Linear Time Constructions of Some d-Restriction Problems

Nader H. Bshouty(✉)

Technion, Haifa, Israel
bshouty@ca.technion.ac.il

Abstract. We give new linear time globally explicit constructions for perfect hash families, cover-free families and separating hash functions.

Keywords: Derandomization · d-Restriction problems · Perfect hash · Cover-free families · Separating hash functions

1 Introduction

A *d-restriction problem* [7,13,58] is a problem of the following form:
Given an alphabet Σ of size $|\Sigma| = q$, an integer n and a class $\mathcal{M}$ of nonzero functions $f : \Sigma^d \to \{0, 1\}$.
Find a small set $A \subseteq \Sigma^n$ such that: For every $1 \le i_1 < i_2 < \cdots < i_d \le n$ and $f \in \mathcal{M}$ there is $a \in A$ such that $f(a_{i_1}, \ldots, a_{i_d}) \ne 0$.

A $(1-\epsilon)$-*dense d-restriction problem* is a problem of the following form:
Given an alphabet Σ of size $|\Sigma| = q$, an integer n and a class $\mathcal{M}$ of nonzero functions $f : \Sigma^d \to \{0, 1\}$.
Find a small set $A \subseteq \Sigma^n$ such that: For every $1 \le i_1 < i_2 < \cdots < i_d \le n$ and $f \in \mathcal{M}$

$$\mathbf{Pr}_{a \in A}[f(a_{i_1}, \ldots, a_{i_d}) \ne 0] > 1 - \epsilon$$

where the probability is over the choice of a from the uniform distribution on A.

We give new constructions for the following three ($(1-\epsilon)$-dense) d-restriction problems: Perfect hash family, cover-free family and separating hash family.

Perfect hash families were introduced by Mehlhorn [50] in 1984 and used as database management. They were used in compiler design to prove lower bounds on the size of a program that constructs a hash function suitable for fast retrieval of fixed data such as library function names [27]. Perfect hash families have been also applied to circuit complexity problems [59], derandomize some probabilistic algorithms [6], broadcast encryption [39] and threshold cryptography [11,12].

Cover-free families were first introduced in 1964 by Kautz and Singleton [47] to investigate superimposed binary codes. Cover-free families have been used to solve some problems in cryptography and communications, including blacklisting, broadcast encryption, broadcast anti-jamming, source authentication in a network setting, group key predistribution and pooling designs over complexes. See [25,29,33,41,42,46,52,62–64,68,69].

V.Th. Paschos and P. Widmayer (Eds.): CIAC 2015, LNCS 9079, pp. 74–88, 2015.
DOI: 10.1007/978-3-319-18173-8_5

A construction is *global explicit* if it runs in deterministic polynomial time in the size of the construction. A *local explicit construction* is a construction where one can find any bit in the construction in time poly-log in the size of the construction. The constructions in this paper are linear time global explicit constructions.

To the best of our knowledge, our constructions have sizes that are less than the ones known from the literature.

1.1 Learning Hypergraphs

In this section we give one application in computational learning theory.

A hypergraph is $H = (V, E)$ where V is the set of vertices and $E \subseteq 2^V$ is the set of edges. The dimension of the hypergraph H is the cardinality of the largest set in E. For a set $S \subseteq V$, the *edge-detecting queries* $Q_H(S)$ is answered "Yes" or "No", indicating whether S contains all the vertices of at least one edge of H. Learning a hidden hypergraph of constant dimension r with s edges using edge-detecting queries is equivalent to another important problem in learning theory [4]: Learning the class s-term r-MDNF (Monotone DNF with s terms of size r) with *membership queries* (the learner can ask about the value of the function in some point).

This problem has many applications in chemical reactions and genome sequencing. In chemical reactions, we are given a set of chemicals, some of which react and some which do not. When multiple chemicals are combined in one test tube, a reaction is detectable if and only if at least one set of the chemicals in the tube reacts. The goal is to identify which sets react using as few experiments as possible. See [2–5,16,18,24,26,28,34,35,40,53,61] for more details on the problem and many other applications.

This problem is also called "sets of positive subsets" [70] "complex group testing" [53] and "group testing in hypergraph" [35].

It is known that a cover-free families can be used as queries to solve the above problem. Several algorithms with non-optimal query complexity are known from the literature. See [28] and references within. Our construction is the first linear time construction that construct an optimal query set for learning hypergraphs.

2 Old and New Results

2.1 Perfect Hash Family

Let H be a family of functions $h : [n] \to [q]$. For $d \le q$ we say that H is an (n, q, d)-*perfect hash family* ((n, q, d)-PHF) [7] if for every subset $S \subseteq [n]$ of size $|S| = d$ there is a *hash function* $h \in H$ such that $h|_S$ is injective (one-to-one) on S, i.e., $|h(S)| = d$.

Blackburn and Wild [23] gave an optimal explicit construction when $q \ge exp\left(\sqrt{d \log d \log n}\right)$. Stinson et al., [65], gave an explicit construction of (n, q, d)-PHF of size $d^{\log^* n} \log n$ for $q \ge d^2 \log n / \log q$. It follows from the technique used in [1] with Reed-Solomon codes that an explicit (n, q, d)-PHF of size$d^2 \log n / \log q$ exist for $q \ge d^2 \log n / \log q$. In [7,9,58] it was shown that there are $(n, \Omega(d^2), d)$-PHF of size $O(d^6 \log n)$ that can be constructed in $poly(n)$ time. Wang and Xing

[71] used algebraic function fields and gave an (n, d^4, d)-PHF of size $O((d^2/\log d)\log n)$ for infinite sequence of integers n. Their construction is not linear time construction. The above constructions are either for large q or are not linear time constructions.

Bshouty in [13] shows that for a constant $c > 1$, the following (third column in the table) (n, q, d)-PHF can be locally explicitly constructed in almost linear time (within $poly(\log)$)

		Linear time.	Upper	Lower
n	q	Size $= O()$	Bound	Bound
I.S.	$q \geq \frac{c}{4}d^4$	$d^2\frac{\log n}{\log q}$	$d\frac{\log n}{\log q}$	$d\frac{\log n}{\log q}$
all	$q \geq \frac{c}{4}d^4$	$d^4\frac{\log n}{\log q}$	$d\frac{\log n}{\log q}$	$d\frac{\log n}{\log q}$
I.S.	$q \geq \frac{c}{2}d^2$	$d^4\frac{\log n}{\log d}$	$d\frac{\log n}{\log(2q/(d(d-1)))}$	$d\frac{\log n}{\log q}$
all	$q \geq \frac{c}{2}d^2$	$d^6\frac{\log n}{\log d}$	$d\frac{\log n}{\log(2q/(d(d-1)))}$	$d\frac{\log n}{\log q}$
I.S.	$q = \frac{d(d-1)}{2} + 1 + o(d^2)$	$d^6\frac{\log n}{\log d}$	$d\log n$	$d\frac{\log n}{\log q}$
all	$q = \frac{d(d-1)}{2} + 1 + o(d^2)$	$d^8\frac{\log n}{\log d}$	$d\log n$	$d\frac{\log n}{\log q}$

The upper bound in the table follows from union bound [13]. The lower bound is from [10,51] (see also [17,22,23,37,44,45,55]). We note here that all the lower bounds in this paper are true even for non-explicit constructions. I.S. stands for "true for infinite sequence of integers n". Here we prove

Theorem 1. *Let q be a power of prime. If $q > 4(d(d-1)/2+1)$ then there is a (n, q, d)-PHF of size*

$$O\left(\frac{d^2\log n}{\log(q/e(d(d-1)/2+1))}\right)$$

that can be constructed in linear time.

If $d(d-1)/2+2 \leq q \leq 4(d(d-1)/2+1)$ then there is a (n, q, d)-PHF of size

$$O\left(\frac{q^2d^2\log n}{(q-d(d-1)/2-1)^2}\right)$$

that can be constructed in linear time.

In particular, for any constants $c > 1$, $\delta > 0$ and $0 \leq \eta < 1$, the following (n, q, d)-PHF can be constructed in linear time (the third column in the following table)

		Linear time.	*Upper*	*Lower*
n	q	*Size* $= O()$	*Bound*	*Bound*
all	$q \geq d^{2+\delta}$	$d^2\frac{\log n}{\log q}$	$d\frac{\log n}{\log q}$	$d\frac{\log n}{\log q}$
all	$q \geq \frac{c}{2}d^2$	$d^2\log n$	$d\log n$	$d\frac{\log n}{\log q}$
all	$q = \frac{d(d-1)}{2} + 1 + d^{2\eta}$	$d^{6-4\eta}\log n$	$d\log n$	$d\frac{\log n}{\log q}$
all	$q = \frac{d(d-1)}{2} + 2$	$d^6\frac{\log n}{\log d}$	$d\log n$	$d\frac{\log n}{\log q}$

Notice that for $q > cd^2/2$, $c > 1$ the sizes in the above theorem is within a factor of d of the lower bound. Constructing almost optimal (within $poly(d)$) (n, q, d)-PHF for $q = o(d^2)$ is still a challenging open problem. Some nearly optimal constructions of (n, q, d)-PHF for $q = o(d^2)$ are given in [49,58].

The (n, q, d)-perfect hash families for $d \le 6$ are studied in [8,10,20,21,23,49,54,65]. In this paper we prove

Theorem 2. *If q is prime power and $d \le \log n/(8 \log\log n)$ then there is a linear time construction of (n, q, d)-PHF of size $O\left(d^3 \log n/g(q, d)\right)$ where $g(q, d) = (1 - 1/q)(1 - 2/q) \cdots (1 - (d - 1)/q)$.*

Using the lower bound in [37] we show that the size in the above theorem is within a factor of d^4 of the lower bound when $q = d + O(1)$ and within a factor of d^3 for $q > cd$ for some $c > 1$.

2.2 Dense Perfect Hash Family

We say that H is an $(1 - \epsilon)$-*dense* (n, q, d)-PHF if for every subset $S \subseteq [n]$ of size $|S| = d$ there are at least $(1 - \epsilon)|H|$ hash functions $h \in H$ such that $h|_S$ is injective on S.

The following improves the results that can be obtained from [13,14]

Theorem 3. *Let q be a power of prime. If $\epsilon > 4(d(d - 1)/2 + 1)/q$ then there is a $(1 - \epsilon)$-dense (n, q, d)-PHF of size*

$$O\left(\frac{d^2 \log n}{\epsilon \log(\epsilon q/e(d(d - 1)/2 + 1))}\right)$$

that can be constructed in linear time.

If $(d(d-1)/2+1)/(q-1) \le \epsilon \le 4(d(d-1)/2+1)/q$ then there is a $(1-\epsilon)$-dense (n, q, d)-PHF of size

$$O\left(\frac{q^2 d^2 \log n}{\epsilon(q - (d(d - 1)/2 + 1)/\epsilon)^2}\right)$$

that can be constructed in linear time.

We also prove (what we believe) two folklore results that show that the bounds on the size and ϵ in the above theorem are almost tight. First, we show that the size of any $(1 - \epsilon)$-dense (n, q, d)-PHF is $\Omega\left(d \log n/(\epsilon \log q)\right)$. Second, we show that no $(1 - \epsilon)$-dense (n, q, d)-PHF exists when $\epsilon < d(d - 1)/(2q) + O((d^2/q)^2)$. Notice that for $q \ge (d/\epsilon)^{1+c}$, where $c > 1$ is any constant, the size of the construction in Theorem 3, $O\left(d^2 \log n/(\epsilon \log q)\right)$, is within a factor d of the lower bound. Also the bound on ϵ is asymptotically tight.

For the rest of this section we will only state the results for the non-dense d-restriction problems. Results similar to Theorem 3 can be easily obtained using the same technique.

2.3 Cover-Free Families

Let X be a set with N elements and let $\mathcal{B}$ be a set of subsets (blocks) of X. We say that $(X,\mathcal{B})$ is (w,r)-*cover-free family* ((w,r)-CFF), [47], if for any w blocks $B_1,\ldots,B_w \in \mathcal{B}$ and any other r blocks $A_1,\ldots,A_r \in \mathcal{B}$, we have

$$\bigcap_{i=1}^{w} B_i \not\subseteq \bigcup_{j=1}^{r} A_j.$$

Let $N((w,r),n)$ denotes the minimum number of points in any (w,r)-CFF having n blocks. Here we will study CFF when $w = o(r)$ (or $r = o(w)$). We will write $(n,(w,r))$-CFF when we want to emphasize the number of blocks.

When $w = 1$, the problem is called *group testing*. The problem of group testing which was first presented during World War II was presented as follows [30,56]: Among n soldiers, at most r carry a fatal virus. We would like to blood test the soldiers to detect the infected ones. Testing each one separately will give n tests. To minimize the number of tests we can mix the blood of several soldiers and test the mixture. If the test comes negative then none of the tested soldiers are infected. If the test comes out positive, we know that at least one of them is infected. The problem is to come up with a small number of tests.

This problem is equivalent to $(n,(1,r))$-CFF and is equivalent to finding a small set $\mathcal{F} \subseteq \{0,1\}^n$ such that for every $1 \le i_1 < i_2 < \cdots < i_d \le n$ and every $1 \le j \le d$ there is $a \in \mathcal{F}$ such that $a_{i_k} = 0$ for all $k \ne j$ and $a_{i_j} = 1$.

Group testing has the following lower bound [31,32,36]

$$N((1,r),n) \ge \Omega\left(\frac{r^2}{\log r}\log n\right). \qquad (1)$$

It is known that a group testing of size $O(r^2 \log n)$ can be constructed in linear time [30,43,60].

An $(n,(w,r))$-CFF can be regarded as a set $\mathcal{F} \subseteq \{0,1\}^n$ such that for every $1 \le i_1 < i_2 < \cdots < i_d \le n$ where $d = w + r$ and every $J \subset [d]$ of size $|J| = w$ there is $a \in \mathcal{F}$ such that $a_{i_k} = 0$ for all $k \not\in J$ and $a_{i_j} = 1$ for all $j \in J$. Then $N((w,r),n)$ is the minimum size of such $\mathcal{F}$.

It is known that, [67],

$$N((w,r),n) \ge \Omega\left(\frac{d\binom{d}{w}}{\log\binom{d}{w}}\log n\right).$$

Using union bound it is easy to show

Lemma 1. *For $d = w+r = o(n)$ we have $N((w,r),n) \le O\left(\sqrt{wrd}\cdot\binom{d}{w}\log n\right)$.*

It follows from [65], that for infinite sequence of integers n, an $(n,(w,r))$-CFF of size $M = O\left((wr)^{\log^* n}\log n\right)$ can be constructed in polynomial time. For constant d, the (n,d)-universal set over $\Sigma = \{0,1\}$ constructed in [57] of size $M = O(2^{3d}\log n)$ (and in [58] of size $M = 2^{d+O(\log^2 d)}\log n$) is $(n,(w,r))$-CFF for any w and r of size $O(\log n)$. See also [48]. In [13], Bshouty gave the following locally explicit constructions of $(n,(w,r))$-CFF that can be constructed in (almost) linear time in their sizes (the third column in the table).

n	w	Linear time Size=	Upper Bound	Lower Bound
I.S	$O(1)$	$\frac{r^{w+2}}{\log r}\log n$	$r^{w+1}\log n$	$\frac{r^{w+1}}{\log r}\log n$
all	$O(1)$	$\frac{r^{w+3}}{\log r}\log n$	$r^{w+1}\log n$	$\frac{r^{w+1}}{\log r}\log n$
I.S.	$o(r)$	$\frac{w^2(ce)^w r^{w+2}}{\log r}\log n$	$\frac{r^{w+1}}{(w/e)^{w-1/2}}\log n$	$\frac{r^{w+1}}{(w/e)^{w+1}\log r}\log n$
all	$o(r)$	$\frac{w^3(ce)^w r^{w+3}}{\log r}\log n$	$\frac{r^{w+1}}{(w/e)^{w-1/2}}\log n$	$\frac{r^{w+1}}{(w/e)^{w+1}\log r}\log n$

In the table, $c > 1$ is any constant. We also added to the table the non-constructive upper bound in the forth column and the lower bound in the fifth column.

In this paper we prove

Theorem 4. *For any constant $c > 1$, the following $(n,(w,r))$-CFF can be constructed in linear time in their sizes*

n	w	*Linear time.* *Size=O( )*	*Upper* *Bound*	*Lower* *Bound*
all	$O(1)$	$r^{w+1}\log n$	$r^{w+1}\log n$	$\frac{r^{w+1}}{\log r}\log n$
all	$o(r)$	$(ce)^w r^{w+1}\log n$	$\frac{r^{w+1}}{(w/e)^{w-1/2}}\log n$	$\frac{r^{w+1}}{(w/e)^{w+1}\log r}\log n$

Notice that when $w = O(1)$ the size of the construction matches the upper bound obtained with union bound and is within a factor of $\log r$ of the lower bound.

See the results for Separating Hash Family in the full paper [15].

3 Preliminary Constructions

A *linear code* over the field $\mathcal{F}_q$ is a linear subspace $C \subset \mathcal{F}_q^m$. Elements in the code are called *words*. A linear code C is called $[m,k,d]_q$ *linear code* if $C \subset \mathcal{F}_q^m$ is a linear code, $|C| = q^k$ and for every two words v and u in the code $\mathrm{dist}(v,u) := |\{i \mid v_i \neq u_i\}| \geq d$. The q-ary entropy function is

$$H_q(p) = p\log_q \frac{q-1}{p} + (1-p)\log_q \frac{1}{1-p}.$$

The following is from [60] (Theorem 2)

Lemma 2. *Let q be a prime power, m and k positive integers and $0 \leq \delta \leq 1$. If $k \leq (1 - H_q(\delta))m$, then an $[m,k,\delta m]_q$ linear code can be globally explicit constructed in time $O(mq^k)$.*

All the results in this paper uses Lemma 2 and therefore they are globally explicit constructions. In the full paper, [15], we prove the following

Lemma 3. *Let q be a prime power, $1 < h < q/4$ and*

$$m = \left\lceil \frac{h \ln(q(n+1))}{\ln q - \ln h - 1} \right\rceil .$$

A

$$\left[m, \left\lceil \frac{\log(n+1)}{\log q} \right\rceil, \left(1 - \frac{1}{h}\right) m \right]_q$$

linear code can be constructed in time $O(hqn \log(qn))$.

When $h = \Theta(q)$ we show

Lemma 4. *Let q be a prime power, $2 \le q/4 \le h \le q-1$ and*

$$m = \left\lceil \frac{4(q-1)^2 h \ln(q(n+1))}{(q-h)^2} \right\rceil .$$

A

$$\left[m, \left\lceil \frac{\log(n+1)}{\log q} \right\rceil, \left(1 - \frac{1}{h}\right) m \right]_q$$

linear code can be constructed in time $O(h(q^2/(q-h)^2) n \log(qn))$.

4 Main Results

In this section we give two main results that will be used throughout the paper

Let $I \subseteq [n]^2$. Define the following homogeneous polynomial $H_I = \prod_{(i_1,i_2)\in I} (x_{i_1} - x_{i_2})$. We denote by $\mathcal{H}_d \subseteq \mathcal{F}_q[x_1, \ldots, x_n]$ the class of all such polynomials of degree at most d. A *hitting set* for $\mathcal{H}_d$ over $\mathcal{F}_q$ is a set of assignment $A \subseteq \mathcal{F}_q^n$ such that for every $H \in \mathcal{H}_d, H \not\equiv 0$, there is $a \in A$ where $H(a) \neq 0$. A $(1-\epsilon)$-*dense hitting set* for $\mathcal{H}_d$ over $\mathcal{F}_q$ is a set of assignment $A \subseteq \mathcal{F}_q^n$ such that for every $H \in \mathcal{H}_d$, $H \not\equiv 0$,

$$\mathbf{Pr}_{a\in A}[H(a) \neq 0] > 1 - \epsilon$$

where the probability is over the choice of a from the uniform distribution on A. When $H(a) \neq 0$ then we say that the assignment *a hits H* and H is *not zero on a*.

We prove

Lemma 5. *Let $n > q, d$. If $q > 4(d+1)$ is prime power then there is a hitting set for $\mathcal{H}_d$ of size*

$$m = \left\lceil \frac{(d+1)\log(q(n+1))}{\log(q/e(d+1))} \right\rceil = O\left(\frac{d \log n}{\log(q/e(d+1))} \right)$$

that can be constructed in time $O(mn) = O(dqn \log(qn))$.

If $d+2 \le q \le 4(d+1)$ is prime power then there is a hitting set for $\mathcal{H}_d$ of size

$$m = \left\lceil \frac{4(q-1)^2(d+1)\ln(q(n+1))}{(q-d-1)^2} \right\rceil = O\left(\frac{dq^2 \log n}{(q-d-1)^2} \right)$$

that can be constructed in time $O(mn) = O(d(q^2/(q-d-1)^2) n \log(qn))$.

Proof. Consider the code C

$$\left[m, \left\lceil \frac{\log(n+1)}{\log q} \right\rceil, \left(1 - \frac{1}{d+1}\right) m\right]_q$$

constructed in Lemma 3 and Lemma 4. The number of non-zero words in the code is at least n. Take any n distinct non-zero words $c^{(1)}, \cdots, c^{(n)}$ in C and define the assignments $a^{(i)} \in \mathcal{F}_q^n$, $i = 1, \ldots, m$ where $a_j^{(i)} = c_i^{(j)}$. Let $H_I \in \mathcal{H}_d, H_I \not\equiv 0$. Then $H_I = \prod_{(i_1,i_2)\in I}(x_{i_1} - x_{i_2}) \not\equiv 0$ where $|I| \le d$. For each $t := x_{i_1} - x_{i_2}$ we have $(t(a^{(1)}), \ldots, t(a^{(m)}))^T = c^{(i_1)} - c^{(i_2)} \in C$ is a non-zero word in C and therefore t is zero on at most $m/(d+1)$ assignments. Therefore H_I is zero on at most $dm/(d+1) < m$ assignment. This implies that there is an assignment in A that hits H_I. □

Notice that the size of the hitting set is mn and therefore the time complexity in the above lemma is linear in the size of the hitting set.

In the same way one can prove

Lemma 6. *Let q be a prime power. If $q > 4(d+1)/\epsilon$ be a prime power. Let $n > q, d$. There is a $(1-\epsilon)$-dense hitting set for $\mathcal{H}_d$ of size*

$$m = \left\lceil \frac{(d+1)\log(q(n+1))}{\epsilon \log(\epsilon q/e(d+1))} \right\rceil = O\left(\frac{d \log n}{\epsilon \log(\epsilon q/e(d+1))}\right)$$

that can be constructed in time $O(dqn \log(qn)/\epsilon)$.

If $(d+1)/\epsilon + 1 \le q \le 4(d+1)/\epsilon$ be a prime power. Let $n > q, d$. There is a $(1-\epsilon)$-dense hitting set for $\mathcal{H}_d$ of size

$$m = \left\lceil \frac{4(q-1)^2(d+1)\ln(q(n+1))}{(q-(d+1)/\epsilon)^2\epsilon} \right\rceil = O\left(\frac{dq^2 \log n}{(q-(d+1)/\epsilon)^2\epsilon}\right)$$

that can be constructed in time $O(d(q^2/(q-d-1)^2)n\log(qn)/\epsilon)$.

We note here that such result cannot be achieved when $q < d/\epsilon$ [13].

5 Proof of the Theorems

5.1 Perfect Hash Family

Here we prove Theorem 1

Proof. Consider the set of functions

$$\mathcal{F} = \{\Delta_{\{i_1,\ldots,i_d\}}(x_1, \ldots, x_n) \mid 1 \le i_1 < \cdots < i_d \le n\}$$

in $\mathcal{F}_q[x_1, x_2, \ldots, x_n]$ where

$$\Delta_{\{i_1,\ldots,i_d\}}(x_1, \ldots, x_n) = \prod_{1\le k<j\le d}(x_{i_k} - x_{i_j}).$$

It is clear that a hitting set for $\mathcal{F}$, when each assignment is regarded as functions $f : [n] \to \mathcal{F}_q$, is (n, q, d)-PHF. Now since $\mathcal{F} \subseteq \mathcal{H}_{d(d-1)/2+1}$ the result follows from Lemma 5. □

When $q > d(d-1)/2$ is not a power of prime number then we can take the nearest prime $q' < q$ and construct an (n, q', d)-PHF that is also (n, q, d)-PHF. It is known that the nearest prime $q' \geq q - \Theta(q^{.525})$, [19], and therefore the result in the above table is also true for any integer $q \geq d(d+1)/2 + O(d^{1.05})$.

5.2 Perfect Hash Family for Small *d*

We now prove Theorem 2

Proof. If $q > d^2$ then the construction in Theorem 1 has the required size. Let $q \leq d^2$. We first use Theorem 1 to construct an (n, d^3, d)-PHF H_1 of size $O(d^2 \log n / \log d)$ in linear time. Then a (d^3, q, d)-PHF H_2 of size$O(d \log d / g(q, d))$ can be constructed in time, [7,58],

$$\binom{d^3}{d} q^{1+\lceil \log d^3 / \log q \rceil (d-1)} \leq d^{3d} q^d d^{3d} \leq d^{8d} < n.$$

Then $H = \{h_2(h_1) \mid h_2 \in H_2, h_1 \in H_1\}$ is (n, q, d)-PHF of the required size. □

It follows from [37] that this bound is within a factor of d^4 of the lower bound when $q = d + O(1)$ and within a factor of $d^3 \log d$ of the lower bound when $q > cd$ for some constant $c > 1$. See details in the following

Lemma 7. *[37] Let $n > d^{2+\epsilon}$ for some constant $\epsilon > 0$. Any (n, q, d)-PHF is of size at least*

$$\Omega\left(\frac{(q-d+1)}{q \log(q-d+2)} \frac{\log n}{g(q, d)}\right).$$

In particular, for $q = d + O(1)$ the bound is

$$\Omega\left(\frac{\log n}{d g(q, d)}\right)$$

and for $q > cd$ for some constant $c > 1$ the bound is

$$\Omega\left(\frac{\log n}{(\log d) g(q, d)}\right).$$

5.3 Dense Perfect Hash

Using Lemma 6 with the same proof as in Theorem 1 we get Theorem 3.

In the appendix we show that the size in the above Theorem and the constraint on ϵ are tight.

For the rest of the paper we will only state the results for the non-dense d-restriction problems. The results for the dense d-restriction problems follows immediately from applying Lemma 6.

5.4 Cover-Free Families

We now prove the following

Theorem 5. *Let $q \geq wr + 2$ be a prime power. Let $S \subseteq \mathcal{F}_q^n$ be a hitting set for $\mathcal{H}_{wr}$. Given a $(q, (w, r))$-CFF of size M that can be constructed in linear time one can construct an $(n, (w, r))$-CFF of size $M \cdot |S|$ that can be constructed in linear time.*

In particular, there is an (w, r)-CFF of size

$$\binom{q}{w} \cdot |S|$$

that can be constructed in linear time in its size.

In particular, for any constant $c > 1$, the following (w, r)-CFF can be constructed in linear time in their sizes

n	w	*Linear time.* *Size=O( )*	*Upper* *Bound*	*Lower* *Bound*
all	$O(1)$	$r^{w+1} \log n$	$r^{w+1} \log n$	$\frac{r^{w+1}}{\log r} \log n$
all	$o(r)$	$(ce)^w r^{w+1} \log n$	$\frac{r^{w+1}}{(w/e)^{w-1/2}} \log n$	$\frac{r^{w+1}}{(w/e)^{w+1} \log r} \log n$

Proof. Let $d = r + w$. Consider the set of non-zero functions

$$\mathcal{M} = \{\Delta_{\mathbf{i}} \mid \mathbf{i} \in [n]^d,\ i_1, i_2, \ldots, i_d \text{ are distinct}\}$$

where

$$\Delta_{\mathbf{i}}(x_1, \ldots, x_n) = \prod_{1 \leq k \leq w \text{ and } w < j \leq d} (x_{i_k} - x_{i_j}).$$

Then S is a hitting set for $\mathcal{M}$.

Let $\mathcal{F} \subseteq \{0,1\}^q$ be a $(q, (w, r))$-CFF of size M. Regard each $f \in \mathcal{F}$ as a function $f : \mathcal{F}_q \to \{0,1\}$. It is easy to see that

$$\{(f(b_1), f(b_2), \ldots, f(b_n)) \mid b \in S, f \in \mathcal{F}\} \subseteq \{0,1\}^n$$

is (w, r)-CFF of size $|\mathcal{F}| \cdot |S| = M \cdot |S|$.

Now for every subset $R \subseteq \mathcal{F}_q$ define the function $\chi_R : \mathcal{F}_q \to \{0,1\}$ where for $\beta \in \mathcal{F}_q$ we have $\chi_R(\beta) = 1$ if $\beta \in R$ and $\chi_R(\beta) = 0$ otherwise. Then $\{\chi_R \mid R \subseteq \mathcal{F}_q, |R| = w\} \subseteq \{0,1\}^{\mathcal{F}_q}$ is a $(q, (w, r))$-CFF of size $\binom{q}{w}$. Therefore

$$C = \{(\chi_R(b_1), \chi_R(b_2), \ldots, \chi_R(b_n)) \mid b \in S, R \subseteq \mathcal{F}_q, |R| = w\}$$

is (w, r)-CFF of size

$$|C| \leq \binom{q}{w} |S|.$$

Now for the results in the table consider a constant $c > c' > 1$ and let q be a power of prime such that $q = c'wr + o(wr)$. This is possible by [19]. By Lemma 5 there is a hitting set S for $\mathcal{H}_{wr}$ of size $O(wr\log n)$. This gives a (w,r)-CFF of size

$$O\left(\binom{q}{w}\cdot wr\log n\right)=O\left(\left(\frac{qe}{w}\right)^w wr\log n\right)=O\left((ce)^w r^{w+1}\log n\right)$$

that can be constructed in linear time in its size. □

5.5 Open Problems

Here we give some open problems

1. Find a polynomial time almost optimal (within $poly(d)$) construction of (n,q,d)-PHF for $q=o(d^2)$. Using the techniques in [58] it is easy to give an almost optimal construction for (n,q,d)-PHF when $q=d^2/c$ for any constant $c>1$. Unfortunately the size of the construction is within a factor of $d^{O(c)}$ of the lower bound.
2. In this paper we gave a construction of $(n,(w,r))$-CFF of size

$$\begin{aligned}&\min((2e)^w r^{w+1},(2e)^r w^{r+1})\log n\\ &=\binom{w+r}{r}2^{\min(w\log w,r\log r)(1+o(1))}\log n\end{aligned}\tag{2}$$

that can be constructed in linear time. Fomin et. al. in [38] gave a construction of size

$$\binom{w+r}{r}2^{O\left(\frac{r+w}{\log\log(r+w)}\right)}\log n\tag{3}$$

that can be constructed in linear time. The former bound, (2), is better than the latter when $w\ge r\log r\log\log r$ or $r\ge w\log w\log\log w$. We also note that the former bound, (2), is almost optimal, i.e.,

$$\binom{w+r}{r}^{1+o(1)}\log n=N^{1+o(1)}\log n,$$

where $N\log n$ is the optimal size, when $r=w^{\omega(1)}$ or $r=w^{o(1)}$ and the latter bound, (3), is almost optimal when

$$o(w\log\log w\log\log\log w)=r=\omega\left(\frac{w}{\log\log w\log\log\log w}\right).$$

Find a polynomial time almost optimal (within $N^{o(1)}$) construction for (w,r)-CFF when $w=\omega(1)$.
3. A construction is global explicit if it runs in deterministic polynomial time in the size of the construction. A local explicit construction is a construction where one can find any bit in the construction in time poly-log in the size of the construction. The constructions in this paper are linear time global explicit constructions. Some almost linear time almost optimal local explicit constructions follows from my recently published paper [14]. It is interesting to find other explicit constructions that are more optimal.

References

1. Alon, N.: Explicit construction of exponential sized families of k-independent sets. Discrete Math. **58**, 191–193 (1986)
2. Alon, N., Asodi, V.: Learning a hidden subgraph. In: Díaz, J., Karhumäki, J., Lepistö, A., Sannella, D. (eds.) ICALP 2004. LNCS, vol. 3142, pp. 110–121. Springer, Heidelberg (2004)
3. Alon, N., Beigel, R., Kasif, S., Rudich, S., Sudakov, B.: Learning a Hidden Matching. SIAM J. Comput. **33**(2), 487–501 (2004)
4. Angluin, D., Chen, J.: Learning a Hidden Hypergraph. Journal of Machine Learning Research **7**, 2215–2236 (2006)
5. Angluin, D., Chen, J.: Learning a hidden graph using $O(\log n)$ queries per edge. J. Comput. Syst. Sci. **74**(4), 546–556 (2008)
6. Alon, N., Naor, M.: Rerandomization, witnesses for Boolean matrix multiplication and construction of perfect hash functions. Algorithmica **16**, 434–449 (1996)
7. Alon, N., Moshkovitz, D., Safra, S.: Algorithmic construction of sets for k-restrictions. ACM Transactions on Algorithms **2**(2), 153–177 (2006)
8. Atici, M., Magliveras, S.S., Stinson, D.R., Wei, W.-D.: Some Recursive Constructions for Perfect Hash Families. Journal of Combinatorial Designs **4**(5), 353–363 (1996)
9. Alon, N., Bruck, J., Naor, J., Naor, M., Roth, R.M.: Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. IEEE Transactions on Information Theory **38**(2), 509–516 (1992)
10. Blackburn, S.R.: Perfect Hash Families: Probabilistic Methods and Explicit Constructions. Journal of Combinatorial Theory, Series A **92**, 54–60 (2000)
11. Blackburn, S.R.: Combinatorics and threshold cryptography. In: Holroyd, F.C., Quinn, K.A.S., Rowley, C., Webb, B.S. (eds.) Combinatorial Designs and Their Applications. Research Notes in Mathematics, vol. 403, pp. 44–70. CRC Press, London (1999)
12. Blackburn, S.R., Burmester, M., Desmedt, Y.G., Wild, P.R.: Efficient multiplicative sharing schemes. In: Maurer, U.M. (ed.) EUROCRYPT 1996. LNCS, vol. 1070, pp. 107–118. Springer, Heidelberg (1996)
13. Bshouty, N.H.: Testers and their applications. In: ITCS 2014, pp. 327–352 (2014). Full version: Electronic Colloquium on Computational Complexity (ECCC) 19, 11 (2012)
14. Bshouty, N.H.: Dense Testers: Almost Linear Time and Locally Explicit Constructions. Electronic Colloquium on Computational Complexity (ECCC) 22: 6 (2015)
15. Bshouty, N.H.: Linear time Constructions of some d-Restriction Problems. CoRR abs/1406.2108. (2014)
16. Beigel, R., Alon, N., Kasif, S., Apaydin, M.S., Fortnow, L.: An optimal procedure for gap closing in whole genome shotgun sequencing. RECOMB **2001**, 22–30 (2001)
17. Blackburn, S.R., Etzion, T., Stinson, D.R., Zaverucha, G.M.: A bound on the size of separating hash families. Journal of Combinatorial Theory, Series A **115**(7), 1246–1256 (2008)
18. Bouvel, M., Grebinski, V., Kucherov, G.: Combinatorial Search on Graphs Motivated by Bioinformatics Applications: A Brief Survey. WG **2005**, 16–27 (2005)
19. Baker, R.C., Harman, G., Pintz, J.: The difference between consecutive primes. II. Proceedings of the London Mathematical Society 83(3), 532–562 (2001)
20. Barwick, S.G., Jackson, W.-A.: Geometric constructions of optimal linear perfect hash families. Finite Fields and Their Applications **14**(1), 1–13 (2008)

21. Barwick, S.G., Jackson, W.-A., Quinn, C.T.: Optimal Linear Perfect Hash Families with Small Parameters. Journal of Combinatorial Designs **12**(5), 311–324 (2004)
22. Bazrafshan, M., van Trung, T.: Bounds for separating hash families. Journal of Combinatorial Theory, Series A **118**(3), 1129–1135 (2011)
23. Blackburn, S.R., Wild, P.R.: Optimal linear perfect hash families. Journal of Combinatorial Theory, Series A **83**(2), 233–250 (1998)
24. Chang, H., Chen, H.-B., Fu, H.-L., Shi, C.-H.: Reconstruction of hidden graphs and threshold group testing. J. Comb. Optim. **22**(2), 270–281 (2011)
25. Canetti, R., Garay, J., Itkis, G., Micciancio, D., Naor, M., Pinkas, B.: Multicast security: a taxonomy and some efficient constructions. In: Proceedings of INFOCOM 1999, vol. 2, pp. 708–716 (1999)
26. Chen, H.-B., Hwang, F.K.: A survey on nonadaptive group testing algorithms through the angle of decoding. J. Comb. Optim. **15**(1), 49–59 (2008)
27. Czech, Z.J., Havas, G., Majewski, B.S.: Perfect hashing. Theoret. Comput. Sci. **182**, 1–143 (1997)
28. Chin, F.Y.L., Leung, H.C.M., Yiu, S.-M.: Non-adaptive complex group testing with multiple positive sets. Theor. Comput. Sci. **505**, 11–18 (2013)
29. Dyer, M., Fenner, T., Frieze, A., Thomason, A.: On key storage in secure networks. J. Cryptol. **8**, 189–200 (1995)
30. Du, D.Z., Hwang, F.K.: Combinatorial group testing and its applications. Series on Applied Mathematics. 2nd edn., vol. 12. World Scientific, New York (2000)
31. Dýachkov, A.G., Rykov, V.V.: Bounds on the length of disjunctive codes. Problemy Peredachi Inf. **18**(3), 7–13 (1982)
32. Dýachkov, A.G., Rykov, V.V., Rashad, A.M.: Superimposed distance codes. Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform. **18**(4), 237–250 (1989)
33. Desmedt, Y., Safavi-Naini, R., Wang, H., Batten, L., Charnes, C., Pieprzyk, J.: Broadcast anti-jamming systems. Comput. Networks 35, 223–236 (2001)
34. D'yachkov, A., Vilenkin, P., Macula, A., Torney, D.: Families of finite sets in which no intersection of ℓ sets is covered by the union of s others. J. Comb. Theory Ser. A. 99, 195–218 (2002)
35. Gao, H., Hwang, F.K., Thai, M.T., Wu, W., Znati, T.: Construction of d(H)-disjunct matrix for group testing in hypergraphs. J. Comb. Optim. **12**(3), 297–301 (2006)
36. Füredi, Z.: On r-cover-free families. Journal of Combinatorial Theory, Series A **73**(1), 172–173 (1996)
37. Fredman, M.L., Komlós, J.: On the size of seperating systems and families of perfect hash function. SIAM J. Algebraic and Discrete Methods **5**(1), 61–68 (1984)
38. Fomin, F.V., Lokshtanov, D., Saurabh, S.: Efficient Computation of Representative Sets with Applications in Parameterized and Exact Algorithms. SODA **2014**, 142–151 (2014)
39. Fiat, A., Naor, M.: Broadcast encryption. In: Stinson, D.R. (ed.) CRYPTO 1993. LNCS, vol. 773, pp. 480–491. Springer, Heidelberg (1994)
40. Grebinski, V., Kucherov, G.: Reconstructing a Hamiltonian Cycle by Querying the Graph: Application to DNA Physical Mapping. Discrete Applied Mathematics **88**(1–3), 147–165 (1998)
41. Garay, J.A., Staddon, J., Wool, A.: Long-Lived Broadcast Encryption. In: Bellare, M. (ed.) CRYPTO 2000. LNCS, vol. 1880, pp. 333–352. Springer, Heidelberg (2000)
42. Huang, T., Wang, K., Weng, C.-W.: A class of error-correcting pooling designs over complexes. J. Comb. Optim. **19**(4), 486–491 (2010)

43. Indyk, P., Ngo, H.Q., Rudra, A.: Efficiently decodable non-adaptive group testing. In: The 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2010), pp. 1126–1142 (2010)
44. Körner, J.: Fredman-Komlós bounds and information theory. SIAM J. Algebraic and Discrete Methods **7**(4), 560–570 (1986)
45. Körner, J., Marton, K.: New bounds for perfect hashing via information theory. Europ. J. of Combinatorics **9**(6), 523–530 (1988)
46. Kumar, R., Rajagopalan, S., Sahai, A.: Coding constructions for blacklisting problems without computational assumptions. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 609–623. Springer, Heidelberg (1999)
47. Kautz, W.H., Singleton, R.C.: Nonrandom binary superimposed codes. IEEE Trans. Inform. Theory **10**(4), 363–377 (1964)
48. Liu, L., Shen, H.: Explicit constructions of separating hash families from algebraic curves over finite fields. Designs, Codes and Cryptography **41**(2), 221–233 (2006)
49. Martirosyan, S.: Perfect Hash Families, Identifiable Parent Property Codes and Covering Arrays. Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften (2003)
50. Mehlborn, K.: Data Structures and Algorithms. 1. Sorting and Searching. Springer, Berlin (1984)
51. Mehlhorn, K.: On the program size of perfect and universal hash functions. In: Proceedings of the 23rd IEEE Symposium on Foundations of Computer Science (FOCS 1982), pp. 170–175 (1982)
52. Mitchell, C.J., Piper, F.C.: Key storage in secure networks. Discrete Appl. Math. **21**, 215–228 (1988)
53. Macula, A.J., Popyack, L.J.: A group testing method for finding patterns in data. Discret Appl Math. **144**, 149–157 (2004)
54. Martirosyan, S., van Trung, T.: Explicit constructions for perfect hash families. Designs, Codes and Cryptography **46**(1), 97–112 (2008)
55. Nilli, A.: Perfect hashing and probability. Combinatorics, Probability and Computing **3**(3), 407–409 (1994)
56. Ngo, H.Q., Du, D.Z.: A survey on combinatorial group testing algorithms with applications to DNA library screening. Theoretical Computer Science **55**, 171–182 (2000)
57. Naor, J., Naor, M.: Small-bias probability spaces: efficient constructions and applications. SIAM J. Comput. **22**(4), 838–856 (1993)
58. Naor, M., Schulman, L.J., Srinivasan, A.: Splitters and Near-optimal Derandomization. FOCS **95**, 182–191 (1995)
59. Newman, I., Wigderson, A.: Lower bounds on formula size of Boolean functions using hypergraph entropy. SIAM J. Discrete Math. **8**, 536–542 (1995)
60. Porat, E., Rothschild, A.: Explicit Nonadaptive Combinatorial Group Testing Schemes. IEEE Transactions on Information Theory **57**(12), 7982–7989 (2011)
61. Reyzin, L., Srivastava, N.: Learning and verifying graphs using queries with a focus on edge counting. In: Hutter, M., Servedio, R.A., Takimoto, E. (eds.) ALT 2007. LNCS (LNAI), vol. 4754, pp. 285–297. Springer, Heidelberg (2007)
62. Stinson, D.R.: On some methods for unconditionally secure key distribution and broadcast encryption. Des. Codes Cryptogr. **12**, 215–243 (1997)
63. Stinson, D.R., Wei, R.: Combinatorial properties and constructions of traceability schemes and frameproof codes. SIAM J. Discrete Math. **11**, 41–53 (1998)
64. Safavi-Naini, R., Wang, H.: Multireceiver authentication codes: models, bounds, constructions, and extensions Inform. Comput. **151**, 148–172 (1999)

65. Stinson, D.R., Wei, R., Zhu, L.: New constructions for perfect hash families and related structures using combintorial designs and codes. J. Combin. Designs. **8**(3), 189–200 (2000)
66. Stinson, D.R., Wei, R., Chen, K.: On generalised separating hash families. Journal of Combinatorial Theory, Series A **115**(1), 105–120 (2008)
67. Stinson, D.R., Wei, R., Zhu, L.: Some new bounds for cover-free families. Journal of Combinatorial Theory, Series A **90**(1), 224–234 (2000)
68. Stinson, D.R., van Trung, T.: Some new results on key distribution patterns and broadcast encryption. Des. Codes Cryptogr. **14**, 261–279 (1998)
69. Stinson, D.R., van Trung, T., Wei, R.: Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. J. Stat. Planning and Inference **86**(2), 595–617 (2000)
70. Torney, D.C.: Sets pooling designs. Ann. Comb. **3**, 95–101 (1999)
71. Wang, H., Xing, C.P.: Explicit Constructions of perfect hash families from algebraic curves over finite fields. J. of Combinatorial Theory, Series A 93(1), 112–124 (2001)