Some intriguing upper bounds for separating hash families

Gennian Ge$^{1,*}$, Chong Shangguan$^2$ & Xin Wang$^3$

$^1$School of Mathematical Sciences, Capital Normal University, Beijing 100048, China;
$^2$School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China;
$^3$Department of Mathematics, Soochow University, Suzhou 215006, China

Email: gnge@zju.edu.cn, theoreming@163.com, xinw@suda.edu.cn

Received September 19, 2017; accepted August 14, 2018; published online September 17, 2018

Abstract An $N \times n$ matrix on $q$ symbols is called $\{w_1, \ldots, w_t\}$-separating if for arbitrary $t$ pairwise disjoint column sets $C_1, \ldots, C_t$ with $|C_i| = w_i$ for $1 \leq i \leq t$, there exists a row $f$ such that $f(C_1), \ldots, f(C_t)$ are also pairwise disjoint, where $f(C_i)$ denotes the collection of components of $C_i$ restricted to row $f$. Given integers $N, q$ and $w_1, \ldots, w_t$, denote by $C(N, q, \{w_1, \ldots, w_t\})$ the maximal $n$ such that a corresponding matrix does exist. The determination of $C(N, q, \{w_1, \ldots, w_t\})$ has received remarkable attention during the recent years. The main purpose of this paper is to introduce two novel methodologies to attack the upper bound of $C(N, q, \{w_1, \ldots, w_t\})$. The first one is a combination of the famous graph removal lemma in extremal graph theory and a Johnson-type recursive inequality in coding theory, and the second one is the probabilistic method. As a consequence, we obtain several intriguing upper bounds for some parameters of $C(N, q, \{w_1, \ldots, w_t\})$, which significantly improve the previously known results.

Keywords separating hash families, Johnson-type recursive bound, graph removal lemma, probabilistic method

MSC(2010) 68R05, 97K20, 94B25

Citation: Ge G, Shangguan C, Wang X. Some intriguing upper bounds for separating hash families. Sci China Math, 2019, 62: 269–282, https://doi.org/10.1007/s11425-017-9378-y

1 Introduction

Separating hash families are useful combinatorial objects introduced by Stinson et al. [31] in 2008. They are generalizations of many previously known and well-studied combinatorial objects. Therefore, during the recent years they have received remarkable attention. Many efforts have been made to determine the upper and lower bounds for separating hash families (see, for example, [5, 9, 26, 31]).

Let $X$ and $Y$ be sets of cardinalities $n$ and $q$, respectively. We call a family $F$ of $N$ functions $f : X \to Y$ an $(N; n, q)$-hash family. Let $f : X \to Y$ be a function and consider pairwise disjoint subsets $C_1, \ldots, C_t \subseteq X$. We say that $f$ separates $C_1, \ldots, C_t$ if $f(C_1), \ldots, f(C_t)$ are pairwise disjoint subsets of $Y$. We further say that $F$ is an $(N; n, q, \{w_1, \ldots, w_t\})$-separating hash family (which will be also denoted as an SHF($N; n, q, \{w_1, \ldots, w_t\}$)) if it satisfies the following property: For all pairwise disjoint
subsets $C_1, \ldots, C_t \subseteq X$ with $|C_i| = w_i$ for $1 \leq i \leq t$, there exists at least one function $f \in \mathcal{F}$ that separates $C_1, \ldots, C_t$. We call the multiset $\{w_1, \ldots, w_t\}$ the type of this separating hash family and denote $u := \sum_{i=1}^t w_i$ throughout this paper. Without loss of generality, we may fix the alphabet set $Y$ to be $[q]$, where $[q] := \{1, \ldots, q\}$. We also assume that $w_1 \leq \cdots \leq w_t$.

When $w_1 = \cdots = w_t = 1$, an SHF($N; n, q, \{1, \ldots, 1\}$) is also known to be a $t$-perfect hash family. Perfect hash families were introduced by Mehlhorn [19] in 1984 and have found applications in cryptography [32, 34], database management [19], and the designs of circuits [20] and algorithms [2]. When $t = 2$, $w_1 = 1$ and $w_2 = w \geq 2$, an SHF($N; n, q, \{1, w\}$) is known as a $w$-frameproof code. When $t = 2$, $w_1 = w_2 = w \geq 2$, an SHF($N; n, q, \{w, w\}$) is known as a $w$-secure-frameproof code. Frameproof and secure-frameproof codes were introduced by Chor et al. [10] in 1994, and by Stinson et al. [30] in 2000, respectively. They can serve as techniques to prevent copyrighted materials from unauthorized use [8, 29]. Moreover, codes with the identifiable parent property [17] are separating hash families which are simultaneously of type $\{1, 1, 1\}$ and $\{2, 2\}$, and partially hash families [4] are separating hash families satisfying $w_1 = \cdots = w_{t-1} = 1$ and $w_t = w \geq 2$. The reader is referred to [27] for recent progress on codes with the identifiable parent property.

The determinations of upper and lower bounds for separating hash families are important open problems in this research area. Given integers $N, q$ and $w_1, \ldots, w_t$, denote by $C(N, q, \{w_1, \ldots, w_t\})$ the maximal $n$ such that there exists an SHF($N; n, q, \{w_1, \ldots, w_t\}$). In the literature, there are two major known approaches to attack the bound $C(N, q, \{w_1, \ldots, w_t\})$. The first one is called the grouping coordinates method, which states that

$$C(aN, q, \{w_1, \ldots, w_t\}) \leq C(N, q^a, \{w_1, \ldots, w_t\})$$

holds for every positive integer $a$. This inequality can be proved directly by regarding a family of $q$-ary length $aN$ vectors as a family of $q^a$-ary length $N$ vectors. One can verify that the new family preserves the $\{w_1, \ldots, w_t\}$-separating property as long as the original family does. Due to this inequality, the problem of bounding $C(N, q, \{w_1, \ldots, w_t\})$ can be reduced to bounding $C(u-1, q, \{w_1, \ldots, w_t\})$, since one can show

$$C(N, q, \{w_1, \ldots, w_t\}) \leq C(u-1, q^{[N/(u-1)]}, \{w_1, \ldots, w_t\}).$$

In 2008, Blackburn et al. [9] obtained a linear bound (linear in $q$) for $C(u-1, q, \{w_1, \ldots, w_t\})$, which states that

$$C(u-1, q, \{w_1, \ldots, w_t\}) \leq (w_1 w_2 + u - w_1 - w_2)q.$$ 

This bound immediately implies the general upper bound

$$C(N, q, \{w_1, \ldots, w_t\}) \leq (w_1 w_2 + u - w_1 - w_2)q^{\left\lceil \frac{N}{w_1} \right\rceil}.$$ 

An improved bound was obtained by Bazrafshan and Trung [5] in 2011, showing that

$$C(u-1, q, \{w_1, \ldots, w_t\}) \leq (u-1)q.$$ 

Moreover, they conjectured that $u-1$ was the smallest linear factor satisfying this bound. One can easily see that the upper bound derived from the grouping coordinates method could never be better than $C(N, q, \{w_1, \ldots, w_t\}) \leq (u-1)q^{\left\lceil \frac{N}{w_1} \right\rceil}$, provided that Bazrafshan and Trung’s conjecture was true. Indeed, recently the correctness of this conjecture was confirmed by the first two authors of this paper in [26]. Moreover, in the same paper the authors have introduced a new method to study $C(N, q, \{w_1, \ldots, w_t\})$. It was shown that $C(N, q, \{w_1, \ldots, w_t\})$ satisfies the following Johnson-type recursive inequality.

**Lemma 1.1** (Johnson-type bound). Let $l$, $1 \leq l \leq N$, be a positive integer. Then it holds that

$$C(N, q, \{w_1, \ldots, w_t\}) \leq q^l + \max\{u-1, C(N-l, q, \{w_1-1, \ldots, w_t\})\}.$$ 

Indeed, in the right-hand side of the inequality we can choose the minus of 1 to be placed after arbitrary $w_i, 1 \leq i \leq t$. 

Using this inequality, the best known general upper bound for \( C(N, q, \{w_1, \ldots, w_t\}) \) was derived in [26], which is stated as follows.

**Theorem 1.2.** Suppose there exists an SHF\((N; n, q, \{w_1, \ldots, w_t\})\). Let \( u = \sum_{i=1}^{t} w_i \) and let \( r, 1 \leq r \leq u - 1 \), be the positive integer such that \( N \equiv r \pmod{u - 1} \). If \( C([N/(u - 1)], q, \{w_1, \ldots, w_t\}) \geq u \), then it holds that

\[
 n \leq rq^{\lceil N/(u-1) \rceil} + (u - 1 - r)q^{[N/(u-1)]}.
\]

Obviously, for \((u - 1) \nmid N\) the above bound is an improvement of the previous ones since

\[
 r < u - 1 \leq w_1w_2 + u - w_1 - w_2.
\]

Indeed, it behaves very well for fixed \(w_1, \ldots, w_t, N\) and sufficiently large \(q\). The exponent \( \lceil \frac{N}{u-1} \rceil \) is realistic in the sense that there are probabilistic constructions (see [7] for a proof using the Lovász local lemma or Theorem 2.1 of [3] for a proof using the alteration method) showing that

\[
 C(N, q, \{w_1, \ldots, w_t\}) \geq \frac{1}{2u} \left( \frac{1}{1 - g(q, u)} \right)^{\frac{N}{u-1}},
\]

where

\[
 g(q, u) = \frac{q(q-1) \cdots (q-u+1)}{q^u},
\]

which implies that

\[
 C(N, q, \{w_1, \ldots, w_t\}) \geq \frac{1}{2u} \left( \frac{q}{u} \right)^{\frac{N}{u-1}} = \Omega_u(N(q^{\frac{N}{u-1}}))
\]

holds for sufficiently large \(q\) and fixed \(w_1, \ldots, w_t, N\). It follows that the exponent \( \lceil \frac{N}{u-1} \rceil \) cannot be further reduced when \((u - 1) \nmid N\). It is an open problem [9] to determine whether the exponent is tight when \((u - 1) \nmid N\). After many efforts [14, 34], the first breakthrough results in this direction were obtained in [26]. Answering an open problem of Walker II and Colbourn [34], it was shown

\[
 q^{2-o(1)} < C(3, q, \{1, 1, 1\}) = o(q^2) \quad (1.1)
\]

and

\[
 q^{2-o(1)} < C(4, q, \{1, 1, 1, 1\}) = o(q^2) \quad (1.2)
\]

hold for sufficiently large \(q\). In this paper we will develop a systematic method to study the upper bound for \( C(N, q, \{w_1, \ldots, w_t\}) \) when \((u - 1) \nmid N\). We will extend the upper bounds of (1.1) and (1.2) to more general parameters.

On the other hand, the bound presented in Theorem 1.2 does not behave quite well for relatively small \(q\) (compared with \(u\) and \(N\)). For example, in [18, 21] it was shown that for \(w_1 = \cdots = w_t = 1\) it holds that

\[
 C(N, q, \{1, \ldots, 1\}) \leq \min_{0 \leq j \leq t-2} (t - j - 1) \left( \frac{q-j}{j} \right)^{(1-o(1))q(j+1)N} \quad (1.3)
\]

where the \(o(1)\) term is induced by the estimation \( \binom{n}{t} \approx \frac{n^t}{t!} \). Setting \(q = t\) one can deduce from (1.3) that

\[
 C(N, t, \{1, \ldots, 1\}) \leq \min \left\{ 2^{(1-o(1))\frac{t^t}{t!}N}, (t - 1) \left( \frac{t}{t-1} \right)^N \right\}, \quad (1.4)
\]

which is much better than the bound \( C(N, t, \{1, \ldots, 1\}) = O(t^{\frac{N}{t^t}}) \) obtained by Theorem 1.2. The paper [12] contains new upper bounds on SHF\((N, n, q, \{w_1, w_2\})\) for some small \(q\) and small \(w_1\) and \(w_2\). In [24] it was shown that \( C(N, 2, \{2, 2\}) \) \(\leq O(2^{0.28N})\) which is superior to the bound \(O(2^{0.33N})\) obtained by Theorem 1.2.

For the very special case \(t = 2\) and \(q = 2\), we use \(N(w)\) to denote the minimal integer \(N\) such that there exists an SHF\((N; n, 2, \{1, w\})\) satisfying \(n > N\). The determination of \(N(w)\) has received considerable
attention. In [16] it was shown $N(w) \geq 3w$. The best known result for $N(w)$ was obtained by Miao and the authors of this paper in [28], which shows that

$$N(w) > \binom{w+1}{2}. \quad (1.5)$$

The main purpose of this paper is to introduce two novel methods, namely, the graph removal lemma (see Lemma 3.3 below) combined with the Johnson-type inequality, and the probabilistic method, to attack the upper bound of $C(N,q,\{w_1,\ldots,w_t\})$. As a result, we can improve the bound presented in Theorem 1.2 in various aspects.

- Firstly, it was widely believed that $C(4,q,\{2,2\}) \leq c_1q^2 + c_2q$ and many efforts had been made to improve the constants $c_1$ and $c_2$ (see, for example, [6,29–31,33]). Recently, the best upper bound along this line is proved by Niu and Cao [22], which states that $C(2w;q,\{w,w\}) \leq (q-1)^2+1$. However, for sufficiently large $q$, using the graph removal lemma we can prove an asymptotically optimal bound showing that $q^{2-o(1)} < C(4,q,\{2,2\}) = o(q^2)$. Our bound is somewhat surprising in the sense that it is the first bound in the literature showing that the order $q^2$ is not achievable. By combining the graph removal lemma and the Johnson-type recursive inequality, we are able to prove a much more general bound (see Theorem 4.1 below), showing that $C(w_1+\cdots+w_t,q,\{w_1,\ldots,w_t\}) = o(q^2)$ holds for arbitrary $w_1,\ldots,w_t$ except for $t = 2$ and $\{w_1,w_2\} = \{1,w\}$.

- Secondly, for the special case $w_1 = \cdots = w_t = w$ and $q = t$, previously known result implies that $C(N,t,\{w,\ldots,w\}) = O(tNw^{-1})$. By employing a probabilistic argument, we can show $C(N,t,\{w,\ldots,w\}) = O(\frac{w^3}{tN}t^2)$ (see Theorem 4.10 below), which substantially improves the bound obtained by Theorem 1.2.

- Thirdly, by relating the determination of $C(N,2,\{1,w\})$ to an old conjecture of Erdős et al. [13] on cover-free families, we improve the result of (1.5) by showing that $N(w) > \frac{15 + \sqrt{33}}{24} (w-2)^2$ (see Theorem 6.1 below), which also implies a tight bound that

$$C(N,2,\{1,w\}) = N \quad \text{for} \quad N \leq \frac{15 + \sqrt{33}}{24} (w-2)^2.$$

The rest of this paper is organized as follows. Section 2 contains the necessary notation and definitions for the proofs in this paper. In Section 3, we use the graph removal lemma to derive an upper bound for a problem in graph theory. The bound itself is of independent interest and can also be used to derive an upper bound for SHF($4;n,q,\{2,2\}$). In Section 4, we use the Johnson-type recursive inequality to prove our general bound for SHF($\sum_{i=1}^{t} w_i;n,q,\{w_1,\ldots,w_t\}$). In Section 5, we use a probabilistic method to study the behavior of $C(N,q,\{w_1,\ldots,w_t\})$ for the special case $q = t$. In Section 6, we relate the determination of $N(w)$ to a conjecture of Erdős et al. [13] and present an improved upper bound for $N(w)$. Section 7 contains some concluding remarks and open problems.

2 Preliminaries
2.1 Representation matrices, separating hash families and hypergraphs

In general, one can regard a separating hash family as a matrix. An $N \times n$ matrix $M$ on $q$ symbols is called $\{w_1,\ldots,w_t\}$-separating if for arbitrary $t$ pairwise disjoint column sets $C_1,\ldots,C_t$ with $|C_i| = w_i$ for $1 \leq i \leq t$, there exists a row $f$ such that $f(C_1),\ldots,f(C_t)$ are also pairwise disjoint, where $f(C_i)$ denotes the collection of components of $C_i$ restricted to row $f$. We call $M$ the representation matrix of an SHF($N;n,q,\{w_1,\ldots,w_t\}$).

We will use hypergraphs to study separating hash families. A hypergraph $\mathcal{H} = (V(\mathcal{H}), E(\mathcal{H}))$ can be viewed as a pair of vertices and edges, where the vertex set $V(\mathcal{H})$ can be regarded as a finite set $X$ and the edge set $E(\mathcal{H})$ can be regarded as a collection of subsets of $X$. For the sake of simplicity, we write $\mathcal{H}$ to represent the edge set $E(\mathcal{H})$, and hence $|\mathcal{H}|$ stands for $|E(\mathcal{H})|$. A hypergraph $\mathcal{H}$ is said to be linear if
for all distinct \( A, B \in \mathcal{H} \) it holds that \( |A \cap B| \leq 1 \). Furthermore, we say \( \mathcal{H} \) is \( r \)-uniform if \( |A| = r \) for every \( A \in \mathcal{H} \).

An \( r \)-uniform hypergraph \( \mathcal{H} \) is \( r \)-partite if its vertex set \( V(\mathcal{H}) \) can be colored in \( r \) colors in such a way that no edge of \( \mathcal{H} \) contains two vertices of the same color. In such a coloring, the color classes of \( V(\mathcal{H}) \), i.e., the sets of all vertices of the same color, are called vertex parts of \( \mathcal{H} \). We use \( V_1, \ldots, V_r \) to denote the \( r \) color classes of \( V(\mathcal{H}) \). Then \( V(\mathcal{H}) \) is a disjoint union of the \( V_i \)'s and for every \( A \in \mathcal{H} \), \( |A \cap V_i| = 1 \) holds for each \( i, 1 \leq i \leq r \).

One can also regard an \( N \)-uniform \( N \)-partite hypergraph \( \mathcal{H} \) with \( |V_1| = \cdots = |V_N| = q \) as an \( N \times |\mathcal{H}| \) \( q \)-ary matrix \( M \), which can be realized as follows. The rows of \( M \) are indexed by the vertex parts of \( \mathcal{H} \) and columns of \( M \) are indexed by the edges of \( \mathcal{H} \). For each \( i, 1 \leq i \leq r \), without loss of generality we can assume that \( V_i = \{ v_{i1}, \ldots, v_{iq} \} \), and then \( V(\mathcal{H}) = \{ v_{ij} \mid 1 \leq i \leq N, 1 \leq j \leq q \} \) is well-defined. Every edge \( A \in \mathcal{H} \) (i.e., column of \( M \)) must have the form \( A = \{ v_{1j_1}, \ldots, v_{Nj_N} \} \), where \( 1 \leq j_k \leq q \) for each \( k, 1 \leq k \leq N \). The entry in row \( i \) and column \( A \) is just the vertex \( V_i \cap A \in V(\mathcal{H}) \). We call \( M \) the representation matrix of the hypergraph \( \mathcal{H} \). Observe that \( |V(\mathcal{H})| = \sum_{i=1}^{N} |V_i| = Nq \), it follows that \( M \) is an \( Nq \)-ary matrix. However, when investigating the separating property of such a matrix, we only want to know whether there exists a row that can separate several sets of columns. Thus two entries of \( M \) located in two different rows will not interact each other. So for simplicity one can just view \( M \) as a \( q \)-ary matrix by setting \( V_i = \{ v_{i1}, \ldots, v_{iq} \} = \{ 1, \ldots, q \} \) for each \( i, 1 \leq i \leq N \).

### 2.2 Hypergraph rainbow cycles

To study the upper bound of \( C(4,q,\{2,2\}) \) we will use the notion of rainbow 4-cycles. The formal definition of rainbow cycles is presented as follows.

**Definition 2.1.** Let \( \mathcal{H} \) be an \( r \)-uniform \( r \)-partite linear hypergraph with vertex parts \( V_1, \ldots, V_r \). A rainbow \( k \)-cycle of \( \mathcal{H} \) is an alternating sequence of vertices and edges of the form \( v_1, E_1, v_2, E_2, \ldots, v_k, E_k, v_1 \) such that

(a) \( v_1, v_2, \ldots, v_k \) are distinct vertices of \( \mathcal{H} \);

(b) \( E_1, E_2, \ldots, E_k \) are distinct edges of \( \mathcal{H} \);

(c) \( v_i, v_{i+1} \in E_i \) for \( 1 \leq i \leq k-1 \) and \( v_k, v_1 \in E_k \);

(d) for \( 1 \leq i \leq k, v_i \in V_{j_i} \) and \( j_i \neq j_{i'} \) for arbitrary \( i \neq i' \).

Condition (d) implies that all \( k \) vertices are located in \( k \) different vertex parts. It is easy to see that for \( r \)-partite hypergraphs, a rainbow \( k \)-cycle exists only if \( k \leq r \). In Section 3, we will present an upper bound for the maximal number of edges of a linear hypergraph that contains no rainbow cycles (see Theorem 3.4 below). In Section 4, this bound will be used to derive an upper bound for separating hash families.

### 2.3 Cover-free families and a conjecture of Erdős, Frankl and Füredi

Let \( X \) be a set of \( N \) elements. A family \( \mathcal{F} \subseteq 2^X \) is said to be \( w \)-cover-free if for arbitrary distinct \( w + 1 \) members \( A_0, A_1, \ldots, A_w \) of \( \mathcal{F} \) it holds that \( A_0 \not\subseteq A_1 \cup A_2 \cup \cdots \cup A_w \). Suppose that \( |\mathcal{F}| = n \) and let us denote \( X = \{ x_1, \ldots, x_N \} \) and \( \mathcal{F} = \{ A_1, \ldots, A_n \} \). \( \mathcal{F} \) will be denoted as a \( \text{CFF}(N; n, w) \). Denote by \( M^* \) the representation matrix of \( \mathcal{F} \), which is an \( N \times n \) binary matrix whose rows are indexed by the elements of \( X \) and whose columns are indexed by the members of \( \mathcal{F} \), such that the entry in the \( i \)-th row and the \( j \)-th column is 1 if and only if \( x_i \in A_j \). In a binary matrix, the weight of a column is simply the number of 1’s contained in it. We will need a property of cover-free families.

**Lemma 2.2** ([See [23]]) Let \( \mathcal{F} \) be a \( \text{CFF}(N; n, w) \) with representation matrix \( M^* \). Fix an arbitrary member \( A \) of \( \mathcal{F} \) and consider the new family \( \mathcal{F}_1 \) defined by

1. \( \mathcal{F}_1 \subseteq 2^{X \setminus A} \);
2. \( \mathcal{F}_1 = \{ B \setminus A : B \in \mathcal{F}, B \neq A \} \).

Then \( \mathcal{F}_1 \) is a \( \text{CFF}(N - |A|; n - 1, w - 1) \).
Lemma 2.4 restated as follows.

Example 3.1. Let $\mathcal{F}$ be an SHF($4; n, q, \{w, w\}$) and $\mathcal{H}_F$ be the hypergraph defined by the representation matrix of $\mathcal{F}$. Assume $\mathcal{H}_F$ contains a rainbow 4-cycle $v_4, A_1, v_1, A_2, v_2, A_3, v_3, A_4, v_4$, which can be depicted in Table 1. Obviously, no row of the representation matrix can separate $\{A_1, A_3\}$ and $\{A_2, A_4\}$, which violates the $\{2, 2\}$-separating property.

Example 3.2. Let $\mathcal{F}$ be an SHF($2w; n, q, \{w, w\}$) and $\mathcal{H}_F$ be the hypergraph defined by the representation matrix of $\mathcal{F}$. Assume $\mathcal{H}_F$ contains a 2w-cycle, which can be denoted as $v_{2w}, A_1, v_1, A_2, v_2, A_3, \ldots, A_{2w-1}, v_{2w-1}, A_{2w}, v_{2w}$. As in Example 3.1, this 2w-cycle can be depicted in Table 2. Obviously, no row of the representation matrix can separate $\{A_1, A_3, \ldots, A_{2w-1}\}$ and $\{A_2, A_4, \ldots, A_{2w}\}$, which violates the $\{w, w\}$-separating property.

Lemma 3.3 (Graph removal lemma [11]). For any graph $G$ and any $\epsilon > 0$, there exists $\delta > 0$ such that any graph on $n$ vertices which contains at most $\delta n^{v(G)}$ copies of $G$ can be made $G$-free by removing at most $\epsilon n^2$ edges.

Using the graph removal lemma described in Lemma 3.3, it is straightforward to deduce the following fact: For any given constant $\epsilon > 0$, there exists some $\delta(\epsilon) > 0$ such that if one must delete at least $\epsilon n^2$ edges to make a graph $H$ with $n$ vertices $G$-free, then $H$ must contain at least $\delta(\epsilon)n^{v(G)}$ copies of $G$.

| Table 1 | Rainbow 4-cycle |
|---------|-----------------|
|         | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
| $V_1$   | $v_1$ |   $v_1$ |
| $V_2$   |   $v_2$ |   $v_2$ |
| $V_3$   |   $v_3$ |   $v_3$ |
| $V_4$   |   $v_4$ |           |

Proof. The first two parameters in CFF($N - |A|; n - 1, w - 1$) are easy to verify. It suffices to prove that $\mathcal{F}_i$ is a ($w - 1$)-cover-free family. Suppose otherwise, there are $w$ different members $B_0, B_1, \ldots, B_{w-1}$ of $\mathcal{F}_i$ such that $B_0 \subseteq B_1 \cup \cdots \cup B_{w-1}$. For each $i$, $0 \leq i \leq w - 1$, denote by $A_i$ the member of $\mathcal{F}$ such that $B_i = A_i \setminus A$. Then it holds that

$$A_0 \subseteq B_0 \cup A \subseteq (B_1 \cup \cdots \cup B_{w-1}) \cup A \subseteq A_1 \cup \cdots \cup A_{w-1} \cup A,$$

which violates the $w$-cover-free property of $\mathcal{F}$.

Denote by $N^*(w)$ the minimal $N$ such that there exists a CFF($N; n, w$) with $n > N$. In 1985, Erdős et al. [13] posed the following conjecture:

Conjecture 2.3 (See [13]). $\lim_{w \to \infty} \frac{N^*(w)}{w} = 1$, or in an even stronger form $N^*(w) \geq (w + 1)^2$.

Note that when $w + 1$ is a prime power, an affine plane of order $w + 1$ induces an example for CFF($w + 1; (w + 1)^2 + (w + 1), w$). The best known lower bound for $N^*(w)$ is proved in [25], which is restated as follows.

Lemma 2.4 (See [25]). It holds that $N^*(w) \geq \frac{15 + \sqrt{245}}{24}w^2$.

By establishing a bridge between cover-free families and separating hash families, this lemma will be used to prove a lower bound for $N(w)$ in Section 6.

3 The maximal number of edges of a linear hypergraph which contains no rainbow cycles

Rainbow cycles are closely related to separating hash families in the sense that an SHF($2w; n, q, \{w, w\}$) cannot contain a rainbow 2w-cycle. In the following we will present two examples to illustrate this observation.

Example 3.1. Let $\mathcal{F}$ be an SHF($4; n, q, \{2, 2\}$) and $\mathcal{H}_F$ be the hypergraph defined by the representation matrix of $\mathcal{F}$. Assume $\mathcal{H}_F$ contains a rainbow 4-cycle $v_4, A_1, v_1, A_2, v_2, A_3, v_3, A_4, v_4$, which can be depicted in Table 1. Obviously, no row of the representation matrix can separate $\{A_1, A_3\}$ and $\{A_2, A_4\}$, which violates the $\{2, 2\}$-separating property.

Example 3.2. Let $\mathcal{F}$ be an SHF($2w; n, q, \{w, w\}$) and $\mathcal{H}_F$ be the hypergraph defined by the representation matrix of $\mathcal{F}$. Assume $\mathcal{H}_F$ contains a 2w-cycle, which can be denoted as $v_{2w}, A_1, v_1, A_2, v_2, A_3, \ldots, A_{2w-1}, v_{2w-1}, A_{2w}, v_{2w}$. As in Example 3.1, this 2w-cycle can be depicted in Table 2. Obviously, no row of the representation matrix can separate $\{A_1, A_3, \ldots, A_{2w-1}\}$ and $\{A_2, A_4, \ldots, A_{2w}\}$, which violates the $\{w, w\}$-separating property.

Lemma 3.3 (Graph removal lemma [11]). For any graph $G$ and any $\epsilon > 0$, there exists $\delta > 0$ such that any graph on $n$ vertices which contains at most $\delta n^{v(G)}$ copies of $G$ can be made $G$-free by removing at most $\epsilon n^2$ edges.

Using the graph removal lemma described in Lemma 3.3, it is straightforward to deduce the following fact: For any given constant $\epsilon > 0$, there exists some $\delta(\epsilon) > 0$ such that if one must delete at least $\epsilon n^2$ edges to make a graph $H$ with $n$ vertices $G$-free, then $H$ must contain at least $\delta(\epsilon)n^{v(G)}$ copies of $G$.
We use a complete graph whose every two distinct vertices form an edge.

Remark 4.2. The only case excluded by Theorem 4.1 is 

\[ \sum_{i=1}^{t} w_i; n, q, \{w_1, \ldots, w_t\} \]

Theorem 3.4. Let \( n \) be a sufficiently large positive integer and \( \epsilon > 0 \) be a given positive constant. Let \( r \) be a fixed positive integer. Assume \( \mathcal{H} \) is an \( r \)-uniform \( r \)-partite linear hypergraph with \( en^2 \) edges. Then \( \mathcal{H} \) must contain a rainbow \( k \)-cycle for every integer \( k, 3 \leq k \leq r \).

Proof. Note that a complete graph is an ordinary graph whose every two distinct vertices form an edge.

The goal of this section is to prove the following general upper bound.

4 An upper bound for \( \text{SHF}(\sum_{i=1}^{t} w_i; n, q, \{w_1, \ldots, w_t\}) \)

Theorem 4.1. Let \( \mathcal{F} \) be an \( \text{SHF}(\sum_{i=1}^{t} w_i; n, q, \{w_1, \ldots, w_t\}) \), where the \( w_i \)'s are fixed positive integers and \( q \) is a sufficiently large integer. Then for either \( t \geq 3 \) or \( t = 2 \) and \( \min\{w_1, w_2\} \geq 2 \), it holds that \( n = o(q^2) \).

Remark 4.2. The only case excluded by Theorem 4.1 is \( t = 2 \) and \( \{w_1, w_2\} = \{1, w\} \). We have mentioned that a \( \{1, w\} \)-separating hash family is equivalent to a \( w \)-frameproof code. We can use...
Reed-Solomon codes (see, for example, [8]) to construct frameproof codes satisfying $C(N, n, q, \{1, w\}) = \Omega(q^{\frac{w}{w+1}})$ for sufficiently large $q$ (say, $q \geq N$).

We will need some lemmas before presenting the proof of Theorem 4.1. We say a separating hash family is linear if in its representation matrix any two distinct columns can be separated by at least $N-1$ rows, i.e., they have identical component in at most one row. The following corollary is a simple consequence of Theorem 3.4 and Example 3.2.

**Corollary 4.3.** Let $F$ be a linear SHF$(2w; n, q, \{w, w\})$. Then for fixed $w$ and sufficiently large $q$, it holds that $n = o(q^2)$.

**Proof.** Note that the hypergraph $H_F$ defined by the representation matrix of $F$ is a $2w$-uniform $2w$-partite linear hypergraph with equal part size $q$. If $n \geq eq^2$ for some positive constant $\epsilon > 0$, then by Theorem 3.4 $H_F$ must contain a rainbow $2w$-cycle. This can not happen according to the observation illustrated by Example 3.2.

Note that Corollary 4.3 only provides an upper bound for linear separating hash families. In order to obtain an upper bound for general separating hash families, a possible strategy is to argue that every separating hash family must contain a sufficiently large linear subfamily. In this way the upper bound for linear families can induce a general upper bound. However, this may not be true for an arbitrary separating hash family. Fortunately, we find that with the help of the Johnson-type lemma, to deduce a general upper bound it suffices to show that every SHF$(4; n, q, \{2, 2\})$ must contain a sufficiently large linear subfamily.

We say a column $x$ (which can be denoted as $x = (x(1), x(2), x(3), x(4))$) of an $F := \text{SHF}(4; n, q, \{2, 2\})$ contains a special coordinate if there exists some row $i$, $1 \leq i \leq 4$, such that there is at most one column $y \in F \setminus \{x\}$ satisfying $x(i) = y(i)$. A column of $F$ is said to be special if it contains a special coordinate. It can be easily seen that if a column $x$ is not special, then for each $i$, $1 \leq i \leq 4$, there exist two distinct columns $y, z \in F \setminus \{x\}$ satisfying $x(i) = y(i) = z(i)$.

**Lemma 4.4.** Given an arbitrary $F := \text{SHF}(4; n, q, \{2, 2\})$, we can obtain a subfamily $F^* \subseteq F$ which contains no special columns with respect to $F^*$ by deleting at most $8q$ columns of $F$.

**Proof.** We will use a greedy algorithm to construct $F^*$. Delete $x_1$ from $F$ if $x_1$ has a special coordinate in $F$. Denote $F_1 = F \setminus \{x_1\}$. In general, if $x_{i+1} \in F_i$ has a special coordinate in $F_i$, we delete $x_{i+1}$ from $F_i$ and then denote $F_{i+1} = F_i \setminus \{x_{i+1}\}$. Continue the deleting procedure until we get an $F^*$ with no columns containing a special coordinate in it. We say that some symbol $a \in \{1, \ldots, q\}$ in row $i$ is responsible for some column $x$ if $x$ is deleted and it contains $x(i) = a$ as a special coordinate when it is deleted. We claim that in order to obtain $F^*$ at most $8q$ columns will be deleted from $F$.

To see this, just notice that for $1 \leq i \leq 4$, any symbol $a \in \{1, \ldots, q\}$ in row $i$ can be responsible for at most two columns, since otherwise let $x, y$ and $z$ be the three columns for which $a$ is responsible (assume we delete $x$ first), by definition it holds that

$$x(i) = y(i) = z(i) = a$$

and hence $x(i)$ is not a special coordinate of $x$ (note that $y$ and $z$ have not yet been deleted) and it cannot be responsible for $x$. On the other hand, any deleted column must contain at least one symbol which is responsible for it. Therefore, a simple double-counting argument yields that we have deleted at most $8q$ columns.

**Remark 4.5.** The following property is straightforward: If $F^*$ is an SHF$(4; n, q, \{2, 2\})$ which contains no special columns, then for any $i$, $1 \leq i \leq 4$, and any $x \in F$, there exist at least two columns $y, z \in F \setminus \{x\}$ satisfying

$$x(i) = y(i) = z(i).$$

**Lemma 4.6.** Given an arbitrary $F := \text{SHF}(4; n, q, \{2, 2\})$, we can obtain a linear subfamily by deleting at most $8q$ columns of it.
Table 3  No row can separate \{x, w\} and \{y, z\}

|   | x   | y   | z   | w   |
|---|-----|-----|-----|-----|
| V1 | x(1)| y(1)|     |     |
| V2 | x(2)| y(2)|     |     |
| V3 | x(3)|     | z(3)|     |
| V4 |     | y(4)|     | w(4)|

Table 4  No row can separate \{x\} and \{y, z\}

|   | x   | y   | z   |
|---|-----|-----|-----|
| V1 | x(1)|     |     |
| V2 | x(2)| y(2)|     |
| V3 | x(3)|     | z(3)|
| V4 | x(4)| y(4)| z(4)|

Proof. By Lemma 4.4 one can deduce that there exists an \(F^* \subseteq F\) with size at least \(n-8q\) which contains no special columns with respect to \(F^*\). We claim that \(F^*\) must be linear. Assume, to the contrary, that there exist two distinct columns \(x, y \in F^*\) which contain at least two identical components. Without loss of generality, assume that \(x(1) = y(1)\) and \(x(2) = y(2)\). Pick a column \(z \in F^* \setminus \{x, y\}\) such that \(x(3) = z(3)\). Note that the existence of such a \(z\) is guaranteed by Remark 4.5. Now let us consider \(y(4)\). On one hand, if there exists some column \(w \in F^* \setminus \{x, y, z\}\) such that \(w(4) = y(4)\), then obviously no row of \(F^*\) can separate \(\{x, w\}\) and \(\{y, z\}\) (see Table 3 for an illustration of the proof), which leads to a contradiction. On the other hand, if such a \(w\) does not exist, then by Remark 4.5 one can infer that \(x(4) = y(4) = z(4)\), which implies that no row can separate \(\{x\}\) and \(\{y, z\}\) (see Table 4 for an illustration of the proof), which leads to a contradiction, too.

The proof is completed. \(\square\)

**Theorem 4.7.** Let \(F\) be an SHF(4; \(n, q\), \{2, 2\}). Then for sufficiently large \(q\), it holds that \(n = o(q^2)\).

**Proof.** By Lemma 4.6 one can deduce that there exists a linear subfamily \(F^* \subseteq F\) such that \(|F^*| \geq n - 8q\). Corollary 4.3 implies that \(n - 8q = o(q^2)\) and hence the theorem follows trivially. \(\square\)

**Remark 4.8.** It is worth mentioning in [1] that it was shown \(C(4, q, \{2, 2\} + \{1, 1, 1\}) = o(q^2)\), where \(\{2, 2\} + \{1, 1, 1\}\) means that a family is simultaneously \(\{2, 2\}\)-separating and \(\{1, 1, 1\}\)-separating. Theorem 4.7 is stronger than their result in the sense that we show \(\{2, 2\}\)-separating property already guarantees the \(o(q^2)\) magnitude.

The following lemma is a simple consequence of Lemma 1.1 by taking \(N = \sum_{i=1}^t w_i\) and \(l = 1\).

**Lemma 4.9.** Let \(u = \sum_{i=1}^t w_i\). Then it holds that

\[
C(u, q, \{w_1, \ldots, w_t\}) \leq q + (u - 1) + C(u - 1, q, \{w_1 - 1, \ldots, w_t\}).
\]

In fact, in the right-hand side of the inequality we can choose the minus of 1 to be after an arbitrary \(w_i\), \(1 \leq i \leq t\).

Now we are able to prove the main theorem of this section.

**Proof of Theorem 4.1.** Let \(u = \sum_{i=1}^t w_i\). For \(t \geq 3\), by applying Lemma 4.9 repeatedly for \(u - 3\) times one can infer

\[
C(u, q, \{w_1, \ldots, w_t\}) \leq (u - 3)q + (u - 3)(u - 1) + C(3, q, \{1, 1, 1\}),
\]

which implies that \(C(u, q, \{w_1, \ldots, w_t\}) = o(q^2)\) since by (1.1) we have \(C(3, q, \{1, 1, 1\}) = o(q^2)\).

For \(t = 2\) and \(\text{min}\{w_1, w_2\} \geq 2\), by applying Lemma 4.9 repeatedly for \(u - 4\) times one can infer

\[
C(u, q, \{w_1, \ldots, w_t\}) \leq (u - 4)q + (u - 4)(u - 1) + C(4, q, \{2, 2\}),
\]
which implies that \( C(u, q, \{w_1, \ldots, w_t\}) = o(q^2) \) since by Theorem 4.7 we have \( C(4, q, \{2, 2\}) = o(q^2) \). 

One may wonder that whether the upper bound obtained by Theorem 4.1 is tight. The following result collects all the known tight cases of Theorem 4.1.

**Proposition 4.10.** For sufficiently large \( q \), it holds that

\[
q^{2-o(1)} < C(3, q, \{1, 1, 1\}) = o(q^2)
\]

and

\[
q^{2-o(1)} < C(4, q, \{1, 1, 1, 1\}) \leq C(4, q, \{1, 1, 2\}) \leq C(4, q, \{2, 2\}) = o(q^2).
\]

**Proof.** Observe that \( \{1, 1, 1, 1\} \)-separating implies \( \{1, 1, 2\} \)-separating and \( \{1, 1, 2\} \)-separating implies \( \{2, 2\} \)-separating. Thus the proposition is a simple consequence of (1.1) and (1.2). \( \square \)

## 5 An upper bound for \( \text{SHF}(N; n, t, \{w_1, \ldots, w_t\}) \)

In this section, we will present an upper bound for \( \text{SHF}(N; n, t, \{w_1, w_2, \ldots, w_t\}) \) (note that here \( q = t \)). In the proof we will use the asymptotic \( \binom{n}{w} \approx \frac{n^w}{w} \), where \( w \) is the largest value among the \( w_i \)'s. Thus our bound is valid for sufficiently large \( n \) (compared with the \( w_i \)'s). The main result can be stated as the following theorem.

**Theorem 5.1.** Given a positive integer \( t \), let \( w_1, \ldots, w_t \) be \( t \) fixed integers such that \( \min \{w_i \mid 1 \leq i \leq t\} \geq 2 \). Denote \( u = \sum_{i=1}^{t} w_i \). Let \( F \) be an \( \text{SHF}(N; n, t, \{w_1, \ldots, w_t\}) \) such that \( n \) is sufficiently large compared with the \( w_i \)'s. Then the following two statements hold:

(a) Denote \( p^* = \max_{\sum_{i=1}^{t} p_i, 0 \leq p_i \leq 1} \sum_{\pi \in S_t} \prod_{i=1}^{t} \text{Pr}_{\pi(i)}(w_i-1) \) and \( g(q, j + 1) = \frac{q(q-1) \cdots (q-j)}{q^{j+1}} \), where \( S_t \) denotes the symmetric group defined on a finite set of \( t \) elements. Then it holds that

\[
C(N, t, \{w_1, \ldots, w_t\}) \leq C(p^* N, t, \{1, \ldots, 1\}) + u - t
\]

\[
\leq \min_{0 \leq j \leq t-2} (t - j - 1) \left( \frac{q - j}{t - j - 1} \right)^{g(q, j+1)p^*N} + u - t.
\]

(b) For \( w_1 = \cdots = w_t = w \geq 2 \), it holds that

\[
C(N, t, \{w, \ldots, w\}) \leq C\left( \left( \frac{1}{t} \right)^{t(w-1)} N, t, \{1, \ldots, 1\} \right) + u - t
\]

\[
\leq \min \left\{ 2^{\frac{(w^2 - w - 1)}{w-1}} (t - 1) \frac{tQ^N}{t^{w^2 - w - 1}} , (t - 1) \left( \frac{t}{t - 1} \right)^{\frac{Q^N}{w}} \right\} + u - t.
\]

**Proof.** Without loss of generality, set the alphabet set of \( F \) to be \( [t] := \{1, \ldots, t\} \). Denote by \( M \) the representation matrix of \( F \). For each \( i, 1 \leq i \leq t \), let \( p_i \) be the fraction of the symbol \( i \) in \( M \). We can also view \( p_i \) as the probability of a randomly chosen entry of \( M \) being equal to \( i \). It is easy to see that \( \sum_{i=1}^{t} p_i = 1 \) and \( 0 \leq p_i \leq 1 \). We pick randomly and independently \( t \) disjoint subsets \( C'_1, \ldots, C'_t \subseteq F \) such that \( |C'_i| = w_i - 1 \) for \( 1 \leq i \leq t \). One can compute that the probability that a row \( f \) of \( M \) separates \( C'_1, \ldots, C'_t \) is at most

\[
f(p_1, \ldots, p_t) := \frac{\sum_{\pi \in S_t} \prod_{i=1}^{t} (p_{\pi(i)}^{w_i-1})}{\binom{n}{w_1-1} \cdots \binom{n-w_i+1}{w_i-1} \cdots \binom{n-w_1-\cdots-w_{i-1}-t+1}{w_i-1}}.
\]

(5.1)

Denote

\[
p^* := \max_{\sum_{i=1}^{t} p_i, 0 \leq p_i \leq 1} f(p_1, \ldots, p_t).
\]

(5.2)

By linearity of expectation, there exist \( t \) column sets \( C'_1, \ldots, C'_t \subseteq F \) which are separated by at most \( p^* N \) rows of \( M \). Let \( T \) be the collection of rows that separate \( C'_1, \ldots, C'_t \subseteq F \) such that \( |T| \leq p^* N \). For \( t \) distinct columns \( c_1, \ldots, c_t \in F \ \backslash \ (C'_1 \cup C'_2 \cup \cdots \cup C'_t) \), there must exist at least one row \( f \in T \) that
separates $c_1, \ldots, c_t$, since otherwise no row of $M$ will separate $C'_1 \cup \{c_1\}, \ldots, C'_t \cup \{c_t\}$, contradicting the \{w_1, \ldots, w_t\}-separating property of $M$. Therefore, one can conclude that the submatrix formed by rows $T$ and columns $F \setminus (C'_1 \cup C'_2 \cup \cdots \cup C'_t)$ must be a representation matrix of an SHF([T]; n−(u−t); t; \{1, \ldots, 1\}) (there are $t$ 1’s in total). Thus one can infer that

$$n - u + t \leq C([T]; t; \{1, \ldots, 1\}),$$

which implies that

$$n \leq C([T]; t; \{1, \ldots, 1\}) + u - t \leq C(p^* N; t; \{1, \ldots, 1\}) + u - t. \quad (5.3)$$

We can express each term in the summation of (5.1) as follows:

$$\left( \frac{p_{\pi(i)}^n}{w_1 - 1} \right)^n \left( \frac{p_{\pi(2)}^n}{w_2 - 1} \right)^{n - w_1 + 1} \cdots \left( \frac{p_{\pi(t)}^n}{w_t - 1} \right)^{n - w_1 - \cdots - w_{t-1} + t - 1}. \quad (5.4)$$

For sufficiently large $n$ one can infer that (5.4) attains the maximal only if $\min\{p_i | 1 \leq i \leq t\} > a > 0$ holds for some constant $a > 0$. Thus we can set

$$\left( \frac{p_{\pi(i)}^n}{w_i - 1} \right) \approx \left( \frac{p_{\pi(i)}^n}{w_i - 1} \right)^{w_i - 1} \quad \text{for each } i, 1 \leq i \leq t.$$ 

One can compute that $f(p_1, \ldots, p_t)$ approximates

$$\sum_{\pi \in S_t} \prod_{i=1}^{t} p_{\pi(i)}^{w_i - 1}, \quad (5.5)$$

which together with (1.3) and (5.3) complete the first part of the theorem. To prove the second part of the theorem for the special case $w_1 = \cdots = w_t = w$, by taking the logarithm of $\prod_{i=1}^{t} \left( \frac{p_{\pi(i)}^n}{w_i - 1} \right)$ and using the concave property of $\log(\cdot)$ one can infer that $f(p_1, \ldots, p_t)$ attains its maximality $p^*$ if and only if $p_1 = \cdots = p_t = \frac{1}{t}$. For sufficiently large $n$, using (5.5) one can infer that $p^*$ approximates

$$t! \left( \frac{1}{t} \right)^{(w-1)}, \quad (5.6)$$

which together with (1.4) completes the proof of the second part of the theorem.

$\square$

6 A tight upper bound for SHF($N; 2, q, \{1, w\}$)

In this section, we provide a lower bound for the minimal $N$ such that there exists an SHF($N; n, 2, \{1, w\}$) with $n > N$, which can be stated as follows.

**Theorem 6.1.** For all $w \geq 3$ and for all $N < \frac{15 + \sqrt{33}}{24} (w - 2)^2$, it holds that $C(N; q, \{1, w\}) \leq N$. Or equivalently, $N(w) \geq \frac{15 + \sqrt{33}}{24} (w - 2)^2$.

**Remark 6.2.** Indeed we have a tight bound $C(N; q, \{1, w\}) = N$ for $N < \frac{15 + \sqrt{33}}{24} (w - 2)^2$, since an $N \times N$ identity matrix satisfies the $\{1, w\}$-separating property.

We will need a few lemmas to prove Theorem 6.1.

**Lemma 6.3.** Every CFF($N; n, w$) is also an SHF($N; n, 2, \{1, w\}$) and every SHF($N; n, 2, \{1, w\}$) induces a CFF($2N; n, w$).

**Proof.** Denote by $M$ and $M^*$ the representation matrices of an SHF($N; n, 2, \{1, w\}$) and a CFF($N; n, w$), respectively. Given $M^*$, by the $w$-cover-free property it holds that for each column and arbitrary $w$ other columns there exists a row in which the first column is 1 and the remaining $w$ columns are all 0. If we view the columns of $M^*$ as columns of some binary separating hash families, then $M^*$ apparently satisfies the $\{1, w\}$-separating property, which implies that $M^*$ also represents an SHF($N; n, 2, \{1, w\}$).
On the other hand, given $M$, by replacing the 0 entry in $M$ by 10 and the 1 entry by 01, we obtain a $2N \times n$ matrix which is denoted by $M_1$. It suffices to verify that $M_1$ is a representation matrix of a CFF($2N; n, w$). For each column and arbitrary $w$ other columns of $M_1$, let us consider the corresponding columns of $M$. By the $\{1, w\}$-separating property, for these $w + 1$ columns there is a row of $M$ having the configuration $10 \cdots 0$ or $01 \cdots 1$, which is translated to

\[
\begin{pmatrix}
0 & 1 & \cdots & 1 \\
1 & 0 & \cdots & 0
\end{pmatrix}
\quad \text{or} \quad
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 1
\end{pmatrix}
\]

in $M_1$. Note that the second row of the first submatrix and the first row of the second submatrix satisfy the $w$-cover-free property. If we view $M_1$ as a representation matrix for some $F \subseteq 2^X$, where $|X| = 2N$, then by the above discussions we can conclude that $F$ is $w$-cover-free.

**Lemma 6.4.** Denote by $N^*(w)$ the minimal $N$ such that there exists a CFF($N; n, w$) with $n > N$. In addition, denote by $N(w)$ the minimal $N$ such that there exists an SHF($N; n, 2, \{1, w\}$) with $n > N$. Then for $w \geq 3$, it holds that $N^*(w - 2) \leq N(w) \leq N^*(w)$.

**Proof.** Denote by $M$ the representation matrix of an SHF($N; n, 2, \{1, w\}$) with $N = N(w)$. Then we have $n > N$ by the definition of $N(w)$. First of all, the upper bound in the inequality follows from the fact that every CFF($N; n, w$) is also an SHF($N; n, 2, \{1, w\}$), which is shown in Lemma 6.3. It remains to prove the lower bound. Replace the 0 entry in $M$ by 10 and the 1 entry by 01. We obtain a $2N \times n$ matrix with constant column weight $N$. Denote this new matrix by $M_1$. By Lemma 6.3, $M_1$ is the representation matrix of a CFF($2N; n, w$).

By Lemma 2.2, deleting from $M_1$ an arbitrary column and the rows containing a 1 in it leads to a new matrix $M_2$, which is the representation matrix of a CFF($N; n - 1, w - 1$).

We claim that there must exist a column in $M_2$ of weight at least two. Denote by $c$ the column deleted from $M_1$. If some column $c' \in M_2$ is of weight one, then one can verify that $c$ and $c'$ have exactly $N - 1$ identical components in $M$. If $M_2$ contains two columns of weight 1, then in $M$ there are two distinct columns that have exactly $N - 1$ identical components with $c$. Then it is not hard to show that no row of $M$ can separate $c$ and these two columns. Therefore, $M_2$ contains at most one column of weight 1. The claim follows from the simple fact that $n - 1 \geq N = N(w) \geq N(3) > 6$, where the last inequality follows from (1.5).

Take an arbitrary column of $M_2$ with weight at least two. Delete from $M_2$ this column and the rows containing a 1 in it. Again, by Lemma 2.2, the new matrix is the representation matrix of a CFF($N; n - 2, w - 2$) satisfying $N' \leq N - 2 < n - 2$ since we have assumed that $n > N$. Thus one can deduce that $N' \geq N^*(w - 2)$ and hence the lower bound $N(w) \geq N^*(w - 2)$ follows immediately.

**Proof of Theorem 6.1.** Theorem 6.1 is a direct consequence of Lemmas 2.4 and 6.4.

**7 Concluding remarks**

This paper provides several new upper bounds for separating hash families. The following remaining open problems seem to be interesting.

Theorem 4.10 implies that Theorem 4.1 is tight for several small parameters. For the general situation, the following conjecture seems reasonable.

**Conjecture 7.1.** Let $w_1, \ldots, w_t$ be fixed integers such that $\{w_1, \ldots, w_t\} \neq \{1, w\}$ for any $w$. Then for sufficiently large $q$ we have $C(\sum_{i=1}^t w_i, q, \{w_1, \ldots, w_t\}) > q^{2-o(1)}$.

Since $u$-perfect hash family satisfies $\{w_1, \ldots, w_t\}$-separating property for arbitrary $\sum_{i=1}^t w_i = u$, to verify Conjecture 7.1, it suffices to show that $C(t, q, \{1, \ldots, 1\}) > q^{2-o(1)}$ (there are $t$ 1’s in total) for sufficiently large $q$ and every fixed integer $t$. Lemma 6.1 of [26] shows that $t$-uniform $t$-partite linear hypergraphs which contain no rainbow cycles of length $k$ for any $3 \leq k \leq t$ are good candidates for $t$-perfect hash families. Thus to prove Conjecture 7.1 it suffices to construct sufficiently large hypergraphs.
containing no rainbow cycles. Unfortunately, this is not a simple task. Some constructions of this type using additive number theory can be found in [15].

For the upper bound of $\text{SHF}(N, n, q, \{w_1, \ldots, w_t\})$ under the situation $(u - 1) \nmid N$, it may be a difficult problem to determine whether $n = \Theta(q^{\frac{u}{u-1}})$ or $n = o(q^{\frac{N}{u-1}})$ for general $w_1, \ldots, w_t$. We have a conjecture for the case $w_1 = \cdots = w_t = 1$, which is stated as follows.

**Conjecture 7.2.** Let $t \geq 3$ be a positive integer. For $w_1 = \cdots = w_t = 1$ and $(t - 1) \nmid N$, it holds that $C(N, q, \{1, \ldots, 1\}) = o(q^{\frac{N}{u-1}})$.

Note that Theorem 4.1 implies that Conjecture 7.2 is true for $N = t \geq 3$. One possible strategy to attack this conjecture is to use the hypergraph removal lemma (see, for example, [11]), which is a generalization of the graph removal lemma. The authors have attempted to use this tool to prove $C(5, q, \{1, 1, 1\}) = o(q^3)$ for sufficiently large $q$. Unfortunately, we failed to accomplish this since the structure of the hypergraphs needed to be considered is too complicated (indeed to prove Theorem 4.1 it suffices to consider linear hypergraphs).

We also have a conjecture for the approximate magnitude of $N(w)$.

**Conjecture 7.3.** It holds that $\lim_{N \to \infty} \frac{N(w)}{w^2} = 1$.

Note that Lemma 6.4 implies that this conjecture is equivalent to the weaker form of Conjecture 2.3.

**Acknowledgements** The first author was supported by National Natural Science Foundation of China (Grant Nos. 11431003 and 61571310), Beijing Scholars Program, Beijing Hundreds of Leading Talents Training Project of Science and Technology, and Beijing Municipal Natural Science Foundation. The third author was supported by the Post-Doctoral Science Foundation of China (Grant No. 2018M632356) and National Natural Science Foundation of China (Grant No. 11801392). The authors are grateful to the anonymous referees for their careful reading of the manuscript and many helpful suggestions.

**References**

1. Alon N, Eldar F, Mario S. Parent-identifying codes. J Combin Theory Ser A, 2001, 95: 349–359
2. Alon N, Moni N. Derandomization, witnesses for Boolean matrix multiplication and construction of perfect hash functions. Algorithmica, 1996, 16: 434–449
3. Alon N, Stav U. New bounds on parent-identifying codes: The case of multiple parents. Combin Probab Comput, 2004, 13: 795–807
4. Barg A, Cohen G, Encheva S, et al. A hypergraph approach to the identifying parent property: The case of multiple parents. SIAM J Discrete Math, 2001, 14: 423–431
5. Bazrafshan M, Trung T. Bounds for separating hash families. J Combin Theory Ser A, 2011, 118: 1129–1135
6. Bazrafshan M, Trung T. Improved bounds for separating hash families. Des Codes Cryptogr, 2013, 69: 369–382
7. Blackburn S R. Perfect hash families: Probabilistic methods and explicit constructions. J Combin Theory Ser A, 2000, 92: 54–60
8. Blackburn S R. Frameproof codes. SIAM J Discrete Math, 2003, 16: 499–510
9. Blackburn S R, Etzion T, Stinson D R, et al. A bound on the size of separating hash families. J Combin Theory Ser A, 2008, 115: 1246–1256
10. Chor B, Fiat A, Naor M. Tracing traitors. In: CRYPTO’94 Proceedings of the 14th Annual International Cryptology Conference on Advances in Cryptology. London: Springer-Verlag, 1994, 257–270
11. Conlon D, Fox J. Graph removal lemmas. In: Surveys in Combinatorics. London Mathematical Society Lecture Note Series, vol. 409. Cambridge: Cambridge University Press, 2013, 1–49
12. Dyachkov A G, Vorobyev I V, Polyanskii N A, et al. Cover-free codes and separating system codes. Des Codes Cryptogr, 2017, 82: 197–209
13. Erdős P, Frankl P, Füredi Z. Families of finite sets in which no set is covered by the union of $r$ others. Israel J Math, 1985, 51: 79–89
14. Fuji-Hara R. Perfect hash families of strength three with three rows from varieties on finite projective geometries. Des Codes Cryptogr, 2015, 77: 351–356
15. Ge G, Shangguan C. Sparse hypergraphs: New bounds and constructions. ArXiv:1706.03306, 2017
16. Guo C, Stinson D R, van Trung T. On tight bounds for binary frameproof codes. Des Codes Cryptogr, 2015, 77: 301–319
17 Hollmann D L, van Lint J H, Linnartz J-P, et al. On codes with the identifiable parent property. J Combin Theory Ser A, 1998, 82: 121–133
18 Körner J, Marton K. New bounds for perfect hashing via information theory. European J Combin, 1988, 9: 523–530
19 Mehlhorn K. Data Structure and Algorithms 1. EATCS Monographs on Theoretical Computer Science. Berlin: Springer-Verlag, 1984
20 Newman I, Wgderston A. Lower bounds on formula size of Boolean functions using hypergraph entropy. SIAM J Discrete Math, 1995, 8: 536–542
21 Nilli A. Perfect hashing and probability. Combin Probab Comput, 1994, 3: 407–409
22 Niu X, Cao H. Constructions and bounds for separating hash families. Discrete Math, 2018, 341: 2627–2638
23 Ruszinkó M. On the upper bound of the size of the r-cover-free families. J Combin Theory Ser A, 1994, 66: 302–310
24 Sagalovich Y L. Separating systems. Problemy Peredachi Informatsii, 1994, 30: 14–35
25 Shangguan C, Ge G. New bounds on the number of tests for disjunct matrices. IEEE Trans Inform Theory, 2016, 62: 7518–7521
26 Shangguan C, Ge G. Separating hash families: A Johnson-type bound and new constructions. SIAM J Discrete Math, 2016, 30: 2243–2264
27 Shangguan C, Ma J, Ge G. New upper bounds for parent-identifying codes and traceability codes. Des Codes Cryptogr, 2018, 86: 1727–1737
28 Shangguan C, Wang X, Ge G, et al. New bounds for frameproof codes. IEEE Trans Inform Theory, 2017, 63: 7247–7252
29 Staddon J N, Stinson D R, Wei R. Combinatorial properties of frameproof and traceability codes. IEEE Trans Inform Theory, 2001, 47: 1042–1049
30 Stinson D R, Trung T, Wei R. Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. J Statist Plann Inference, 2000, 86: 595–617
31 Stinson D R, Wei R, Chen K. On generalized separating hash families. J Combin Theory Ser A, 2008, 115: 105–120
32 Stinson D R, Wei R, Zhu L. New constructions for perfect hash families and related structures using combinatorial designs and codes. J Combin Des, 2000, 8: 189–200
33 Stinson D R, Zaverucha G M. Some improved bounds for secure frameproof codes and related separating hash families. IEEE Trans Inform Theory, 2008, 54: 2508–2514
34 Walker II R A, Colbourn C J. Perfect hash families: Constructions and existence. J Math Cryptol, 2007, 1: 125–150