerable attention in complexity theory, and support vector machines use threshold functions as a hypothesis in many learning scenarios. Aside from the intrinsic interest in studying LTFs, our work is motivated by the desire to build methodically a theory of pseudorandom generators for geometric functions. In the algebraic setting (over GF[2]), $\varepsilon$-biased sample spaces fool linear functions [NN93]; they were recently composed to construct pseudorandom generators for low-degree polynomials [Vio08]. Analogously, we hope that dealing with LTFs is a good starting point for the gradual construction of more complicated pseudorandom generators for nonlinear geometric functions, which are needed to resolve some of the questions mentioned earlier.

[1] Such a set $S$ is also called a *hitting set* in the CS literature, however in the geometric setting we prefer the notion of $\varepsilon$-nets.

[2] This follows, for example, from applying a VC dimension argument; see [AS08].

[3] Whenever we say that an "explicit construction" exists we mean that there is a polynomial time algorithm that on input $1^n$ outputs the required construction. It is not difficult to verify that the constructions in this work can also be performed by a log-space machine.

### 1.1 Proof technique

Our constructions use several ideas from derandomization theory. The first is the notion of a $k$-wise independent distribution. A set of $m$ random variables on a sample space $\Omega$ is $k$-wise independent iff every subset of the random variables of cardinality at most $k$ is independent. There are numerous applications in computer science for $k$-wise independent distributions with small support. In particular, $\mathsf{poly}(n)$ size $k$-wise independent distributions on $\mathcal{B}_n$ give a construction of a covering code with covering radius $\frac{n}{2} - \Omega(\sqrt{n})$. We improve the covering radius of a $\mathsf{poly}(n)$ size set to $\frac{n}{2} - \Omega(\sqrt{n \log n})$. The idea is to concatenate $O(\log n)$ samples from a 4-wise independent distribution with $m = n/O(\log n)$ random variables. In order to restrict the size of the constructed set, we need to consider only a subset of all possible concatenations. In the case of the covering code we actually concatenate the same element with itself $O(\log n)$ times. More accurately, for every element $s \in S$ and every sequence of $t = O(\log n)$ signs $\alpha_1, \ldots, \alpha_t$ we consider $\alpha_1 \cdot s \circ \cdots \circ \alpha_t \cdot s$. Thus, every element of $S$ gives rise to $2^t = \mathsf{poly}(n)$ vectors in the Boolean cube.

Note that this idea does not yield an $\varepsilon$-net. Indeed, a covering code is an $\varepsilon$-net for LTFs of the form $f(x) = \mathrm{sign}(\langle x, v \rangle - \theta)$ (for an appropriate $\theta$) when $v \in \mathcal{B}_n$. However, when $v$ is taken from the unit sphere and all the weight of $v$ is concentrated on the first $n/O(\log n)$ coordinates, the inner product of $v$ and the "self-concatenated" $s$ will be off by a factor of $O(\sqrt{\log n})$. To overcome this and to ensure that the weight of $v$ is "spread" we first hash the coordinates of $v$ into $O(\log n)$ "buckets" such that each of them contains approximately the same weight of coefficients as the other sets. To get a small set of partitions, we use certain explicit constructions of perfect hash functions. Once we have this "reordering" of the coordinates of $v$ we would like to repeat the idea from before. However, an additional component is needed. Instead of concatenating each $s \in S$ to itself (with different signs) we instead pick a subset of $O(\log n)$ elements of $s$ and concatenate them together. As we do not wish to go over all such $O(\log n)$-tuples we use walks on an expander to pick those sets (in fact, we could have used any sampler and not just expander walks here).

The analysis of the construction of $\varepsilon$-nets is different from the analysis in the case of covering codes. The main difference is that the distribution of an LTF $f(x) = \mathrm{sign}(\langle x, v \rangle - \theta)$ depends on the way the weight of $v$ is distributed among its coordinates. If no subset of coordinates contains too much weight, then the analysis is similar to before. However, if there is a small subset of coordinates (say, of size $O(\log n)$) that contains most of the weight, then we need to have the correct sign on those coordinates. This reasoning gives rise to a case analysis in the spirit of an earlier work of Servedio [Ser06] where the notion of *critical index* was first used to obtain small weight approximators for LTFs. Specifically, assume that the coordinates of $v$ satisfy $|v_1| \geq |v_2| \geq \cdots \geq |v_n|$. Consider the first index $t$ such that $v_t^2 \leq O((v_{t+1}^2 + \cdots + v_n^2)/t)$. Intuitively, if $t$ is large (say, $t > \log(1/\varepsilon)$), then $v$ is roughly concentrated on its first $t$ coordinates and by hashing them to different buckets we just have to go over all possible sign assignments (to the buckets) in order to get an inner product of (roughly) $|v_1| + \cdots + |v_t|$, which is the maximum one can hope for. On the other hand, if $t$ is small, then it means that except for a few large coordinates the weight of $v$ is "spread" among many coordinates, which is similar to the case of covering codes discussed above where one studies (normalized) sign vectors.

*Organization*. In section 2 we give some formal definitions and the necessary background on $k$-wise independent distributions, expander graphs, and perfect hash functions. We also give some concentration results for threshold functions. In section 3 we give the construction of a covering code. In section 4 we give our main construction for linear threshold functions and in section 5 we give the construction for spherical caps.

### 1.2 Subsequent works

Following our work [RS09] several other papers looked at the problem of obtaining pseudorandom generators for LTFs on the Boolean cube. [DGJ+09] showed that $k$-wise distributions $\varepsilon$-fool LTFs, where $k = O(\log^2(1/\varepsilon)/\varepsilon^2)$, namely, a $k$-wise independent distribution contains the "correct" number of accepting inputs of any LTF up to error (roughly) $\exp(-\sqrt{k})$. Note however that the size of such sets is $n^{\Omega(k)}$ and so this gives a polynomial size construction only when $\varepsilon$ is a constant (and of course, the exponent of the construction depends on $\varepsilon$).[4] More recently, [MZ09] gave a construction of a pseudorandom generator for *polynomial* threshold functions, namely, functions that are the sign of a low-degree polynomial. The size of their construction is around $n^{1/\varepsilon^d}$ for error $\varepsilon$. For the special case of LTFs they obtain a pseudorandom generator of size $n^{O(\log 1/\varepsilon)}$ for $\varepsilon > 1/\mathsf{poly}(n)$ and of polynomial size whenever $\varepsilon > 1/\mathsf{poly}\log(n)$. Compared to our construction they obtain a pseudorandom generator where we obtain only a hitting set. On the other hand, our construction is of polynomial size even for a polynomially small $\varepsilon$.

## 2 Preliminaries

We will use the following notation. The $n$-dimensional binary cube is $\mathcal{B}_n = \{-1,1\}^n$. The $(n-1)$-dimensional unit sphere in $\mathbb{R}^n$ is $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. The Hamming distance on $\mathbb{R}^n$ is denoted by $\mathcal{H}$, so $\mathcal{H}(x,y)$ is the number of coordinates $i$ for which $x_i \neq y_i$. For $x \in \mathbb{R}^n$ and $J = \{i_1, \ldots, i_{|J|}\} \subseteq [n]$ we denote $x_J = (x_{i_1}, \ldots, x_{i_{|J|}})$. We will abuse notation and use (for $A \subset \mathbb{R}^n$) $\mathcal{H}(x,A)$ to denote $\min_{y \in A} \mathcal{H}(x,y)$. For $A \subseteq \mathcal{B}_n$ and $\rho > 0$, we put $A_\rho = \{x \in \mathcal{B}_n : \mathcal{H}(x,A) \leq \rho\}$. The covering radius of a set $C \subset \mathcal{B}_n$ is the minimum $\rho$ such that $C_\rho = \mathcal{B}_n$, namely, it is the minimal $\rho$ such that for every $x \in \mathcal{B}_n$ there is $y \in A$ with $\mathcal{H}(x,y) \leq \rho$.

In this paper we focus on LTFs. A vector $v \in \mathbb{R}^n$ and a real number $\theta \in \mathbb{R}$ define an LTF $L_{v,\theta} : \mathcal{B}_n \to \{-1,1\}$ by $L_{v,\theta}(x) = \mathrm{sign}(\langle v, x\rangle - \theta)$. In other words, $L_{v,\theta}(x) = 1$ if $\langle v, x\rangle \geq \theta$ and $L_{v,\theta}(x) = -1$ otherwise. For a linear function $L_{v,\theta}$ we define by $A_{v,\theta}$ its set of accepting inputs, namely, $A_{v,\theta} = L_{v,\theta}^{-1}(1) = \{x \in \mathcal{B}_n : \langle v, x\rangle \geq \theta\}$. A spherical cap in $\mathbb{R}^n$ is a subset of $\mathbb{S}^{n-1}$ that is contained in a half-space, namely, for every $v \in \mathbb{R}^n$ and $\theta > 0$ the cap $C_{v,\theta}$ is defined as $C_{v,\theta} = \{x \in \mathbb{S}^{n-1} : \langle v, x\rangle \geq \theta\}$. Stated differently, $C_{v,\theta} = L_{v,\theta}^{-1}(1) \cap \mathbb{S}^{n-1}$ (we now think of $L_{v,\theta}$ as a function from $\mathbb{R}^n$ to $\{-1,1\}$).

### 2.1 $k$-wise independent distributions

A multiset $I \subset \{-1,1\}^n$ that, for every $j \in \{1,2,\ldots,k\}$, for every $\{i_1, i_2, \ldots, i_j\} \subset \{1,2,\ldots,n\}$, and for every $z_1, z_2, \ldots, z_j \in \{-1,1\}$, satisfies that

$$\left|\left\{x \in I : \ (x_{i_1}, x_{i_2}, \ldots, x_{i_j}) = (z_1, z_2, \ldots, z_j)\right\}\right| = \frac{|I|}{2^j}$$

is called a $k$-wise independent sample space. Many explicit constructions of small $k$-wise independent sample spaces are known. For example, extended binary BCH codes of length $n = 2^m - 1$ and designed distance $2t+2$ can be used to construct a $(2t+1)$-wise independent sample space of size $2^{mt+1} = 2(n+1)^t$ (see [AS08, Chapter 16]).

[4] When speaking of pseudorandom generators one usually considers the seed length. However, to ease the comparison to our result we consider the size of the image of the pseudorandom generator.

**Fact 2.1.** *For every integer $k > 0$ there exists an explicit construction of a sample space of size $O(n^{k/2})$ that is $k$-wise independent.*

Let a multiset $S \subseteq \{-1,1\}^n$ be a $k$-wise independent sample space. The following is an easy observation.

**Observation 2.2.** *For $i \in [n]$ and $\alpha \in \{-1,1\}$, restricted to coordinates $[n] \setminus \{i\}$, the multiset $S_{i,\alpha} := \{x \in S : x_i = \alpha\}$ is a $(k-1)$-wise independent sample space.*

The following result was proved by Berger in [Ber97].

**Lemma 2.3** (see Lemma 3.1 in [Ber97])**.** *Let $S \subset \{-1,1\}^n$ be a 4-wise independent sample space. Then for every $x \in \mathbb{S}^{n-1}$ we have that $\mathsf{EE}[\langle s,x\rangle] = 0$, $\mathsf{EE}[\langle s,x\rangle^2] = 1$, and $\mathsf{EE}[\langle s,x\rangle^4] \leq 3$, where all expectations are with respect to a uniform choice of $s \in S$. Moreover, for every $x \in \mathbb{R}^n$ we have that*

$$\Pr_{s\in S}\left[|\langle s,x\rangle| > \frac{\|x\|_2}{\sqrt{3}}\right] \geq \frac{2}{11}.$$

The following lemma is a special case of a lemma of Alon, Gutin, and Krivelevich [AGK04].

**Lemma 2.4** (see Lemma 3.2 in [AGK04])**.** *Let $X$ be a real random variable and suppose that its first, second, and fourth moments satisfy $\mathsf{EE}[X] = 0$, $\mathsf{EE}[X^2] = 1$, and $\mathsf{EE}[X^4] \leq 3$. Then $\Pr[X > 1/7] \geq 1/20$. Consequently, if $S \subset \{-1,1\}^n$ is a 4-wise independent sample space, then for every $x \in \mathbb{S}^{n-1}$ we have that*

$$\Pr_{s\in S}[\langle s,x\rangle > 1/7] \geq 1/20.$$

Next is an easy corollary of Observation 2.2 and Lemma 2.4 that gives an anticoncentration result for LTFs. As the distribution of LTFs on the Boolean cube is very different from their distribution on the sphere (e.g., compare the distribution of $f(x) = \langle x,v\rangle$ for $v = (1,0,\ldots,0)$ on the sphere and cube), we need to separately handle the large coordinates of $v$ and its small coordinates.

**Lemma 2.5.** *Let $k > 4$ be an integer, $S \subseteq \{-1,1\}^n$ a $k$-wise independent sample space, and $v = (v_1,\ldots,v_n) \in \mathbb{R}^n$ a unit vector. Let $M \subset [n]$ be such that $|M| = k-4$ and the entries of $v$ corresponding to the coordinates in $M$ are the $k-4$ largest entries of $v$ (namely, for every $j \notin M$ and every $i \in M$ we have that $|v_j| \leq |v_i|$). Then*

$$\Pr_{x\in S}\left[\langle x,v\rangle \geq \|v_M\|_1 + \frac{1}{7}\|v_{[n]\setminus M}\|_2\right] \geq \frac{4}{5}\cdot 2^{-k}.$$

*Proof.* Let $S' \subset S$ be the set of all $s \in S$ such that $\operatorname{sign} s_i = \operatorname{sign} v_i$ for every $i \in M$. By definition we have that $|S'| = 2^{-|M|}\cdot|S| = |S|/2^{k-4}$. Moreover, by Observation 2.2 we get that $S'$ is 4-wise independent. Let $v' = (v'_1,\ldots,v'_n)$ be defined as $v'_i = 0$ for $i \in M$ and $v'_i = v_i$ for $i \notin M$. By Lemma 2.4 we have that

$$\Pr_{s\in S'}\left[\langle s,v'\rangle > \frac{1}{7}\|v'\|_2\right] > \frac{1}{20}.$$

By definition of $v'$ we get that $\langle s,v\rangle = \sum_{i\in M} s_i\cdot v_i + \langle s,v'\rangle = \|v_M\|_1 + \langle s,v'\rangle$. Thus,

$$\Pr_{x\in S}\left[\langle x,v\rangle \geq \|v_M\|_1 + \frac{1}{7}\|v_{[n]\setminus M}\|_2\right] \geq \frac{1}{20\cdot 2^{k-4}} = \frac{4}{5}\cdot 2^{-k}.$$

□

## 2.2 Expander graphs

An undirected graph $G = (V, E)$ is called an $(n, d, \lambda)$-expander if $|V| = n$, the degree of each node is $d$, and the second largest eigenvalue, in absolute value, of the adjacency matrix of $G$ is $\lambda$. For every $d = p + 1$, where $p$ is a prime congruent to 1 modulo 4, there are explicit constructions for infinitely many $n$ of $(n, d, \lambda)$-expanders, where $\lambda \leq 2\sqrt{d-1}$ [Mar88, LPS88].

A random walk of length $\ell$ on $G$ is the following random process. First pick a vertex of $G$ uniformly at random. Denote this vertex with $v_1$. At the $i$th step (for $1 < i \leq \ell$) we pick a neighbor of $v_{i-1}$ uniformly at random and label it with $v_i$. The walk is the ordered list $(v_1, v_2, \ldots, v_\ell)$. We shall need the following theorem of Alon et al. [AFWZ95].

**Theorem 2.6.** *Let $G$ be an $[n, d, \lambda]$-expander. Let $W_1, \ldots, W_\ell \subset V(G)$ be some subsets of $G$, each of size at least $\mu n \geq 6\lambda n/d$. The probability that a random walk of length $\ell$ stays inside $W_1, W_2, \ldots, W_\ell$ is at least $\mu(\mu - 2\lambda/d)^{\ell-1}$.*

## 2.3 Perfect hash functions

A set $H$ of functions $h : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, m\}$ such that for every $S \subset \{1, 2, \ldots, n\}$ with $|S| = s$ there exists $h \in H$ such that $|h(S)| = s$ is called an $(n, m, s)$-perfect hash family. For all $n, s \in \mathbb{N}$, $s \leq n$, there are explicit constructions of $(n, O(s), s)$-perfect hash families $H$ with $|H| = 2^{O(s+\log\log n)}$ (see Theorem 6 in [SS90]). Lemma 2.7 is a strengthening of the above requirement. Informally, the strengthened version says that we can construct $H$ to have the following property. For every vector $v = (v_1, \ldots, v_n)$ there is $h \in H$ that maps its "heaviest" $s$ coordinates (in absolute value) to different locations, and furthermore, if the remaining coordinates have sufficient $L_2$ mass, then it is distributed by $h$ roughly evenly among the $O(s)$ locations.

**Lemma 2.7** (perfect hash functions)**.** *There exists a universal constant $A$ such that the following holds. For every integers $s, n$ such that $s \leq n$, there is an explicit family $\mathcal{H}$ of hash functions $h : [n] \to [8s]$ of cardinality $|\mathcal{H}| = 2^{(4+o(1))\cdot s + A\cdot\log 2s\log\log n + O(1)}$ such that the following holds for every unit vector $v \in \mathbb{S}^{n-1}$. Let $i_1, i_2, \ldots, i_n$ be an enumeration of $[n]$ such that $|v_{i_1}| \geq |v_{i_2}| \geq \cdots \geq |v_{i_n}|$, and let $I_t$ denote the set $\{i_1, i_2, \ldots, i_t\}$. For every $t \in [s-1]$, there exists some $h \in \mathcal{H}$ such that*

1. *The map $h$ is an injection on $I_s$.*
2. *If $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then*

$$\sum_{r\in[8s]} \min\left\{\|v_{h^{-1}(r)\setminus I_t}\|_2^2, \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\} \geq \frac{1}{2}\cdot\|v_{[n]\setminus I_t}\|_2^2. \tag{2.8}$$

*Furthermore, the $o(1)$ term in the exponent of $|\mathcal{H}|$ depends only on $s$.*

For completeness we give the proof in the appendix. The following is an easy corollary.

**Corollary 2.9.** *Let $24 \leq s \leq n$ be integers and $\mathcal{H}$ the hash family guaranteed by Lemma 2.7. There exists constants $c_1$ and $c_2$ such that one of the following conditions holds (using the same notation as in Lemma 2.7):*

1. *either $\sum_{q=\lceil 2s/3\rceil}^{s-1} |v_{i_{q+1}}| \geq \frac{\sqrt{s}}{32}\|v_{[n]\setminus I_s}\|_2$ ;*

2. *or, there exists* $\lceil 2s/3 \rceil \leq q \leq s-1$ *and* $h \in \mathcal{H}$ *such that* $h$ *is an injection on* $I_s$ *and for at least* $c_1 \cdot 8s$ *buckets* $r$ *it holds that* $\|v_{h^{-1}(r)\setminus I_q}\|_2^2 \geq \frac{c_2}{s} \cdot \|v_{[n]\setminus I_q}\|_2^2$.

*Proof.* We consider two cases.

*Case* 1. There is some $\lceil 2s/3 \rceil \leq q \leq s-1$ such that $v_{i_{q+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_q}\|_2^2$.

*Case* 2. For every $\lceil 2s/3 \rceil \leq q \leq s-1$ we have that $v_{i_{q+1}}^2 > \frac{1}{64s} \cdot \|v_{[n]\setminus I_q}\|_2^2$.

Consider Case 1. By the assumption in Case 1 we get from Lemma 2.7 that there exists $h \in \mathcal{H}$ such that Eq (2.8) is satisfied. We will show that for some constants $c_1, c_2$ at least $c_1 \cdot 8s$ buckets satisfy that $\|v_{h^{-1}(r)\setminus I_q}\|_2^2 \geq \frac{c_2}{s} \cdot \|v_{[n]\setminus I_q}\|_2^2$. Assume for a contradiction that less than $c_1 \cdot 8s$ buckets have high norm. Hence,

$$\begin{aligned}\frac{1}{2} \cdot \|v_{[n]\setminus I_q}\|_2^2 &\leq \sum_{r \in [8s]} \min\left\{ \|v_{h^{-1}(r)\setminus I_q}\|_2^2, \frac{2}{s} \cdot \|v_{[n]\setminus I_q}\|_2^2 \right\} \\ &\leq c_1 \cdot 8s \cdot \frac{2}{s} \cdot \|v_{[n]\setminus I_q}\|_2^2 + 8s \cdot \frac{c_2}{s} \cdot \|v_{[n]\setminus I_q}\|_2^2 = (16c_1 + 8c_2) \cdot \|v_{[n]\setminus I_q}\|_2^2.\end{aligned}$$

Therefore, for $c_1 = \frac{1}{48}$ and $c_2 = \frac{1}{49}$ we get a contradiction, unless $\|v_{[n]\setminus I_q}\|_2^2 = 0$. However, the claim is trivial if this is the case.

Let us now assume that we are in Case 2. It follows that

$$\sum_{q=\lceil 2s/3 \rceil}^{s-1} |v_{i_{q+1}}| \geq \sum_{q=\lceil 2s/3 \rceil}^{s-1} \frac{1}{8\sqrt{s}} \cdot \|v_{[n]\setminus I_q}\|_2 \geq \sum_{q=\lceil 2s/3 \rceil}^{s-1} \frac{1}{8\sqrt{s}} \cdot \|v_{[n]\setminus I_s}\|_2 \geq \frac{\sqrt{s}}{32} \|v_{[n]\setminus I_s}\|_2,$$

where in the last inequality we used the assumption that $s \geq 24$. □

## 2.4 Concentration of threshold functions

In order to construct an $\varepsilon$-net for LTFs we need to understand, for every LTF $L_{v,\theta}$, for which values of $\theta$ it holds that $\Pr_{x \in \mathcal{B}_n}[L_{v,\theta}(x) = 1] > \varepsilon$. The following theorem is a standard application of the Bernstein–Chernoff–Hoeffding bound. A proof can be found, e.g., in Chapter 1 of [DP09].

**Theorem 2.10** (Bernstein–Chernoff–Hoeffding)**.** *For* $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ *and* $\theta \in (0, \infty)$ *we have that*

$$\Pr_{x \in \mathcal{B}_n}[\langle x, v \rangle > \theta] \leq \exp\left( -\frac{1}{2} \left( \frac{\theta}{\|v\|_2} \right)^2 \right).$$

The following result will be used to determine how large $\theta$ can be for a given $v \in \mathbb{R}^n$ so that $L_{v,\theta}$ accepts an $\varepsilon$ fraction of the inputs.

**Corollary 2.11.** *Let* $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ *and* $\delta \in \mathbb{R}^+$*. Assume that* $|v_1| \geq |v_2| \geq \cdots \geq |v_n|$*. Let* $1 \leq k \leq n$ *be an integer. Assume further that* $|v_k| > 0$*. Then*

$$\Pr_{x \in \mathcal{B}_n}\left[ \langle x, v \rangle \geq \|v_{[\lceil 2k/3 \rceil]}\|_1 + \delta \cdot \|v_{[n]\setminus[k]}\|_2 \right] \leq \exp(-k/18) + \exp(-\delta^2/2).$$

*Proof.* We have that

$$\begin{aligned}&\Pr_{x \in \mathcal{B}_n}\left[ \langle x, v \rangle \geq \|v_{[\lceil 2k/3 \rceil]}\|_1 + \delta \cdot \|v_{[n]\setminus[k]}\|_2 \right] \\ \leq &\Pr_{x_{[k]} \in \mathcal{B}_k}\left[ \left\langle x_{[k]}, v_{[k]} \right\rangle \geq \|v_{[\lceil 2k/3 \rceil]}\|_1 \right] + \Pr_{x_{[n]\setminus[k]} \in \mathcal{B}_{n-k}}\left[ \left\langle x_{[n]\setminus[k]}, v_{[n]\setminus[k]} \right\rangle \geq \delta \cdot \|v_{[n]\setminus[k]}\|_2 \right].\end{aligned}$$

As $|v_1| \geq |v_2| \geq \cdots \geq |v_k| > 0$ we see that in order for the inequality

$$\left\langle x_{[k]}, v_{[k]} \right\rangle \geq \|v_{[\lceil 2k/3 \rceil]}\|_1$$

to hold we must have that $\text{sign}(x_i) = \text{sign}(v_i)$ for at least $2k/3$ of the indices. Using the Bernstein–Chernoff–Hoeffding bound, we bound this probability with

$$\Pr_{x_{[k]} \in \mathcal{B}_k} \left[ \left\langle x_{[k]}, v_{[k]} \right\rangle \geq \|v_{[\lceil 2k/3 \rceil]}\|_1 \right] \leq \exp(-k/18).$$

The upper estimate

$$\Pr_{x_{[n]\setminus[k]} \in \mathcal{B}_{n-k}} \left[ \left\langle x_{[n]\setminus[k]}, v_{[n]\setminus[k]} \right\rangle \geq \delta \cdot \|v_{[n]\setminus[k]}\|_2 \right] \leq \exp(-\delta^2/2)$$

also follows immediately from the Bernstein–Chernoff–Hoeffding bound. □

When considering caps and not LTFs the results are somewhat easier. Recall that $C_{v,\theta}$ is defined as $C_{v,\theta} = \{x \in \mathbb{S}^{n-1} : \langle v, x \rangle \geq \theta\}$. For a proof of the next lemma, see, e.g., [Mat02].

**Lemma 2.12.** *Let $v \in \mathbb{S}^{n-1}$ be a unit vector. Then*

$$\Pr_{x \in \mathbb{S}^{n-1}}[x \in C_{v,\theta}] \leq \exp\left(-\frac{1}{2}n\theta^2\right),$$

*where we consider the uniform probability measure on $\mathbb{S}^{n-1}$.*

## 3 Construction of a covering code

As a warm up for the proof of Theorem 1.1 we give an explicit construction of a covering code of covering radius $\frac{n}{2} - c\sqrt{n \log n}$ for $\mathcal{B}_n$. Later we will build on the ideas of the proof to get the more general result.[5] For convenience we repeat the claim of Corollary 1.3 here.

COROLLARY 1.3. *There exists $a > 0$ such that for every $c > 0$ there is an explicit construction of a set $C \subset \mathcal{B}_n$ of size $|C| = n^2 \cdot (n^c)^a$ such that for every $z \in \mathcal{B}_n$ there is some $x \in C$ with $\mathcal{H}(z, x) \leq \frac{n}{2} - \sqrt{cn \log n}$.*

As described in the introduction the construction is based on first picking a 4-wise independent distribution $S$ on vectors of length $n/t$ (where $t$ is roughly $\log n$) and then constructing vectors of length $n$, using concatenation, from them. In this simple case we actually concatenate each vector with itself $\log n$ times, but in each copy we may take a different sign flip, so that eventually each vector in $S$ contributes $2^t$ different vectors to the covering code.

*Proof.* Fix $c > 0$, and let $n \in \mathbb{N}$. Put $t = \lceil c_1 \log n \rceil$ for a sufficiently large constant $c_1$ that will be determined later. For simplicity we assume that $t$ divides $n$. Let $J_1, J_2, \ldots, J_t$ be the partition of $[n]$ defined by $J_i = \{(i-1) \cdot n/t + 1, \ldots, i \cdot n/t\}$ (in fact, we can take the $J_i$'s to be any partition of the coordinates into $t$ disjoint sets, each of size $n/t$). Let $S \subset \{-1, 1\}^{n/t}$ be a 4-wise independent distribution. Let $m = |S|$ and recall that by Fact 2.1 we can assume that $m = O((n/t)^2)$. Denote $S = \{s_0, \ldots, s_{m-1}\}$. The set $C$ is defined as follows. For every sequence of signs $\alpha = (\alpha_1, \ldots, \alpha_t) \in$

[5] The corollary does follow immediately from Theorem 1.1 but we prove it separately to give some intuition for the proof of the theorem.

$\{-1,1\}^t$ and every $0 \le j \le m-1$, let $x^{\alpha,j} \in \mathcal{B}_n$ be defined as the concatenation $(\alpha_1 \cdot s_j) \circ \cdots \circ (\alpha_t \cdot s_j)$. That is, $x^{\alpha,j}_{J_i} = \alpha_i \cdot s_j$. In other words, we concatenate $t$ copies of the same element of $S$, with possibly different signs, for each of the $2^t$ sign patterns. The set $C$ is the collection of all the $x^{\alpha,j}$'s, i.e., $C = \{x^{\alpha,j} : \alpha \in \{-1,1\}^t, 0 \le j < m\}$. Hence, the size of $C$ is $2^t \cdot m = O((\frac{n}{t})^2 \cdot 2^t) \le n^2 \cdot n^{c_1}$.

We now proceed with the analysis of this construction. As $S$ is 4-wise independent we get by Lemma 2.3 that for every $y \in \{-1,1\}^{n/t}$

$$\Pr\left[|\langle y, s\rangle| > \sqrt{n/3t}\right] \ge \frac{2}{11}.$$

Fix $z \in \mathcal{B}_n$. Let $X_i$ denote the indicator of the event that $|\langle z_{J_i}, s_{j+i-1 \bmod m}\rangle| > \sqrt{n/3t}$ (where $0 \le j \le m-1$ is picked uniformly at random). Recall that $\mathsf{EE}[X_i] \ge 2/11$ and so, by linearity of expectation, we get that $\mathsf{EE}\left[\sum_{i=1}^t X_i\right] \ge 2t/11$. Therefore, for every $z \in \mathcal{B}_n$ there exists $j_z \in \{0,\ldots,m-1\}$ such that

$$\left|\left\{i : \ |\langle z_{J_i}, s_{j_z+i-1 \bmod m}\rangle| \ge \sqrt{n/3t}\right\}\right| \ge \frac{2t}{11}.$$

Set $\alpha \in \{-1,1\}^t$ as $\alpha_i = \mathrm{sign}(\langle z_{J_i}, s_{j_z+i-1 \bmod m}\rangle)$. It follows that

$$\left\langle z, x^{\alpha,j}\right\rangle = \sum_{i=1}^t |\langle z_{J_i}, s_{j_z+i-1 \bmod m}\rangle| \ge \frac{2t}{11}\sqrt{n/3t} \ge \frac{2\sqrt{c_1}}{11\sqrt{3}}\sqrt{n\log n}.$$

To complete the proof, set $c_1 = 400c$ to get $\left\langle z, x^{\alpha,j}\right\rangle > 2\sqrt{cn\log n}$. We thus obtain that

$$\mathcal{H}(z, x^{\alpha,j}) = \frac{n}{2} - \frac{1}{2}\left\langle z, x^{\alpha,j}\right\rangle \le \frac{n}{2} - \sqrt{cn\log n}.$$

Moreover, $|C| \le n^2 \cdot n^{c_1} = n^2 \cdot (n^c)^{400}$, as required. □

We note that by a simple application of the Chernoff bound one can show that this result is essentially tight (up to the exact setting of $a$). Indeed, given a set $C \subset \mathcal{B}_n$ and a point $s \in C$ it holds that $\Pr[|\langle x, s\rangle| \ge c\sqrt{\log n}] \le n^{-O(c^2)}$. Thus, for any fixed set $C$ of size $|C| = n^b$, if we let $c$ be $O(\sqrt{b})$, then by the union bound we get that there is some $x \in \mathcal{B}_n$ that has distance larger than $n/2 - c\sqrt{\log n}$ from all points in $S$.

# 4 The main construction

We now give an explicit construction of an $\varepsilon$-net set $N_\varepsilon \subset \mathcal{B}_n$ for LTFs. In particular we will prove Theorem 1.1. For convenience we repeat it here.

THEOREM 1.1. *There exist two universal constants $a, b > 0$ such that for every $\varepsilon > 0$ there is an explicit construction of an $\varepsilon$-net, $N_\varepsilon \subset \mathcal{B}_n$, for LTFs of size $|N_\varepsilon| = O(\varepsilon^{-b} \cdot n^a)$.*

Before giving the construction we explain what changes are needed from the earlier construction of the covering code. Consider a vector $v' \in \{-1,1\}^{n/\log n}$ and let $v$ be the unit vector in $\mathbb{R}^n$ having $v'/\|v'\|_2$ in its first $n/\log n$ coordinates and zeros elsewhere. Consider the linear function $L_{v,\sqrt{\log n}} : \mathcal{B}_n \to \{-1,1\}$. It is not hard to see that with probability $1/\mathsf{poly}(n)$ over the choice of $x \in \mathcal{B}_n$ we have that $L_{v,\sqrt{\log n}}(x) = 1$ for every such $v$. On the other hand, there exists a $v'$

(and actually a random $v'$ will have the required property) such that for every $y \in C$, where $C$ is the covering code constructed in section 3, we will have that $|\langle y, v\rangle| = O(1)$. Thus, for every $y \in C$ we have that $L_{v,\sqrt{\log n}}(y) = 0$. Therefore $C$ is not a $1/\mathsf{poly}(n)$-net. The reason for the failure of $C$ is that all the large coordinates of $v$ were concentrated on a set of size $n/\log n$ that was one of the sets in the partition of the coordinates with respect to which we constructed $C$. To overcome this difficulty we construct sets in a way analogous to the construction of $C$ but with respect to different partitions of the $n$ coordinates. These partitions will come from the family of perfect hash functions discussed in section 2.3. Another change that we will have to make is in the way that we concatenate short strings (of length $O(n/\log n)$ in order to get length $n$ strings. Previously we simply concatenated consecutive strings. Now we will have to concatenate them according to an expander walk, the reason being that there will be $O(\log n)$ sets in the partitions from which we will have to make sure that we get the "correct" contribution. We now turn to the actual construction (also replacing $1/\mathsf{poly}(n)$ with $\varepsilon$).

*Proof.* Let $\varepsilon > 0$ be given. We assume that $\varepsilon > 2^{-n/100}$ as otherwise we can pick $N_\varepsilon = \mathcal{B}_n$. Let $t = \lceil c \log 2/\varepsilon \rceil$ for some absolute constant $c$ that will be determined later. We assume w.l.o.g. that $t \geq 24$. We will later need this assumption (without explicitly referring to it) for applying the result of Corollary 2.9. Set $k = 5$ and $d = 2^{18}$. Similarly to the case of covering codes, let $S \subset \{-1, 1\}^n$ be a $k$-wise independent sample space. Let $m = |S|$. By Fact 2.1 we can assume that $m = |S| = O(n^{k/2})$. Denote $S = \{s_i\}_{i=1}^m$. As mentioned above we will need to consider many different partitions of the coordinates, so let $\mathcal{H}$ be the $(n, 8t, t)$-perfect hash family guaranteed by Lemma A.1. We think of every $h \in \mathcal{H}$ as partitioning the coordinates to $8t$ sets $\{J_{h,1}, \ldots, J_{h,8t}\}$ with $J_{h,i} = h^{-1}(i)$. Let $J_h = \{J_{h,1}, \ldots, J_{h,8t}\}$ be the collection of the sets in the partition. Note that the sets in $J_h$ are not necessarily of the same size. In order to concatenate elements of $S$ to create a word in $\mathcal{B}_n$ we need to consider walks on an expander graph. Let $G$ be an $(m, d, d/1000)$-expander with node set $S$. In other words, we identify the $i$th node of $G$ with $s_i$. In particular a walk $(w_1, \ldots, w_\ell)$ on $G$ is a sequence of $\ell$ elements from $S$. We now explain how to mix all these ingredients together to get the final construction.

The set $N_\varepsilon$ contains all the points $x^{h,w}$ (that will soon be defined), where $h \in H$ and $w$ is a walk of length $8t$ in $G$. We now explain how to construct $x^{h,w}$. Let $h \in \mathcal{H}$ be a hash function and let $w = (w_1, \ldots, w_{8t}) \in S^{8t}$ be a walk on $G$. Let $i \in \{1, 2, \ldots, 8t\}$. Let $w_i'$ be the first $|J_{h,i}|$ bits of $w_i$. The reason for this is that it may be the case (and it is most likely the case) that $|J_{h,i}| < n$ and so we need to cut the last bits of $w_i$ to get a vector of length exactly $|J_{h,i}|$. We now define

$$x^{h,w}|_{J_{h,i}} = w_i' = \text{ first } |J_{h,i}| \text{ bits of } w_i.$$

As the collection $\{J_{h,i}\}_{i=1}^{8t}$ is a partition of $[n]$ we get that indeed $x^{h,w} \in \mathcal{B}_n$.

A good way to understand the construction is the following. We would like to define a point $x = x^{h,w} \in N_\varepsilon$. To do so we first map the coordinates of $x$ to $8t$ buckets according to $h$. Assume that the set $J_{h,i}$ was mapped to the $i$th bucket. Now, we would like to assign a value to $x_{J_{h,i}}$ from the $k$-wise independent set $S$, and we would like to do so for every $i \in [8t]$. As there are $m^{8t}$ possibilities for such assignments we have to pick a small subset of all possible assignments. We do so by taking an expander walk on an expander with $m$ vertices. Given a walk $w = (w_1, \ldots, w_{8t})$ of length $8t$ we would like to consider the assignment $x_{J_{h,i}} = w_i$. The final thing to notice is that $|J_{h,i}|$ may be smaller than $n$ and so we consider only the first $|J_{h,i}|$ bits of $w_i$. Going over all $i \in [8t]$

we get the vector $x^{h,w}$. An easy bound on the size of $N_\varepsilon$ is

$$\begin{aligned}|N_\varepsilon| = |\mathcal{H}| \cdot d^{8t-1} \cdot m &= O\left(2^{(4+o(1))\cdot t + A\log 2t \log\log n} \cdot d^7 \cdot (2/\varepsilon)^{8c\log d} \cdot n^{k/2}\right) \\ &= O\left(n^a \cdot (1/\varepsilon)^b\right),\end{aligned}$$

for any constants $a > k/2$ and $b > 4c + 8c\log d$. We now show that $N_\varepsilon$ is an $\varepsilon$-net for LTFs. Let $L_{v,\theta}$ be an LTF, where $\|v\|_2 = 1$, such that

$$\Pr_{x\in\mathcal{B}_n}[L_{v,\theta}(x) = 1] \geq \varepsilon.$$

Let $i_1, i_2, \ldots, i_n$ be an enumeration of $[n]$ such that $|v_{i_1}| \geq |v_{i_2}| \geq \cdots \geq |v_{i_n}|$, and let $I_r$ denote the set $\{i_1, i_2, \ldots, i_r\}$. We now show that there exists $x^{h,w}$ in $N_\varepsilon$ for which $L_{v,\theta}(x^{h,w}) = 1$, which implies that $N_\varepsilon$ is an $\varepsilon$-net for LTFs.

We analyze three different cases. The first is when the support of $v$ is small. The second is when the support is not too small, but most of the mass of $v$ is concentrated on a few coordinates (this case corresponds to the first point in Corollary 2.9). The last case is when the mass of $v$ is "nicely" spread. We shall make use of the following notations. Given the $k$-wise independent set $S$ and an index $i \in [8t]$, consider coordinates $J_{h,i}$ of every element in $S$. Denote this multiset with $S_{h,i}$. Clearly $S_{h,i}$ is $k$-wise independent.

*Case* 1. Assume that the size of the support of $v$ is at most $t$. Clearly, for every $x \in \mathcal{B}_n$ we have that $\langle x, v\rangle \leq \|v\|_1$. We now show that there is some $x^{h,w} \in N_\varepsilon$ with $\langle x^{h,w}, v\rangle = \|v\|_1$. This clearly implies that $L_{v,\theta}(x^{h,w}) = 1$. Indeed, Lemma A.1 guarantees that there is some $h \in \mathcal{H}$ that is injective on $I_t$. Namely, it maps all the nonzero coordinates of $v$ to different buckets. As a bucket now contains at most one nonzero element, we see that for each $i \in [8t]$ we have that

$$\Pr_{s\in S_{h,i}}\left[\left\langle s, v_{J_{h,i}}\right\rangle = \|v_{I_t\cap J_{h,i}}\|_1\right] \geq \frac{1}{2}, \tag{4.1}$$

where we used the fact that each bucket contains at most one nonzero element so we need only one entry of $s$ to have the correct sign. For every $i \in [8t]$ denote with $A_i \subseteq S_{h,i}$ the set of $s \in S_{h,i}$ that belongs to the "good" sets defined in (4.1), namely, those elements from $S_{h,i}$ that have a large inner product with $v_{J_{h,i}}$. Clearly, for every $i$ we have that $|A_i|/|S_{h,i}| \geq \frac{1}{2}$. We will now show that there exists a walk on $G$ such that for every $i$, $w_i \in A_i$. Indeed, $G$ is an $[m, d, \lambda]$-expander and so Theorem 2.6 guarantees that if $\frac{1}{2} > 6\lambda/d$, then there exists a walk that hits all the $A_i$'s. As we picked a graph $G$ with $\lambda \leq d/1000$ we have the required property. Thus, there exists a walk $w = (w_1, \ldots, w_{8t})$ such that for every $i$, $w_i \in A_i$. Calculating, we get that

$$\left\langle x^{h,w}, v\right\rangle = \sum_{i=1}^{8t}\left\langle w_i, v_{J_{h,i}}\right\rangle = \sum_{i\in h(I_t)}\left\langle w_i, v_{J_{h,i}}\right\rangle = \sum_{i\in I_t}|v_i| = \|v\|_1,$$

as required. This completes the analysis of the first case.

*Case* 2. Assume that $\sum_{r=\lceil 2t/3\rceil}^{t-1}|v_{i_{r+1}}| \geq \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2$ (this is the first bullet of Corollary 2.9).[6] Similarly to the first case (and, using Lemma 2.5) we get that there is $x^{h,w} \in N_\varepsilon$ such that

$$\left\langle x^{h,w}, v\right\rangle \geq \|v_{I_t}\|_1 \geq \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2. \tag{4.2}$$

[6]Clearly this case subsumes Case 1, but we decided to give both of them as the first case is easier to handle and gives an intuition for the second case.

Indeed, as in the first case we consider the different buckets into which coordinates from $I_t$ fell. Lemma 2.5 guarantees that for each bucket a $1/40$ fraction of all elements in $S$ gives the required inner product. Theorem 2.6 guarantees the existence of such a good $x^{h,w}$. We now show that $\theta$ is smaller than the right-hand side of (4.2) and hence our chosen $x^{h,w}$ satisfies that $L_{v,\theta}(x^{h,w}) = 1$ as required. By Corollary 2.11 we get that

$$\Pr_{x\in\mathcal{B}_n}\left[\langle x, v\rangle \geq \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2\right]$$
$$\leq \exp(-t/18) + \exp\left(-\frac{1}{2}\left(\frac{\sqrt{t}}{32}\right)^2\right) = \exp(-\gamma t)$$

for some absolute constant $\gamma > 0$. If we pick $c$ large enough (i.e., $c \geq 1/\gamma$), then for $t = \lceil c\log(2/\varepsilon)\rceil$ we get that

$$\Pr_{x\in\mathcal{B}_n}\left[\langle x, v\rangle \geq \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2\right] \leq \exp(-\gamma t) < \varepsilon.$$

As we assumed that $\Pr_{x\in\mathcal{B}_n}[L_{v,\theta}(x) = 1] \geq \varepsilon$, we have that

$$\theta < \|v_{I_{\lceil 2t/3\rceil}}\|_1 + \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2. \tag{4.3}$$

This completes the analysis of the second case.

*Case* 3. We now assume that $\sum_{r=\lceil 2t/3\rceil}^{t-1}|v_{i_{r+1}}| < \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2$. Hence, Corollary 2.9 implies that there exist $\lceil 2t/3\rceil \leq q \leq t-1$ and some $h \in \mathcal{H}$ such that $h$ is an injection on $I_t$, and for at least $c_1 \cdot 8t$ buckets $r \in [8t]$ it holds that $\|v_{h^{-1}(r)\setminus I_q}\|_2^2 \geq \frac{c_2}{t}\cdot\|v_{[n]\setminus I_q}\|_2^2$ for two universal constants $c_1$ and $c_2$. Denote the set of $\geq c_1\cdot 8t$ "good" buckets $r$ with $R \subset [8t]$. We also define, for every $i \in [8t]$, $J'_{h,i} = h^{-1}(i)\setminus I_q$. It follows that for every $i \in R$

$$\|v_{J'_{h,i}}\|_2 \geq \sqrt{\frac{c_2}{t}}\cdot\|v_{[n]\setminus I_q}\|_2.$$

By Lemma 2.5, specialized to $k = 5$, we get that for every $i \in h(I_q)$

$$\Pr_{s\in S_{h,i}}\left[\left\langle s, v_{J_{h,i}}\right\rangle \geq \|v_{I_q}\cap J_{h,i}\|_1 + \frac{1}{7}\|v_{J'_{h,i}}\|_2\right] \geq \frac{4}{5}\cdot 2^{-5} = \frac{1}{40}, \tag{4.4}$$

where we recall that by our assumption on $h$ we have that $|I_q \cap J_{h,i}| = 1$. In addition, Lemma 2.4 implies that for $i \notin h(I_q)$ (this actually holds for every $i$)

$$\Pr_{s\in S_{h,i}}\left[\left\langle s, v_{J_{h,i}}\right\rangle \geq \frac{\|v_{J_{h,i}}\|_2}{7}\right] \geq \frac{1}{20}. \tag{4.5}$$

For every $i \in [8t]$ denote with $A_i \subseteq S_{h,i}$ the set of $s \in S_{h,i}$ that belong to the "good" sets defined in (4.4), (4.5), namely, those elements from $S_{h,i}$ that have large inner product with $v_{J_{h,i}}$. Clearly, for every $i$ we have that $|A_i|/|S_{h,i}| \geq \min(\frac{1}{40}, \frac{1}{20}) = \frac{1}{40}$. We will now show that there exists a walk on $G$ such that for every $i$, $w_i \in A_i$. Indeed, $G$ is an $[m, d, \lambda]$-expander and so Theorem 2.6 guarantees that if $\frac{1}{40} > 6\lambda/d$, then there exists a walk that hits all the $A_i$'s. As we picked a graph $G$ with

$\lambda \leq d/1000$ we have the required property. Thus, there exists a walk $w = (w_1, \ldots, w_{8t})$ such that for every $i$, $w_i \in A_i$. Calculating, we get that

$$\begin{aligned}
\left\langle x^{h,w}, v \right\rangle &= \sum_{i=1}^{8t} \langle w_i, v_{J_{h,i}} \rangle \\
&= \sum_{i \in h(I_q)} \langle w_i, v_{J_{h,i}} \rangle + \sum_{i \notin h(I_q)} \langle w_i, v_{J_{h,i}} \rangle \\
&\geq \sum_{i \in h(I_q)} \left( \|v_{I_q} \cap J_{h,i}\|_1 + \frac{1}{7} \|v_{J'_{h,i}}\|_2 \right) + \sum_{i \notin h(I_q)} \frac{\|v_{J_{h,i}}\|_2}{7} \\
&= \|v_{I_q}\|_1 + \frac{1}{7} \sum_{i \in [8t]} \|v_{J'_{h,i}}\|_2 \geq \|v_{I_q}\|_1 + \frac{1}{7} \sum_{i \in R} \|v_{J'_{h,i}}\|_2 \\
&\geq \|v_{I_q}\|_1 + \frac{1}{7} \sum_{i \in R} \sqrt{\frac{c_2}{t}} \cdot \|v_{[n] \setminus I_q}\|_2 \\
&\geq^{\ddagger} \|v_{I_q}\|_1 + \frac{8c_1\sqrt{c_2}}{7} \cdot \sqrt{t} \cdot \|v_{[n] \setminus I_q}\|_2 \\
&\geq \|v_{I_q}\|_1 + \frac{8c_1\sqrt{c \cdot c_2}}{7} \cdot \sqrt{\log(2/\varepsilon)} \cdot \|v_{[n] \setminus I_q}\|_2 \\
&\geq^{*} \|v_{I_q}\|_1 + \sqrt{2\log(2/\varepsilon)} \cdot \|v_{[n] \setminus I_q}\|_2 >^{\dagger} \theta,
\end{aligned}$$

where inequality $(\ddagger)$ follows from the fact that $|R| \geq c_1 \cdot 8t$, inequality $(*)$ holds for a large enough universal constant $c$, and inequality $(\dagger)$ holds from the same argument as in case 2 (for $c$ large enough), recalling that $q \geq \lceil 2t/3 \rceil = \lceil \frac{2}{3} c \log 2/\varepsilon \rceil$. Thus, $L_{v,\theta}(x^{h,w}) = 1$ as required. This concludes the proof of Theorem 1.1. □

# 5 Construction of $\varepsilon$-nets for spherical caps

In this section we show how to construct $\varepsilon$-nets for spherical caps. In particular we prove Theorem 1.2.

THEOREM 1.2. *There exist two universal constants $a, b > 0$ such that for every $\varepsilon = \exp(-O(\sqrt{n}))$ there is an explicit construction of an $\varepsilon$-net, $S_\varepsilon \subset \mathbb{S}^{n-1}$, for spherical caps of size $|S_\varepsilon| = O(\varepsilon^{-b} \cdot n^a)$.*

A first natural attempt is to check whether the $\varepsilon$-net for threshold functions is also an $\varepsilon$-net for spherical caps. As we are looking for subsets of the sphere $\mathbb{S}^{n-1}$, we consider the natural embedding of $\mathcal{B}_n$ in $\mathbb{S}^{n-1}$ that shrinks every vector by a factor of $\sqrt{n}$, i.e., we set $\overline{\mathcal{B}_n} = \{-1/\sqrt{n}, 1/\sqrt{n}\}^n$. In this section whenever we discuss the Boolean cube we will refer to the set $\overline{\mathcal{B}_n}$. In particular we will view every subset of $\mathcal{B}_n$ as a subset of $\overline{\mathcal{B}_n}$. To see that the Boolean cube (as a subset of $\mathbb{S}^{n-1}$) is not an $\varepsilon$-net for a polynomially small $\varepsilon$ consider the cap defined by $v = (1, 0, \ldots, 0)$ and $\theta = \sqrt{\log(1/\varepsilon)/n}$. We see that $L_{v,\theta}(\mathcal{B}_n) = 0$ whereas the cap $C_{v,\theta} = L_{v,\theta}^{-1}(1) \cap \mathbb{S}^{n-1}$ has measure $\mathrm{poly}(\varepsilon)$. However, it turns out that if an $\varepsilon$-net for LTFs does not hit a large enough cap, then a "rotation" of it does hit the cap. Therefore, the union of an $\varepsilon'$-net for LTFs and its rotation yields an $\varepsilon$-net for spherical caps. Indeed, the reason that $v = (1, 0, \ldots, 0)$ and $\theta = \sqrt{\log(1/\varepsilon)/n}$ show that the Boolean cube is not an $\varepsilon$-net is because all the mass of $v$ is concentrated on a few coordinates (actually only one coordinate). On the other hand, if it was the case that no set of $O(\log(1/\varepsilon))$ coordinates contains more than, say, $3/4$ of the total mass of $v$, then the set $N_\varepsilon$ guaranteed by Theorem 1.1 will hit the cap $C_{v,\sqrt{2\log(1/\varepsilon^{1/16})/n}}$, which by Lemma 2.12 is of weight at most $\varepsilon^{1/16}$.

Indeed, the proof of Theorem 1.1 shows that there is an element $x \in N_\varepsilon$ such that if $M \subset [n]$ is the set of $O(\log 1/\varepsilon)$ largest coordinates of $v$, then[7]

$$\langle x, v\rangle > \sqrt{2\log(1/\varepsilon)/n} \cdot \|v_{[n]\setminus M}\|_2 \geq^{(*)} (1/4) \cdot \sqrt{2\log(1/\varepsilon)/n} = \sqrt{2\log(1/\varepsilon^{1/16})/n},$$

where inequality $(*)$ follows from the fact that at least $1/4$ of the mass of $v$ is supported on the set of coordinates $[n] \setminus M$. Hence, all that we have to do is find a way of spreading out the coordinates of $v$ so that the mass is "nicely" distributed on many coordinates. Our approach to solving this problem is the following: We show that for the Fourier matrix $F$, either $Fv$ has the property that its mass is "well spread" or $v$ itself is well spread. Then we simply let $S_\varepsilon = N_{\varepsilon'} \cup F(N_{\varepsilon'})$ for some $\varepsilon' = \mathsf{poly}(\varepsilon)$, where $N_{\varepsilon'}$ is an $\varepsilon'$-net for LTFs. We now give the formal proof.

*Proof of Theorem 1.2.* As before, we let $i_1, i_2, \ldots, i_n$ be an enumeration of $[n]$ such that $|v_{i_1}| \geq |v_{i_2}| \geq \cdots \geq |v_{i_n}|$, and let $I_r$ denote the set $\{i_1, i_2, \ldots, i_r\}$. Assume that[8] $n = 2^k$ for some integer $k$. Let $F$ be the $n \times n$ Fourier matrix. In other words, each coordinate of $F$ is in $\{-1/\sqrt{n}, 1/\sqrt{n}\}$ and the rows of $F$ are orthogonal. The following lemma shows that $Fv$ or $v$ are "well spread."

**Lemma 5.1.** *For every two subsets $M_1, M_2 \subset [n]$ of size $|M_1|, |M_2| \leq \sqrt{n}/20$ and any unit vector $v \in \mathbb{R}^n$ we have that $\|(Fv)_{M_1}\|_2 \leq 3/4$ or $\|v_{M_2}\|_2 \leq 3/4$.*

*Proof.* The proof follows the following lemma of [Ind07] (specialized for $L = 2$).

**Lemma 5.2** (see Lemma 4.2 of [Ind07])**.** *Let $T$ be a matrix obtained by concatenating rows of two unitary $n \times n$ matrices $H_1$ and $H_2$ with coherence[9] $\delta$. Then, for any set of coordinates $M \subset [2n]$ of size $|M| = s$ and any unit vector $v \in \mathbb{R}^n$ we have that $\|(Tv)_M\|_2^2 \leq \frac{1}{2}(1 + \delta s) \cdot \|Tv\|_2^2$.*

Indeed, let $T$ be the matrix whose first $n$ rows are the identity matrix and the last $n$ rows are $F$. Then, the coherence of $T$ is $\delta = 1/\sqrt{n}$. Given two subsets $M_1, M_2 \subset [n]$ of size $|M_1|, |M_2| \leq \sqrt{n}/20$, let $M_2'$ be the subset of $\{n+1, \ldots, 2n\}$ obtained by adding $n$ to each element of $M_2$. Let $M = M_1 \cup M_2'$. Then for any unit vector $v \in \mathbb{R}^n$ it holds that

$$\|(Fv)_{M_1}\|_2^2 + \|v_{M_2}\|_2^2 = \|(Tv)_M\|_2^2 \leq \frac{1}{2}(1 + \delta|M|) \cdot \|Tv\|_2^2 \leq 1.1/2 \cdot \|Tv\|_2^2 < 1.1.$$

This completes the proof of Lemma 5.1. □

Let $N_{\varepsilon'} \subset \{-1/\sqrt{n}, 1/\sqrt{n}\}^n$ be an $\varepsilon'$-net for LTFs for some $\varepsilon'$ that will be determined later. Define $S_\varepsilon = N_{\varepsilon'} \cup F(N_{\varepsilon'})$. In other words, $S_\varepsilon$ is the union of $N_{\varepsilon'}$ with the rotation of $N_{\varepsilon'}$ by $F$. Note that as $F$ is unitary we have that $S_\varepsilon \subset \mathbb{S}^{n-1}$. We now show that $S_\varepsilon$ is indeed an $\varepsilon$-net for spherical caps. Let $C_{v,\theta}$ be a spherical cap of weight $\varepsilon$. By Lemma 2.12 we see that $\theta \leq \sqrt{2\log(1/\varepsilon)/n}$. Let $u = Tv$, where $T$ is the matrix defined in the proof of Lemma 5.1. As $u = Tv = (v, Fv)$ (the concatenation of $v$ and $Fv$) and $\|v\| = \|Fv\|$ we get by Lemma 5.1 that either in $v$ or in $Fv$, no set of $\sqrt{n}/40$ coordinates contains more than $3/4$ of the total mass (as, if there were two such sets, then their union contradicts the lemma). Assume w.l.o.g. that in $Fv$ no set of $\sqrt{n}/40$ coordinates contains more than $3/4$ of the total mass (the analysis for $v$ is similar). Let $I_t \subset [n]$ be the set of largest[10] $t = \lceil c\log(1/\varepsilon')\rceil \leq \sqrt{n}/40$ coordinates of $Fv$ (note that $c$, $t$, and $I_t$ are chosen as in the

[7] This inequality (or a stronger one) is reached in both Case 2 and Case 3 before using the fact that $x$ belongs to the Boolean cube.

[8] If it is not the case, then we can work with $n' = 2^k$ such that $n < n' < 2n$.

[9] The coherence of $H_1$ and $H_2$ is the largest inner product between a row of $H_1$ and a row of $H_2$.

[10] Recall our assumption that $\varepsilon = \exp(-O(\sqrt{n}))$.

proof of Theorem 1.1). In particular, no coordinate in $I_t$ is the zero coordinate. Following the proof of Theorem 1.1, we note that we are in either Case 2 or Case 3 there and hence, for a large enough $c$, $N_{\varepsilon'}$ contains an element $x \in N_{\varepsilon'}$ such that[11]

$$\begin{aligned}\langle x, Fv\rangle &\geq^{(\dagger)} \frac{1}{\sqrt{n}} \cdot \sqrt{2\log(1/\varepsilon')} \cdot \|(Fv)_{[n]\setminus I_t}\|_2 \\ &\geq^{(*)} \frac{1}{\sqrt{n}} \cdot \sqrt{2\log(1/\varepsilon')} \cdot \frac{1}{4} = \sqrt{2\log(1/\varepsilon'^{1/16})/n},\end{aligned}$$

where inequality $(\dagger)$ is implied either by (4.2) (in Case 2) or by the conclusion of Case 3. Inequality $(*)$ follows from the fact that $\lceil c\log(1/\varepsilon')\rceil < \sqrt{n}/40$ and the assumption that every subset of $\sqrt{n}/40$ coordinates of $Fv$ contains at most $3/4$ of the mass of $Fv$. Hence, $Fx \in F(N_{\varepsilon'}) \subset S_\varepsilon$ and

$$\langle Fx, v\rangle = \langle x, Fv\rangle \geq \sqrt{2\log(1/\varepsilon'^{1/16})/n} = \sqrt{2\log(1/\varepsilon)/n} \geq \theta$$

for $\varepsilon' = \varepsilon^{16}$. This shows that $S_\varepsilon$ is indeed an $\varepsilon$-net for spherical caps. Moreover, we have that

$$|S_\varepsilon| \leq 2|N_{\varepsilon'}| = O(\varepsilon^{-b'} \cdot n^{a'})$$

for absolute constants $a'$ and $b'$. This completes the proof of Theorem 1.2. □

# A Regarding an error in [RS10]

Lemma 2.7 and its proof below fix an error in the original version of this paper [RS10]. The error was pointed out to us by William Hoza. Lemma 2.7 in [RS10] states:

**Lemma A.1** (Lemma 2.7 in [RS10])**.** *For every integer $s$, there is an explicit family $\mathcal{H}$ of hash functions $h : [n] \to [8s]$ of cardinality $|\mathcal{H}| = 2^{(4+o(1))\cdot s + \log 2s \log\log n}$ [12] such that the following holds for every unit vector $v \in \mathbb{S}^{n-1}$. Let $i_1, i_2, \ldots, i_n$ be an enumeration of $[n]$ such that $|v_{i_1}| \geq |v_{i_2}| \geq \cdots \geq |v_{i_n}|$, and let $I_t$ denote the set $\{i_1, i_2, \ldots, i_t\}$. There exists some $h \in \mathcal{H}$ such that the following hold:*

1. *The map $h$ is an injection on $I_s$.*
2. *Let $t \in [s-1]$. If $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then*

$$\sum_{r\in[8s]} \min\left\{\|v_{h^{-1}(r)\setminus I_t}\|_2^2, \frac{2}{s}\cdot\|v_{[n]\setminus I_t}\|_2^2\right\} \geq \frac{1}{2}\cdot\|v_{[n]\setminus I_t}\|_2^2. \tag{A.2}$$

The flawed proof in the appendix of [RS10] constructs $\mathcal{H}$ in two steps. The error is that the first step uses known constructions of pairwise independent hash families mapping $[n]$ to $[s]$. However, the cardinality of such families is $2^{O(\log(sn))}$, which is too large to give the bounds stated in the lemma (in the paper we wrongly claimed that the size is much smaller). If we were to use

[11] The factor of $\frac{1}{\sqrt{n}}$ comes from viewing $\mathcal{B}_n$ as a subset of $\mathbb{S}^{n-1}$. In fact, we can get a much better inner product but we do not try to optimize.

[12] The $\log 2s$ factor can be eliminated at the expense of a slight complication of the construction (adding a preliminary phase that maps $[n]$ to $[s^2]$ and replacing the maps from $[n]$ in the two-phase construction by maps from $[s^2]$). In our application, this does not improve the exponent beyond an $o(1)$ factor, as we use $s = \Theta(\log n)$.

this bound in the construction of the hitting set then we would get a set of size $(n/\varepsilon)^{O(\log\log 1/\varepsilon)}$, which is larger than the claimed size (alternatively, in terms of seed length, we get seed of length $O(\log(n/\varepsilon)\log\log 1/\varepsilon)$ instead of the optimal $O(\log(n/\varepsilon))$. In a footnote we also proposed to add a preliminary step that maps $[n]$ to $[s^2]$ and then to map $[s^2]$ to $[s]$. This, too, is a flawed construction as it does not guarantee that we get from this a family of pairwise independent hash functions. Our fix is to apply a different preliminary step that reduces the domain size, using the construction of lossless condensers of [GUV09]. This leads to a weaker statement, that nonetheless is good enough to derive the main result (Theorem 1.1).

Observe that there are two differences between Lemma 2.7 and Statement A.1. Firstly, there is the term $A \cdot \log 2s \log\log n$, which was $\log 2s \log\log n$ in the original statement (i.e. $A = 1$ was claimed). This slight change does not affect the main result as it only affects the $o(1)$ term in the exponent of $n$. This can be seen by redoing the computation of $N_\varepsilon$ on page 3510 of the original manuscript (as we did here). Secondly, the order of quantifiers on $t$ and $h$ is reversed. This implies the same change in the statement of Corollary 2.8 This does not affect the rest of the paper, because the use of Corollary 2.8 in the proofs of Theorems 1.1 and 1.2 in [RS10] does not require the same hash function $h$ for all $t$, just a hash function $h$ for a specific value of $t$.

# B Perfect hashing

## B.1 Lossless condensers

An important ingredient in the proof is the following construction of Guruswami, Umans, and Vadhan [GUV09] (see also Chapter 6 in [Vad12]) of lossless condensers.

For completeness we first give some basic definitions and then discuss lossless condensers.

**Definition B.1.** *Let $D$ be a distribution on $\{0,1\}^a$. We say that $D$ is a $k$-source if every point in the hypercube has probability at most $2^{-k}$. We use $U_d$ to denote the uniform distribution on $\{0,1\}^d$. We say that two distributions $D_1, D_2$ are $\varepsilon$-close if their statistical distance (half their $L_1$ distance) is at most $\varepsilon$.* ♢

**Definition B.2.** *A function $Con : \{0,1\}^a \times \{0,1\}^d \to \{0,1\}^b$ is a $k \to_\varepsilon k'$ condenser iff for every $k$-source $X$ on $\{0,1\}^a$, there exists a $k'$-source $Z$ on $\{0,1\}^b$ such that $Con(X, U_d)$ is $\varepsilon$-close to $Z$. The function $Con$ is lossless iff $k' = k + d$.* ♢

The main result that we shall need is Theorem 1.7 of [GUV09].[13]

**Theorem B.3** (Theorem 1.7 of [GUV09])**.** *There exists an absolute constant $\beta > 0$ such that: for all positive $k$, all $a \geq k$, where $a$ is an integer, and all $\varepsilon > 0$, there is an explicit $k \to_\varepsilon k + d$ lossless condenser $Con : \{0,1\}^a \times \{0,1\}^d \to \{0,1\}^b$ with $d = \lceil 3(\log a + \log k + \log(1/\varepsilon)) + \beta \rceil$ and $b \leq 2(k + d)$.*

## B.2 Proof of Lemma 2.7

First, note that when $s = 1$ the statement of the claim is trivial and so we only consider the case $s \geq 2$ in the proof.

Our proof consists of three steps. In the first step we use the lossless condenser, of Theorem B.3, to map $[n]$ to $[\log(n) \cdot \mathsf{poly}(s)]$. Then we use the oblivious implementation due to Schmidt and Siegel [SS90] of the Fredman, Komlós, and Szemerédi (FKS) adaptive hashing scheme [FKS84].[14]

[13] We consider the special case $\alpha = 1/2$ of Theorem 1.7 of [GUV09].

[14] For the construction of our hitting set we need the hash family to be fixed and to not depend on the input.

The Schmidt and Siegel implementation consists of two steps. In the first step (second step of our proof), they use a universal family of pairwise-independent hash function to reduce the domain size further to size $O(s)$. The last step repairs the few collisions that may exist.

To ease the readability of the proof we shall assume that $n$ is a power of 2. This has no effect on the claim or the result.

**Step 1** Let $C : \{0,1\}^a \times \{0,1\}^d \to \{0,1\}^b$ be the condenser promised in Theorem B.3 for parameters $a = \log n$, $k = \log(64s)$, $\varepsilon = s^{-100}$, $d = \lceil 3(\log a + \log k + \log(1/\varepsilon)) + \beta \rceil$ and $b \leq 2(k+d)$. (Notice that if $s > n/64$ and hence $k > a$, we can simply skip Step 1.) For each seed $y \in \{0,1\}^d$ denote $C_y(x) = C(x,y)$. We think of the family $C_y$ as a family of hash functions from $[n]$ to $[2^b]$.

We next show that for a random seed $y \in \{0,1\}^d$, with high probability, $C_y$ is one to one on the set $I_s$.

**Claim B.4.** *Let $X \subseteq \{0,1\}^a$ be a $k$-source. Then for all but $\sqrt{\varepsilon}$ of the seeds $y \in \{0,1\}^d$ it holds that $C_y(X)$ is $\sqrt{\varepsilon}$-close to a $k$-source.*

*Proof.* The proof is an easy application of Markov's inequality. □

**Corollary B.5.** *Let $I \subseteq [n]$ be a set of size $|I| \leq 2^k$. Then, except with probability $\sqrt{\varepsilon}$ over $y \in \{0,1\}^d$, the map $C_y$ is injective on $I$.*

*Proof.* Let $X$ be a random variable that is uniformly distributed over a set of size exactly $2^k$ that contains $I$. Let $y$ be such that $C_y(X)$ is $\sqrt{\varepsilon}$-close to a $k$-source $Z$. Then,

$$\forall z \in \{0,1\}^b \,, \quad \Pr[C_y(X) = z] \leq \Pr[Z = z] + \sqrt{\varepsilon} \leq 2^{-k} + \sqrt{\varepsilon} < 2 \cdot 2^{-k} \,,$$

where the third inequality follows from the choice of $\varepsilon$. In particular, no two elements of $I$ were mapped to the same element $z$. □

We next show that, with high probability, $C_y$ distributes the weight "nicely".

**Claim B.6.** *For all but $\sqrt{\varepsilon}$ fraction of $y \in \{0,1\}^d$ the following holds. If for $t \in [s-1]$, we have that $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then for every $z \in \{0,1\}^b$,*

$$\|v_{C_y^{-1}(z)\setminus I_t}\|^2 \leq \left(\frac{1}{64s} + \sqrt{\varepsilon}\right) \cdot \|v_{[n]\setminus I_t}\|_2^2 \,.$$

*Proof.* Consider the following distribution on $[n]$:

$$\Pr[X = i_j] = \begin{cases} \frac{v_{i_j}^2}{\|v_{[n]\setminus I_t}\|_2^2} & \text{if } j > t \\ 0 & \textit{otherwise} \end{cases} \,.$$

By our assumption, $X$ is a $\log(64s)$-source ($k$-source). Claim B.4 implies that except for a $\sqrt{\varepsilon}$ fraction of the seeds $y$, $C_y(X)$ is $\sqrt{\varepsilon}$-close to a $k$-source $Z$ (note that $Z$ may depend on $y$). For such

a good $y$ and for $z \in \{0,1\}^b$ we have that

$$\begin{aligned} 2^{-k} + \sqrt{\varepsilon} \geq \Pr[C_y(X) = z] &= \sum_{j \in [n]\setminus[t] : C_y(i_j) = z} \Pr[X = i_j] \\ &= \sum_{j \in [n]\setminus[t] : C_y(i_j) = z} \frac{v_{i_j}^2}{\|v_{[n]\setminus I_t}\|_2^2} \\ &= \frac{\|v_{C_y^{-1}(z)\setminus I_t}\|_2^2}{\|v_{[n]\setminus I_t}\|_2^2} \end{aligned}$$

as claimed. □

**Corollary B.7.** *With probability at least $1 - s\sqrt{\varepsilon}$ a random seed $y$ satisfies that:*

1. *$C_y$ is one-to-one on $I_s$.*
2. *For every $t \in [s-1]$, if $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then for every $z \in \{0,1\}^b$,*

$$\|v_{C_y{}^{-1}(z)\setminus I_t}\|^2 \leq \left(\frac{1}{64s} + \sqrt{\varepsilon}\right) \cdot \|v_{[n]\setminus I_t}\|_2^2 .$$

*Proof.* Follows immediately from applying the union bound to Corollary B.5 and Claim B.6. □

We say that a seed $y$ is "good" if it is one of the $1 - s\sqrt{\varepsilon}$ fraction of seeds in the statement of Corollary B.7.

Fix such a good $y$. We now have that $C_y$ has reduced the domain size to $2^b \leq 2^{2d+2k} = (\log(n) \cdot s)^{O(1)}$.

**Step 2** We now proceed as in the construction of perfect hash families of Schmidt and Siegel [SS90]. We first apply a map $f : [2^b] \to [s]$, taken from a pairwise independent family of hash functions $\mathcal{F}$. There are known explicit constructions of $\mathcal{F}$ with $|\mathcal{F}| = 2^{b + \log s + O(1)}$ (see Theorem 3.26 in [Vad12] and the historical discussion there).

A pairwise independent family of hash functions $\mathcal{F}$ has the following property. If $f$ is chosen uniformly at random from $\mathcal{F}$, then for every $x, y \in [2^b]$, $x \neq y$, it holds that $f(x)$ is distributed uniformly in $[s]$, even when conditioned on $f(y)$. In particular, $\Pr[f(x) = f(y)] = \frac{1}{s}$.

Let $S \subseteq [2^b]$ be an arbitrary set of size $|S| \leq s$. Consider the following event.

$$\sum_{j=1}^{s} |f^{-1}(j) \cap S|^2 < 4s. \tag{B.8}$$

We now show that the probability of this event, when $f$ is chosen uniformly at random from $\mathcal{F}$, is more than $\frac{1}{2}$. Indeed, denoting by $\chi_p$ the indicator of an event $p$, we have that

$$\begin{aligned} \mathsf{EE}\left[\sum_{j=1}^{s} |f^{-1}(j) \cap S|^2\right] &= \mathsf{EE}\left[\sum_{x,x' \in S} \chi_{f(x)=f(x')}\right] = \sum_{x,x' \in S} \mathsf{EE}\left[\chi_{f(x)=f(x')}\right] \\ &= \sum_{x \neq x' \in S} \mathsf{EE}\left[\chi_{f(x)=f(x')}\right] + s = s \cdot (s-1) \cdot (1/s) + s = 2s - 1 . \end{aligned}$$

By applying Markov's inequality we conclude that

$$\Pr\left[\sum_{j=1}^{s} |f^{-1}(j) \cap S|^2 \geq 4s\right] < \frac{1}{2}. \tag{B.9}$$

Thus, the average square of the number of pre-images of a bucket, is of size at most 4.

**Step 3** The second phase of the Fredman, Komlós, and Szemerédi hashing scheme is adaptive, and depends on the hashed set $S$. The idea is the following. If $c_i$ elements of $S$ landed in bucket $i \in [s]$, then by mapping this bucket to $c_i^2$ buckets using a pairwise independent family of hash functions, it is likely that no collision between the elements of $S$ occurs. As the first phase guarantees that $\sum_{i\in[s]} c_i^2 = O(s)$, we end up with a hash table of size $O(s)$. Note that for this construction to work, we need to know the values $\{c_i\}$ which is the reason for the adaptiveness. The Schmidt and Siegel implementation proceeds as follows. It uses a pairwise independent family of hash functions $\mathcal{G}$. Here it will be convenient to assume that $g \in \mathcal{G}$ maps $[2^b]$ to bit vectors. So every $g \in \mathcal{G}$ is a function $g : [2^b] \to \{0,1\}^{2+\log s}$. We can take $|\mathcal{G}| = 2^{\log s + b + O(1)}$. The second phase uses a selection of $s$ (not necessarily distinct) hash functions from $\mathcal{G}$. The hash functions are selected and used as follows. Take a sequence of $\log s$ hash functions $g_1, g_2, \ldots, g_{\log s} \in \mathcal{G}$. Notice that there are at most $|\mathcal{G}|^{\log s} = 2^{\log^2 s + b\log s + O(\log s)}$ such sequences. In addition, take a sequence of $s$ non-negative integers $\bar{c} = (c_1, c_2, \ldots, c_s)$ that satisfy $\sum_{j=1}^{s} c_j = s$ and $\sum_{j=1}^{s} c_j^2 < 4s$. There are at most $2^{2s}$ such sequences (easily bounded by writing the sequence elements in unary notation, separated by zeros). This sequence is our guess of the bucket loads due to $S$ after the first phase. Finally, use an assignment $\rho : [s] \to [\log s]$, that assigns values from $[\log s]$ to elements of $[s]$ in the following way: $1 \in [\log s]$ is assigned to $\frac{s}{2}$ elements of $[s]$, $2 \in [\log s]$ is assigned to $\frac{s}{4}$ elements of $[s]$, and in general $i \in [\log s]$ is assigned to $\frac{s}{2^i}$ elements of $[s]$. Exceptionally, $\log s$ is assigned to 2 elements of $[s]$, in order to cover the entire set. The number of such assignments is at most $2^{s\cdot(1+\sum_{i=1}^{\log s} 2^{-i})} < 2^{2s}$ (write the $s$ assigned values in unary, separated by zeros). The assignment $\rho$ is our guess as to which of the $\log s$ selected hash functions should be used for each bucket.

Each setting of $y$, $f$, $\bar{g}$, $\bar{c}$ and $\rho$ defines a hash function $h \in \mathcal{H}$ as follows. For every $x \in [n]$,

$$h(x) = \left(\sum_{i<f(C_y(x))} 2^{\lceil 2\log c_i\rceil}\right) + \bar{g}_{\rho(f(C_y(x)))}(C_y(x)),$$

where for $i = f(C_y(x))$, $\bar{g}_{\rho(i)}(C_y(x))$ is the first $\lceil 2\log c_i\rceil$ bits of $g_{\rho(i)}(C_y(x))$. We shall also think of $\bar{g}_{\rho(i)}(C_y(x))$ as the binary expansion of an integer number. Notice that

$$|\mathcal{H}| \leq 2^d \cdot |\mathcal{F}| \cdot |\mathcal{G}|^{\log s} \cdot \#\{\bar{c}\} \cdot \#\{\rho\} \leq 2^{4s + O(\log^2 s) + O(\log 2s \log\log n) + O(\log s)}, \tag{B.10}$$

implying the claim in the lemma.[15] Also notice that each $h \in \mathcal{H}$ maps $[n]$ to

$$\sum_{i=1}^{s} 2^{\lceil 2\log c_i\rceil} \leq 2 \cdot \sum_{i=1}^{s} c_i^2 < 8s,$$

as required.

[15] A careful calculation shows that the constant $A$ in the statement of Lemma 2.7 is at most 5.

**Wrapping up** Recall that we still work with a fixed "good" $y$.

**Claim B.11.** *For every vector $v \in \mathbb{S}^{n-1}$, the probability that when we pick $f$ at random there is a choice of $\bar{g}$ and $\rho$ such that $h = h_{y,f,\bar{g},\rho}$ is injective on $I_s$ is at least $\frac{1}{2}$.*

*Proof.* Let $S = I_s$. Since $y$ is good, $C_y$ is injective on $S$. Denote $S_y = C_y(S)$. For this set $S_y$, Equation (B.8) holds for at least half of the choices of $f$ (by Equation (B.9)). Fix any such choice $f$. For $i = 1, 2, \ldots, s$, let $C_i = \{x \in S_y : f(x) = i\}$. Consider the choice of $c_i = |C_i|$, for $i = 1, 2, \ldots, s$. Fix $i$. For every $g \in \mathcal{G}$ and $x \in [2^b]$, let $\bar{g}(x)$ denote the first $\lceil 2 \log c_i \rceil$ bits of $g(x)$. Consider the "bad" event

$$A_i = A_i(g) = \exists x, x' \in C_i,\ x \neq x' :\ \bar{g}(x) = \bar{g}(x') \ .$$

As $\mathcal{G}$ is a pairwise independent family of hash functions, if $g$ is chosen uniformly at random in $\mathcal{G}$, then $\Pr[A_i] \leq \binom{c_i}{2} \cdot \frac{1}{c_i^2} < \frac{1}{2}$. Therefore, there exists a choice of $g_1$ that is good for a set $J_1 \subset [s]$ of buckets of cardinality $|J_1| = \frac{s}{2}$. Similarly, for $j = 2, 3, \ldots, \log s - 1$, there exists a choice of $g_j$ that is good for a set $J_j \subset [s] \setminus \bigcup_{j'<j} J_{j'}$ of cardinality $|J_j| = \frac{s}{2^j}$. Similarly, there exists a choice of $g_{\log s}$ that is good for both elements in $[s] \setminus \bigcup_{j<\log s} J_j$. So, for every $f$ that satisfies Equation (B.8), there is a choice of $g$, $c$, and $\rho$ such that the resulting hash function $h$ is an injection on $I_s$. □

**Claim B.12.** *For every $t \in [s-1]$, if $t$ satisfies that $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then with probability at least $\frac{2}{3}$, $f$ satisfies that*

$$\sum_{r \in [s]} \min\left\{ \|v_{(f \circ C_y)^{-1}(r) \setminus I_t}\|_2^2, \frac{2}{s} \cdot \|v_{[n]\setminus I_t}\|_2^2 \right\} \geq \frac{1}{2} \cdot \|v_{[n]\setminus I_t}\|_2^2. \tag{B.13}$$

Observe that Equation (B.13) implies Equation (2.8), as the $g_i$-s only further split hash buckets.

The intuition behind this claim is simple: If $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then no $i \in [n] \setminus I_t$ has $v_i^2$ very large relative to $\|v_{[n]\setminus I_t}\|_2^2$, so $\|v_{[n]\setminus I_t}\|_2^2$ is spread roughly evenly on many coordinates. As $f \circ C_y$ is likely to map the coordinates of $v_{[n]\setminus I_t}$ roughly evenly, it also maps the weight $\|v_{[n]\setminus I_t}\|_2^2$ roughly evenly.

*Proof.* To ease the reading, let us use the following notation. Let $u \in \mathbb{R}^{2^b}$ be defined as follows. For $z \in \{0,1\}^b$,

$$u_z = \|v_{C_y^{-1}(z) \setminus I_t}\|_2^2 \ .$$

Let us also denote

$$W = \|u\|_1 = \|v_{[n]\setminus I_t}\|_2^2 \ .$$

With these notations, what we wish to prove is that with probability at least $\frac{2}{3}$, over the choice of $f$,

$$\sum_{r \in [s]} \min\left\{ \|u_{f^{-1}(r)}\|_1, \frac{2}{s} W \right\} \geq \frac{1}{2} W \ . \tag{B.14}$$

Since $y$ is good, Corollary B.7 guarantees that for all $z$,

$$u_z \leq \left( \frac{1}{64s} + \sqrt{\varepsilon} \right) \cdot \|v_{[n]\setminus I_t}\|_2^2 = \left( \frac{1}{64s} + \sqrt{\varepsilon} \right) \cdot W < \frac{W}{7.99^2 s} \ .$$

Let $X_z^i$ be the indicator random variable for the event that $f(z) = i$. As $\Pr[f(z) = i] = \frac{1}{s}$, we have that

$$\mathsf{EE}\left[\|u_{f^{-1}(i)}\|_1\right] = \sum_{z=1}^{2^b} \mathsf{EE}[X_z^i] \cdot u_z = \frac{1}{s}\|u\|_1 = \frac{1}{s}W\,.$$

Moreover, as $f$ comes from a pairwise independent family of hash functions, for fixed $i$ the random variables $X_z^i$ are pairwise independent, so

$$\begin{aligned}\sigma^2\left[\|u_{f^{-1}(i)}\|_1\right] &= \mathrm{Var}\left[\|u_{f^{-1}(i)}\|_1\right] = \mathrm{Var}\left[\sum_{z=1}^{2^b} X_z^i \cdot u_z\right] \\ &= \sum_{z=1}^{2^b} \mathrm{Var}[X_z^i] \cdot u_z^2 = \left(1 - \frac{1}{s}\right) \cdot \frac{1}{s} \cdot \|u\|_2^2\,.\end{aligned}$$

Thus, as $u_z \leq \frac{W}{7.99^2 s}$, we get that

$$\sigma\left[\|u_{f^{-1}(i)}\|_1\right] \leq \frac{1}{\sqrt{s}} \cdot \|u\|_2 \leq \frac{1}{\sqrt{s}} \cdot \sqrt{\frac{W}{7.99^2 s}} \cdot \sqrt{\|u\|_1} = \frac{W}{7.99s}\,.$$

By Chebyshev's inequality, for every $r > 1$,

$$\Pr\left[\|u_{f^{-1}(i)}\|_1 \geq \frac{r}{s}W\right] \leq \frac{1}{7.99^2(r-1)^2}\,. \tag{B.15}$$

For each value of $r$, consider all values $\lambda$ in the interval $[2^r, 2^{r+1}]$ such that $\Pr\left[\|u_{f^{-1}(i)}\|_1 = \frac{\lambda}{s}W\right] \neq 0$. Clearly there are finitely many such values. From Equation (B.15) we get that

$$\begin{aligned}\sum_{\lambda \in [2^r, 2^{r+1}]} \lambda \cdot \Pr\left[\|u_{f^{-1}(i)}\|_1 = \frac{\lambda}{s}W\right] &\leq 2^{r+1} \Pr\left[\|u_{f^{-1}(i)}\|_1 \geq \frac{2^r}{s} \cdot W\right] \\ &\leq \frac{2^{r+1}}{7.99^2(2^r-1)^2}\,.\end{aligned}$$

Thus,

$$\begin{aligned}\mathsf{EE}\left[\max\left\{0, \|u_{f^{-1}(i)}\|_1 - \frac{2}{s}W\right\}\right] &\leq \frac{W}{7.99^2 s} \cdot \sum_{r=1}^{\infty} \frac{2^{r+1}}{(2^r-1)^2} \\ &= \frac{4W}{7.99^2 s} \cdot \sum_{r=1}^{\infty} \frac{2^{r-1}}{(2^r-1)^2} \\ &< \frac{4W}{7.99^2 s} \cdot \sum_{r=1}^{\infty} \frac{1}{2^{r-1}} \\ &= \frac{8W}{7.99^2 s}\,.\end{aligned}$$

Let $Y^i = \max\left\{0, \|u_{f^{-1}(i)}\|_1 - \frac{2}{s}W\right\}$. We just showed that $\mathsf{EE}\left[\sum_{i\in[s]} Y^i\right] < \frac{8W}{7.99^2}$, so by Markov's Inequality, $\Pr\left[\sum_{i\in[s]} Y^i > \frac{1}{2}W\right] < \frac{1}{3}$.

We next show that when $\sum_{i\in[s]} Y^i \leq \frac{1}{2}W$, Equation (B.14) holds, and thus it holds with probability at least $\frac{2}{3}$ over the choice of $f \in \mathcal{F}$, which is what we wanted to prove.

Let $m$ be the number of $i \in [s]$ such that $Y^i > 0$. We now get that

$$\frac{1}{2}W \geq \sum_{i\in[s]} Y^i = \sum_{i:Y^i>0} \left( \|u_{f^{-1}(i)}\|_1 - \frac{2}{s}W \right) = \left( \sum_{i:Y^i>0} \|u_{f^{-1}(i)}\|_1 \right) - \frac{2m}{s}W \,.$$

Hence,

$$\begin{aligned} \sum_{i=1}^{s} \min\left\{ \|u_{f^{-1}(i)}\|_1, \frac{2}{s}W \right\} &= \left( \sum_{i:Y^i=0} \|u_{f^{-1}(i)}\|_1 \right) + \frac{2m}{s}W \\ &= \left( W - \sum_{i:Y^i>0} \|u_{f^{-1}(i)}\|_1 \right) + \frac{2m}{s}W \\ &\geq W - \frac{1}{2}W = \frac{1}{2}W \,, \end{aligned}$$

and Equation (B.14) holds. □

To conclude the proof of Lemma 2.7 we recall that $y$ is good with probability at least $1 - s\sqrt{\varepsilon} > 0$ and that for each good $y$, Claim B.11 holds for a random choice of $f$ with probability at least $\frac{1}{2}$. Furthermore, for a good $y$, we have that Equation (B.14) holds for at least $\frac{2}{3}$ of the choices of $f \in \mathcal{F}$. As $\frac{2}{3} + \frac{1}{2} > 1$, we get that for each good $y$, there is a good choice of $f$, so that both Equation (B.14) and the condition in the statement of Claim B.11 hold. This is exactly what Lemma 2.7 claims.


### Acknowledgments

We thank Noga Alon and Avi Wigderson for helpful discussions and for bringing [SS90] to our attention. We also thank Noga for sharing his proof of Corollary 1.3 with us.

We are grateful to William Hoza who found a mistake in our original argument. We also wish to thank William Hoza and the anonymous reviewers for pointing out several other inaccuracies and for comments that helped us significantly improve the presentation of the proof.


## References


[AFWZ95] N. ALON, U. FEIGE, A. WIGDERSON, AND D. ZUCKERMAN, *Derandomized graph products*, Comput. Complexity, 5 (1995), pp. 60–75.

[AGK04] N. ALON, G. GUTIN, AND M. KRIVELEVICH, *Algorithms with large domination ratio*, J. Algorithms, 50 (2004), pp. 118–131.

[AKN+08] N. ALON, H. KAPLAN, G. NIVASCH, M. SHARIR, AND S. SMORODINSKY, *Weak ε-nets and interval chains*, in Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), San Francisco, 2008, pp. 1194–1203.

[Alo08] N. ALON, *private communication*, 2008.

[AS08] N. ALON AND J. SPENCER, *The Probabilistic Method*, 3rd ed., Wiley, 2008.

[Ber97] B. BERGER, *The fourth moment method*, SIAM J. Comput., 26 (1997), pp. 1188–1207.

[Cha94] B. CHAZELLE, *Computational geometry: A retrospective*, in Proceedings of the 26th Annual ACM Symposium on Theory of Computing (STOC), Montreal, 1994, pp. 75–94.

[CM96] B. CHAZELLE AND J. MATOUSEK, *On linear-time deterministic algorithms for optimization problems in fixed dimension*, J. Algorithms, 21 (1996), pp. 579–597.

[CT06] E. J. CANDÉS AND T. TAO, *Near-optimal signal recovery from random projections: Universal encoding strategies*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.

[CW79] J. L. CARTER AND M. N. WEGMAN, *Universal classes of hash functions*, J. Comput. System Sci., 18 (1979), pp. 143–154.

[DGJ+09] I. DIAKONIKOLAS, P. GOPALAN, R. JAISWAL, R. A. SERVEDIO, AND E. VIOLA, *Bounded independence fools halfspaces*, SIAM J. Comput., to appear.

[Don06] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[DP09] D. P. DUBHASHI AND A. PANCONESI, *Concentration of Measure for the Analysis of Randomized Algorithms*, Cambridge University Press, Cambridge, UK, 2009.

[EGL+92] G. EVEN, O. GOLDREICH, M. LUBY, N. NISAN, AND B. VELICKOVIC, *Approximations of general independent distributions*, in Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC), Victoria, BC, 1992, pp. 10–16.

[EIO02] L. ENGEBRETSEN, P. INDYK, AND R. O'DONNELL, *Derandomized dimensionality reduction with applications*, in Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), San Francisco, 2002, pp. 705–712.

[FKS84] M. L. FREDMAN, J. KOMLÓS, AND E. SZEMERÉDI, *Storing a sparse table with $O(1)$ worst case access time*, J. ACM, 31 (1984), pp. 538–544.

[GLR08] V. GURUSWAMI, J. R. LEE, AND A. A. RAZBOROV, *Almost Euclidean subspaces of $l_1^n$ via expander codes*, in Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), San Francisco, 2008, pp. 353–362.

[GLW08] V. GURUSWAMI, J. R. LEE, AND A. WIGDERSON, *Euclidean sections of $\ell_1^N$ with sublinear randomness and error-correction over the reals*, in Approximation, Randomization and Combinatorial Optimization: Algorithms and Techniques, Lecture Notes in Comput. Sci. 5171, Springer, Berlin, Heidelberg, 2008, pp. 444–454.

[GUV09] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *J. ACM*, 56(4):1–34, 2009.

[Ind07] P. INDYK, *Uncertainty principles, extractors, and explicit embeddings of $\ell_2$ into $\ell_1$*, in Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC), San Diego, 2007, pp. 615–620.

[JL84] W. B. JOHNSON AND J. LINDENSTRAUSS, *Extensions of Lipschitz maps into a Hilbert space*, in Contemp. Math. 26, AMS, Providence, 1984, pp. 189–206.

[LLSZ97] N. LINIAL, M. LUBY, M. E. SAKS, AND D. ZUCKERMAN, *Efficient construction of a small hitting set for combinatorial rectangles in high dimension*, Combinatorica, 17 (1997), pp. 215–234.

[LPS88] A. LUBOTZKY, R. PHILLIPS, AND P. SARNAK, *Ramanujan graphs*, Combinatorica, 8 (1988), pp. 261–277.

[Mar88] G. A. MARGULIS, *Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators*, Probl. Inf. Transm., 24 (1988), pp. 39–46.

[Mat02] J. MATOUSEK, *Lectures on Discrete Geometry*, Grad. Texts in Math. 212, Springer, New York, 2002.

[MZ09] R. MEKA AND D. ZUCKERMAN, *Pseudorandom generators for polynomial threshold functions*, in Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC), Cambridge, MA, 2010, pp. 427–436.

[NN93] J. NAOR AND M. NAOR, *Small-bias probability spaces: Efficient constructions and applications*, SIAM J. Comput., 22 (1993), pp. 838–856.

[RS09] Y. RABANI AND A. SHPILKA, *Explicit construction of a small epsilon-net for linear threshold functions*, in Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC), Bethesda, MD, 2009, pp. 649–658.

[RS10] Y. Rabani and A. Shpilka. Explicit construction of a small $\varepsilon$-net for linear threshold functions. *SIAM J. on Computing*, 39(8):3501–3520, 2010.

[Ser06] R. A. SERVEDIO, *Every linear threshold function has a low-weight approximator*, in Proceedings of the 21st Annual IEEE Conference on Computational Complexity (CCC), Prague, 2006, pp. 18–32.

[Siv02] D. SIVAKUMAR, *Algorithmic derandomization via complexity theory*, in Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC), Montreal, 2002, pp. 619–626.

[SS90] J. P. Schmidt and A. Siegel. The Spatial Complexity of Oblivious k-Probe Hash Functions. *SIAM J. on Computing*, 19(5):775–786, 1990.

[Vad12] S. P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

[Vio08] E. VIOLA, *The sum of d small-bias generators fools polynomials of degree d*, in Proceedings of the 23rd Annual IEEE Conference on Computational Complexity (CCC), College Park, MD, 2008, pp. 124–127.