Des. Codes Cryptogr.
https://doi.org/10.1007/s10623-018-0468-3

Multiset combinatorial batch codes

Hui Zhang[1] · Eitan Yaakobi[1] · Natalia Silberstein[2]

Received: 11 January 2017 / Revised: 7 February 2018 / Accepted: 8 February 2018

Abstract *Batch codes*, first introduced by Ishai, Kushilevitz, Ostrovsky, and Sahai, mimic a distributed storage of a set of n data items on m servers, in such a way that any batch of k data items can be retrieved by reading at most some t symbols from each server. *Combinatorial batch codes*, are replication-based batch codes in which each server stores a subset of the data items. In this paper, we propose a generalization of combinatorial batch codes, called *multiset combinatorial batch codes* (*MCBC*), in which n data items are stored in m servers, such that any multiset request of k items, where any item is requested at most r times, can be retrieved by reading at most t items from each server. The setup of this new family of codes is motivated by recent work on codes which enable high availability and parallel reads in distributed storage systems. The main problem under this paradigm is to minimize the number of items stored in the servers, given the values of n, m, k, r, t, which is denoted by $N(n, k, m, t; r)$. We first give a necessary and sufficient condition for the existence of MCBCs. Then, we present several bounds on $N(n, k, m, t; r)$ and constructions of MCBCs. In particular, we determine the value of $N(n, k, m, 1; r)$ for any $n \geq \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - (m-k+1)A(m, 4, k-2)$, where $A(m, 4, k-2)$ is the maximum size of a binary constant weight code of length m, distance four and weight $k-2$. We also determine the exact value of $N(n, k, m, 1; r)$ when $r \in \{k, k-1\}$ or $k = m$.

Communicated by M. Paterson.

Parts of the presented work [22] were published in the proceedings of the IEEE International Symposium on Information Theory (ISIT) 2017.

✉ Hui Zhang
huizhang@cs.technion.ac.il

Eitan Yaakobi
yaakobi@cs.technion.ac.il

Natalia Silberstein
natalys@cs.technion.ac.il

1 Computer Science Department, Technion—Israel Institute of Technology, Haifa, Israel

2 Yahoo! Research, Haifa, Israel

Published online: 23 February 2018

Keywords Batch code · Dual set system · Hall's condition

Mathematics Subject Classification 05C65 · 05D05 · 68P20 · 68R05

1 Introduction

1.1 Background and definitions

Batch codes were first introduced by Ishai et al. [13] as a method to represent the distributed storage of a set of n data items on m servers. These codes were originally motivated by several applications such as load balancing in distributed storage, private information retrieval, and cryptographic protocols. Formally, these codes are defined as follows [13].

Definition 1
1. An (n, N, k, m, t) *batch code* over an alphabet Σ, encodes a string $x \in \Sigma^n$ into an m-tuple of strings $y_1, \ldots, y_m \in \Sigma^*$ (called *buckets* or *servers*) of total length N, such that for each k-tuple (called *batch* or *request*) of distinct indices $i_1, \ldots, i_k \in [n]$, the k data items $x_{i_1}, \ldots, x_{i_k}$ can be decoded by reading at most t symbols from each server.
2. An (n, N, k, m, t) *multiset batch code* is an (n, N, k, m, t) batch code which also satisfies the following property: for any multiset request of k indices $i_1, \ldots, i_k \in [n]$ there is a partition of the buckets into k subsets $S_1, \ldots, S_k \subseteq [m]$ such that each item x_{i_j}, $j \in [k]$, can be retrieved by reading at most t symbols from each bucket in S_j.

Yet another class of codes, called *combinatorial batch codes* (CBC), is a special type of batch codes in which all encoded symbols are copies of the input items, i.e., these codes are replication-based. Several works have considered codes under this setup; see e.g. [2,3,5–10,17,19,20]. However, note that combinatorial batch codes are not multiset batch codes and don't allow to request an item more than once.

Motivated by the works on codes which enable parallel reads for different users in distributed storage systems, for example, codes with locality and availability [18,21], we introduce a generalization of CBCs, named *multiset combinatorial batch codes*.

Definition 2 An $(n, N, k, m, t; r)$ *multiset combinatorial batch code (MCBC)* is a collection of subsets of $[n]$, $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ (called *servers*) where $N = \sum_{j=1}^{m} |C_j|$, such that for each multiset request $\{i_1, i_2, \ldots, i_k\}$, in which every element in $[n]$ has multiplicity at most r, there exist subsets $D_1, \ldots, D_m$, where for all $j \in [m]$, $D_j \subseteq C_j$ with $|D_j| \leq t$, and the multiset union[1] of D_j for $j \in [m]$ contains the multiset request $\{i_1, i_2, \ldots, i_k\}$.

In other words, an $(n, N, k, m, t; r)$-MCBC is a coding scheme which encodes n items into m servers, with total storage of N items, such that any multiset request of items of size at most k, where any item can be repeated at most r times, can be retrieved by reading at most t items from each server. In particular, when $r = 1$ we obtain a combinatorial batch code, and when $r = k$ and $t = 1$ we obtain a multiset batch code based on replication. Note that the constraint t = 1 guarantees that the servers retrieving different items in a multiset request are distinct.

Example 1 Let us consider the following $(n = 5, N = 15, k = 5, m = 5, t = 1; r = 2)$ MCBC,

[1] For any $i \in [n]$, the multiplicity of i in the multiset union of the sets D_j for $j \in [m]$ is the number of subsets that contain i, that is $|\{j \in [m] : i \in D_j\}|$.

1	1	2	2	3
3	4	3	4	4
5	5	5	5	5

where the i-th column contains the indices of items stored in the server $C_i \in \mathcal{C}$, $i \in [5]$. It is possible to verify that the code $\mathcal{C}$ satisfies the requirements of a (5, 15, 5, 5, 1; 2)-MCBC. For example, the multiset request $\{3, 3, 4, 4, 5\}$ can be read by taking the subsets $D_1 = \{3\}$, $D_2 = \{4\}$, $D_3 = \{3\}$, $D_4 = \{4\}$, $D_5 = \{5\}$.

Similarly to the original problem of combinatorial batch codes, the goal in this paper is to minimize the total storage N given the parameters n, m, k, t and r of an MCBC. Let $N(n, k, m, t; r)$ be the smallest N such that an $(n, N, k, m, t; r)$-MCBC exists. An MCBC is called *optimal* if N is minimal given n, m, k, t, r. In this paper, we focus on the case $t = 1$, and thus omit t from the notation and write it as an $(n, N, k, m; r)$-MCBC and its minimum storage by $N(n, k, m; r)$. In case $r = 1$, i.e. an MCBC is a CBC, we further omit r and write it as an (n, N, k, m)-CBC and its minimum storage as $N(n, k, m)$.

In [17], the authors studied another class of CBCs, called *uniform combinatorial batch codes* (*uniform CBCs*), in which each item is stored in the same number of servers. Several constructions of optimal uniform CBCs were given in [2,3,17,20]. In this paper we consider a slightly different class of MCBCs, in which each server stores the same number of items, and call these codes *regular multiset combinatorial batch codes* (*regular MCBCs*).

A regular $(n, N, k, m; r)$-MCBC is an MCBC in which each server stores the same number μ of items, where $\mu = N/m$. Given n, m, k, r, let $\mu(n, k, m; r)$ denote the smallest number of items stored in each server, then the optimal value of N is determined by $\mu(n, k, m; r)$, that is, $N = m\mu(n, k, m; r)$. Note that the (5, 15, 5, 5; 2)-MCBC in Example 1 is regular. Removing "5" from the last server, we get an irregular (5, 14, 4, 5; 2)-MCBC, because the request $\{1, 1, 2, 2, 5\}$ can not be retrieved.

1.2 Previous results on CBCs

For CBCs, a significant amount of work has been done to study the value $N(n, k, m)$, and the exact value has been determined for a large range of parameters. We list some of the known results below (for more details see [3,5,7,8,17,20]).

Theorem 1 *For CBCs,*

(i) $N(n, k, k) = kn - k(k-1)$.
(ii) *If* $n \geq (k-1)\binom{m}{k-1}$, *then* $N(n, k, m) = kn - (k-1)\binom{m}{k-1}$.
(iii) *If* $\binom{m}{k-2} \leq n \leq (k-1)\binom{m}{k-1}$, *then* $N(n, k, m) = (k-1)n - \left\lfloor \frac{(k-1)\binom{m}{k-1}-n}{m-k+1} \right\rfloor$.
(iv) *If* $\binom{m}{k-2} - (m-k+1)A(m, 4, k-3) \leq n \leq \binom{m}{k-2}$, *then* $N(n, k, m) = (k-2)n - \left\lfloor \frac{2(\binom{m}{k-2}-n)}{m-k+1} \right\rfloor$ *for* $0 \leq (\binom{m}{k-2} - n) \mod (m-k+1) < \frac{m-k+1}{2}$.
(v) $N(m+1, k, m) = m + k$.
(vi) *Let* k *and* m *be integers with* $2 \leq k \leq m$, *then*

$$N(m+2, k, m) = \begin{cases} m+k-2+\lceil 2\sqrt{k+1}\rceil & \text{if } m+1-k \geq \lceil \sqrt{k+1}\rceil, \\ 2m-2+\left\lceil 1+\frac{k+1}{m+1-k}\right\rceil & \text{if } m+1-k < \lceil \sqrt{k+1}\rceil. \end{cases}$$

(vii) *For all integers* $n \geq m \geq 3$, $N(n,3,m) = \begin{cases} 2n - m + \left\lfloor \frac{n-3}{m-2} \right\rfloor & \text{if } n \leq m^2 - m, \\ 3n - m^2 + m & \text{if } n \geq m^2 - m. \end{cases}$

(viii) *For all integers* $n \geq m \geq 4$,

$$N(n,4,m) = \begin{cases} n & \text{if } n = m, \\ 2n - m + \left\lceil \frac{1+\sqrt{8n-8m+1}}{2} \right\rceil & \text{if } m < n \leq \frac{m^2+6m}{8} \text{ and } m \text{ is even} \\ & \text{or if } m < n \leq \frac{m^2+4m+3}{8} \text{ and } m \text{ is odd,} \\ 2n - m + \left\lceil \frac{5+\sqrt{8n-16m+25}}{2} \right\rceil & \text{if } \frac{m^2+6m+8}{8} \leq n < \binom{m}{2} \text{ and } m \text{ is even} \\ & \text{or if } \frac{m^2+4m+11}{8} < n < \binom{m}{2} \text{ and } m \text{ is odd,} \\ 2n - \frac{m-1}{2} & \text{if } n = \frac{m^2+4m+11}{8} \text{ and } m \text{ is odd,} \\ 3n - \left\lfloor \frac{m^2}{2} - \frac{n-m}{m-3} \right\rfloor & \text{if } \binom{m}{2} \leq n < 3\binom{m}{3}, \\ 4n - 3\binom{m}{3} & \text{if } 3\binom{m}{3} \leq n. \end{cases}$$

(ix) *For any prime power* $q \geq 3$, $N(q^2+q-1, q^2-q-1, q^2-q) = q^3 - q$.

1.3 Our contributions

From the definition of MCBCs, one can observe that $r \leq k \leq tm$ and $n \leq N$. If $m \geq nr$, the trivial construction where each server stores a single item is optimal. Since we only consider the case $t = 1$, we always assume $r \leq k \leq m < nr$ throughout the paper.

In this paper, we study the properties of MCBCs, and give a necessary and sufficient condition for the existence of MCBCs. Using this condition, we are able to obtain the following bounds on the value of $N(n,k,m;r)$.

Theorem 2 *For MCBCs,*

(i) $N(n,k,m;r) \geq rn$.
(ii) $N(n,k,m;r) \geq N(n,k,m;i)$ *for* $i \in [r-1]$.
(iii) $\frac{1}{r}N(rn,k,m) \leq N(n,k,m;r) \leq N(rn,k,m)$.
(iv) $N(n,k,m;r) \leq rN(n, \lceil \frac{k}{r} \rceil, \lfloor \frac{m}{r} \rfloor)$.
(v) *Let* $r \leq k-1$. *For any* $c \in [r, k-1]$, $N(n,k,m;r) \geq nc - \left\lfloor \frac{k-c}{m-k+1} \left\lceil \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}}{\binom{m-c}{k-1-c}} - n \right\rceil \right\rfloor$.

We also provide several constructions of $(n,N,k,m;r)$-MCBC and determine the exact value of $N(n,k,m;r)$ for some specific parameters.

Theorem 3 *For MCBCs,*

(i) *If* $n \geq \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$, *then* $N(n,k,m;r) = kn - \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$.
(ii) $N(n,k,m;k) = kn$, $N(n,k,m;k-1) = \begin{cases} kn - \binom{m}{k-1} & \text{if } n \geq \binom{m}{k-1}, \\ (k-1)n & \text{if } n < \binom{m}{k-1}. \end{cases}$
(iii) *If* $\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - (m-k+1)A(m,4,k-2) \leq n \leq \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$ *and* $r \leq k-2$, *then*
$N(n,k,m;r) = (k-1)n - \left\lfloor \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - n}{m-k+1} \right\rfloor$.
(iv) $N(n,k,k;r) = kn - \lfloor \frac{k-1}{r} \rfloor k$ *if* $r \mid k$, $n \geq \frac{k}{r}$ *or* $r \nmid k$, $n \geq \lfloor \frac{k}{r} \rfloor + r$.
(v) *For any prime power* q, $N(q^2+q, k, q^2; r) \leq q^3 + q^2$, *where* (k,r) *satisfies* $\lfloor \frac{q}{2} \rfloor + 1 \leq r \leq q$, $k \leq (q-r+1)(2r-1)$ *or* $r = 1$, $k \leq q^2$. *Especially, when* $(k,r) \in \{(q^2,1), (2q-1,q)\}$, $N(q^2+q,k,q^2;r) = q^3+q^2$.

By Theorem 3, we can see: when $r \geq k-1$ for all n, or when $r \leq k-2$ and $n \geq \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - (m-k+1)A(m, 4, k-2)$, the exact value of $N(n, k, m; r)$ is determined.

For a regular $(n, N, k, m; k)$-MCBC, every item has to be stored in at least k different servers and so $\mu(n, k, m; k) \geq kn/m$. Our contribution in this part is finding a necessary and sufficient condition for equality in the last inequality. This result is summarized in the following theorem.

Theorem 4 *For regular MCBCs,* $\mu(n, k, m; k) = \frac{kn}{m}$ *if and only if* $n = c \cdot \frac{m}{\gcd(m,k)}$ *for some integer* $c \geq 0$.

The rest of the paper is organized as follows. In Sect. 2, we give a necessary and sufficient condition for the existence of MCBCs. In Sects. 3 and 4, we give several bounds and constructions for MCBCs, and establish the results of $N(n, k, m; r)$ in Theorems 2 and 3. In Sect. 5, we analyze regular MCBCs, and determine the value of $\mu(n, k, m; k)$ in Theorem 4.

2 Set systems and the multiset Hall's condition

A *set system* is a pair $(V, \mathcal{C})$, where V is a finite set of *points* and $\mathcal{C}$ is a collection of subsets of V (called *blocks*). Given a set system $(V, \mathcal{C})$ with a point set $V = \{v_1, v_2, \ldots, v_n\}$ and a blocks set $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$, its *incidence matrix* is an $m \times n$ matrix M, given by

$$M_{i,j} = \begin{cases} 1 & \text{if } v_j \in C_i, \\ 0 & \text{if } v_j \notin C_i. \end{cases}$$

If M is the incidence matrix of the set system $(V, \mathcal{C})$, then the set system having incidence matrix $M^\top$ is called the *dual set system* of $(V, \mathcal{C})$.

Let $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ be an $(n, N, k, m; r)$-MCBC. Similarly to the study of CBCs, by setting $V = [n]$, we consider the set system $(V, \mathcal{C})$ of the MCBC. In addition, we denote the set system $(X, \mathcal{B})$ which is given by $X = [m]$ and $\mathcal{B} = \{B_1, B_2, \ldots, B_n\}$ where for each $i \in [n]$, $B_i \subseteq X$ consists of the servers that store the i-th item. Then, it is readily verified that $(X, \mathcal{B})$ is the dual set system of $(V, \mathcal{C})$. We note that a set system $(V, \mathcal{C})$ of this form or its dual set system $(X, \mathcal{B})$ uniquely determines an MCBC and thus in the rest of the paper we will usually refer to an MCBC by its set system or its dual set system.

Example 2 The following is a (20, 80, 16, 16)-CBC given in [20] based on an *affine plane* of order 4. Here, $V = [20]$, each column contains the indices of items stored in a server $C_i \in \mathcal{C}$ and also forms a block of the set system $(V, \mathcal{C})$.

1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
5	6	7	8	6	5	8	7	7	8	5	6	8	7	6	5
9	10	11	12	12	11	10	9	10	9	12	11	11	12	9	10
13	14	15	16	15	16	13	14	16	15	14	13	14	13	16	15
17	17	17	17	18	18	18	18	19	19	19	19	20	20	20	20

Here, the dual set system of $(V, \mathcal{C})$ is $(X, \mathcal{B})$ in which $X = [16]$ and $\mathcal{B} = \{\{1, 5, 9, 13\}, \{2, 6, 10, 14\}, \{3, 7, 11, 15\}, \{4, 8, 12, 16\}, \{1, 6, 11, 16\}, \{2, 5, 12, 15\}, \{3, 8, 9, 14\}, \{4, 7, 10, 13\}, \{1, 8, 10, 15\}, \{2, 7, 9, 16\}, \{3, 6, 12, 13\}, \{4, 5, 11, 14\}, \{1, 7, 12, 14\}, \{2, 8, 11, 13\}, \{3, 5, 10, 16\}, \{4, 6, 9, 15\}, \{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{9, 10, 11, 12\}, \{13, 14, 15, 16\}\}$.

In the rest of this section we let $(V, \mathcal{C})$ with $V = [n]$ and $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ be a set system, and $(X, \mathcal{B})$ with $X = [m]$ and $\mathcal{B} = \{B_1, B_2, \ldots, B_n\}$ be its dual set system. The following theorem states a necessary and sufficient condition on the dual set system to form a construction of CBCs.

Theorem 5 [17] *The set system $(V, \mathcal{C})$ is an (n, N, k, m)-CBC if and only if its dual set system $(X, \mathcal{B})$ satisfies the following* Hall's condition*: for all $h \in [k]$, and any h distinct blocks $B_{i_1}, B_{i_2}, \ldots, B_{i_h} \in \mathcal{B}$, $|\cup_{j=1}^{h} B_{i_j}| \geq h$.*

The Hall's condition was generalized in several ways, see e.g. [1,15,16]. For example, in [9], the authors explored the value of $N(n, k, m, t)$ for $t > 1$ with a generalization of the Hall's condition, that is, the set system $(V, \mathcal{C})$ is an (n, N, k, m, t)-CBC if and only if its dual set system satisfies the (k, t)*-Hall's condition: for all $h \in [k]$, and any h distinct blocks* $B_{i_1}, B_{i_2}, \ldots, B_{i_h} \in \mathcal{B}$, $|\cup_{j=1}^{h} B_{i_j}| \geq h/t$. In this paper, we present another generalization of the Hall's condition, named the *multiset Hall's condition*, and provide a necessary and sufficient condition for the construction of MCBCs.

Given a set system $(V, \mathcal{C})$, we construct a new set system $(U, \mathcal{F})$ with $U = [rn]$ and $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ where for $i \in [m]$, $F_i = \{c + jn : c \in C_i, j \in [0, r-1]\}$. Let $U^{(\ell)} = \{\ell + jn : j \in [0, r-1]\}$ for $\ell \in [n]$. We first show the following claim.

Claim 1 *$(V, \mathcal{C})$ is an $(n, N, k, m; r)$-MCBC if and only if $(U, \mathcal{F})$ is an (rn, rN, k, m)-CBC.*

Proof Suppose that $(V, \mathcal{C})$ is an $(n, N, k, m; r)$-MCBC. For any request $P = \{i_1, i_2, \ldots, i_k\} \subseteq [rn]$ of $(U, \mathcal{F})$, let r_ℓ, for $\ell \in [n]$, be the number of elements requested from the set $U^{(\ell)}$, and note that $0 \leq r_\ell \leq r$ and $\sum_{\ell=1}^{n} r_\ell = k$. Consider the multiset request Q of $(V, \mathcal{C})$ where each $\ell \in [n]$ appears r_ℓ times. Since $(V, \mathcal{C})$ is an $(n, N, k, m; r)$-MCBC, Q can be read by choosing subsets $D_j \subseteq C_j$, $|D_j| \leq 1$ for $j \in [m]$. Then P can also be read from the set of servers $\{F_j : j \in [m], |D_j| = 1\}$. Therefore, $(U, \mathcal{F})$ is an (rn, rN, k, m)-CBC.[2]

The reverse is similar. Suppose that $(U, \mathcal{F})$ is an (rn, rN, k, m)-CBC. For any multiset request Q of $(V, \mathcal{C})$ where each $\ell \in [n]$ appears r_ℓ times, consider the request P of $(U, \mathcal{F})$ which contain any r_ℓ distinct elements in $U^{(\ell)}$. Since $(U, \mathcal{F})$ is an (rn, rN, k, m)-CBC, P can be read by taking $D_j \subseteq F_j$, $|D_j| \leq 1$ for $j \in [m]$. Then Q can be read from the servers $\{C_j : j \in [m], |D_j| = 1\}$. Therefore, $(V, \mathcal{C})$ is an $(n, N, k, m; r)$-MCBC. □

Now, we are in a position to give the *multiset Hall's condition* for MCBCs.

Theorem 6 *The set system $(V, \mathcal{C})$ is an $(n, N, k, m; r)$-MCBC if and only if its dual set system $(X, \mathcal{B})$ satisfies the following* multiset Hall's condition*: for all $h \in \left[\lceil \frac{k}{r} \rceil\right]$, and any h distinct blocks $B_{i_1}, B_{i_2}, \ldots, B_{i_h} \in \mathcal{B}$, $|\cup_{j=1}^{h} B_{i_j}| \geq \min\{hr, k\}$.*

Proof ($\Rightarrow$) Assume that $(V, \mathcal{C})$ is an $(n, N, k, m; r)$-MCBC, and let $i_1, i_2, \ldots, i_h \in V$ for some $h \in \left[\lceil \frac{k}{r} \rceil\right]$ be the indices of some h different items. Then, the set $\cup_{j \in [h]} B_{i_j}$ corresponds to the indices of all the servers that contain these items.

If $h \leq \lfloor \frac{k}{r} \rfloor$, let us consider the multiset request $\{i_1, \ldots, i_1, i_2, \ldots, i_2, \ldots, i_h, \ldots, i_h\}$ where each of the h elements is requested r times. Since it is possible to read from each server at most one item, the number of servers that contain these h items has to be at least hr, that is $|\cup_{j \in [h]} B_{i_j}| \geq hr$. Similarly, if $h = \lceil \frac{k}{r} \rceil$, then we need k servers for the multiset request of size k on $i_1, i_2, \ldots, i_h$ where each i_j, for $j \in [h]$, is requested at most r times, and so $|\cup_{j \in [h]} B_{i_j}| \geq k$. Together we conclude that $|\cup_{j=1}^{h} B_{i_j}| \geq \min\{hr, k\}$.

[2] We notice that this direction is not needed in the proof. But we still prove it here because we will use it in Lemma 7 below.

($\Leftarrow$) Suppose $(U, \mathcal{F})$ is defined as above. Let $(X = [m], \mathcal{G} = \{G_1, \dots, G_{nr}\})$, be the dual set system of $(U, \mathcal{F})$, so G_i, for $i \in [rn]$, is the set of servers that contain the i-th item in $(U, \mathcal{F})$, and the block G_i for all $i \in U^{(\ell)}$ are the same as B_ℓ for $\ell \in [n]$. We show that $(X, \mathcal{G})$ satisfies the Hall's condition. For any $i_1, i_2, \dots, i_h \in [rn]$, $h \in [k]$, let r_ℓ denote the number of elements in $U^{(\ell)}$ for $\ell \in [n]$. Then $\left|\cup_{j=1}^{h} G_{i_j}\right| = \left|\cup_{\ell:r_\ell \neq 0} B_\ell\right|$.

Let $a = |\{\ell : r_\ell \neq 0\}|$. By the multiset Hall's condition, when $a \leq \lceil \frac{k}{r} \rceil$, $\left|\cup_{\ell:r_\ell \neq 0} B_\ell\right| \geq \min\{ar, k\}$; when $\lceil \frac{k}{r} \rceil < a \leq k$,

$$\left|\bigcup_{\ell:r_\ell \neq 0} B_\ell\right| \geq \min\left\{r\left\lceil\frac{k}{r}\right\rceil, k\right\} \geq k = \min\{ar, k\}.$$

Note that the first inequality is due to the fact that the union of a sets is at least size of the union of $\lceil k/r \rceil$, since $a > \lceil k/r \rceil$.

Since $h = \sum_{\ell:r_\ell \neq 0} r_\ell \leq ar$ and $h \leq k$, we always have $| \cup_{j=1}^{h} G_{i_j}| \geq h$ for any $h \in [k]$, that is $(X, \mathcal{G})$ satisfies the Hall's condition. Hence, $(U, \mathcal{F})$ is an (rn, rN, k, m)-CBC by Theorem 5, and by Claim 1 $(V, \mathcal{C})$ is an $(n, N, k, m; r)$-MCBC. □

Theorem 5 is a special case of Theorem 6 for $r = 1$. In the following, when constructing an MCBC, we always construct its dual set system $(X, \mathcal{B})$, and check if it satisfies the multiset Hall's condition from Theorem 6. By adding an asterisk, we let $(X, \mathcal{B})^*$ denote its dual set system $(V, \mathcal{C})$. The following properties are obvious.

Remark 1 (i) If there exists some $h_0 < \lceil \frac{k}{r} \rceil$ such that for any h_0 blocks $B_{i_1}, B_{i_2}, \dots, B_{i_{h_0}}$, $| \cup_{j=1}^{h_0} B_{i_j}| \geq k$, then for any h such that $h_0 < h \leq \lceil \frac{k}{r} \rceil$, the multiset Hall's condition is also satisfied.

(ii) An $(n, N, k, m; r)$-MCBC is also an $(n, N, k', m; r)$-MCBC for any $k' \leq k$.

Example 3 By checking the multiset Hall's condition, it is possible to verify that Example 2 gives a construction of $(20, 80, k, 16; r)$-MCBC for any pair $(k, r) \in \{(16, 1), (11, 2), (10, 3), (7, 4)\}$. Especially, as will be shown in Construction 13 in the sequel, the code is optimal when $(k, r) \in \{(16, 1), (7, 4)\}$.

In the following sections, we will give several bounds and constructions of MCBCs.

3 Bounds of MCBCs

In this section, we give several bounds of MCBCs, which provide the results stated in Theorem 2. Note that we always assume $r \leq k \leq m < nr$ throughout the paper.

Lemma 1 (i) $N(n, k, m; r) \geq rn$.
(ii) $N(n, k, m; r) \geq N(n, k, m; i)$ *for* $i \in [r - 1]$.
(iii) $N(n, k, m; k) = kn$.

Proof (i) This inequality holds since each item has to be stored in at least r servers. (ii) This inequality holds from the definition of MCBCs. (iii) By (i), $N(n, k, m; k) \geq kn$. The trivial construction where each item is stored in arbitrary k servers gives an optimal code construction. □

Lemma 2 (i) $\frac{1}{r} N(nr, k, m) \leq N(n, k, m; r) \leq N(rn, k, m)$.

(ii) $N(n,k,m;r) \le rN(n, \lceil \frac{k}{r} \rceil, \lfloor \frac{m}{r} \rfloor)$.

Proof (i) From the proof of Claim 1 in Theorem 6, we can see that if there exists an $(n,N,k,m;r)$-MCBC, then there exists an (rn,rN,k,m)-CBC, and therefore $N(nr,k,m) \le rN(n,k,m;r)$.

Assume that there exists an (rn,N,k,m)-CBC given by the set system $(U=[rn], \mathcal{F}=\{F_1,F_2,\dots,F_m\})$. We construct a new set system $(V=[n], \mathcal{C}=\{C_1,\dots,C_m\})$ as follows. The i-th server contains the items given by the set $C_i=\{(\ell-1)(\bmod n)+1 : \ell \in F_i\}$. That is, each item in the set $U^{(\ell)}=\{\ell+jn : j \in [0,r-1]\}$ in each server is replaced with ℓ for $\ell \in [n]$ (without repetitions). The new set system $(V,\mathcal{C})$ defines an $(n,N',k',m;r')$-MCBC with storage $N' \le N$. To complete this proof we will show that $k'=k$ and $r'=r$.

Let Q be a multiset request for $(V,\mathcal{C})$ where the ℓ-th element, $\ell \in [n]$, is requested r_ℓ times, so $0 \le r_\ell \le r$ and $\sum_{\ell=1}^{n} r_\ell = k$. Consider the request P of $(U,\mathcal{F})$ which contains any r_ℓ distinct elements from $U^{(\ell)}$. Since $(U,\mathcal{F})$ is an (rn,N,k,m)-CBC, P can be read by taking subsets $D_j \subseteq F_j$, $|D_j| \le 1$ for $j \in [m]$. Then Q can be read from the servers $\{C_j : j \in [m], |D_j|=1\}$. Hence, $(V,\mathcal{C})$ is an $(n,N',k,m;r)$-MCBC with $N' \le N$, and $N(n,k,m;r) \le N(rn,k,m)$.

(ii) Assume that there exists an $(n,N,\lceil \frac{k}{r} \rceil, \lfloor \frac{m}{r} \rfloor)$-CBC given by the set system $(V=[n], \mathcal{F}=\{F_1,F_2,\dots,F_{\lfloor \frac{m}{r} \rfloor}\})$. We construct a new code by the following set system $(V=[n], \mathcal{C}=\{C_1,C_2,\dots,C_m\})$, such that $C_{i+j\lfloor \frac{m}{r} \rfloor}=F_i$ for any $i \in [\lfloor \frac{m}{r} \rfloor]$ and $j \in [0,r-1]$, and $C_\ell=\emptyset$ for any $r\lfloor \frac{m}{r} \rfloor+1 \le \ell \le m$. That is, each F_i for $i \in [\lfloor \frac{m}{r} \rfloor]$ is repeated r times.

Assume that the dual set system of $(V,\mathcal{F})$ is $(Y=[\lfloor \frac{m}{r} \rfloor], \mathcal{G}=\{G_1,\dots,G_n\})$, and the dual set system of $(V,\mathcal{C})$ is $(X=[m], \mathcal{B}=\{B_1,\dots,B_n\})$. Then $|B_i|=r|G_i|$ for $i \in [n]$. Since $(V,\mathcal{F})$ is an $(n,N,\lceil \frac{k}{r} \rceil, \lfloor \frac{m}{r} \rfloor)$-CBC, for any $1 \le h \le \lceil \frac{k}{r} \rceil$, and distinct $i_1,\dots,i_h \in [n]$, $|\cup_{j=1}^{h} G_{i_j}| \ge h$ by Theorem 5. Then for any $1 \le h \le \lceil \frac{k}{r} \rceil$, $|\cup_{j=1}^{h} B_{i_j}| \ge hr \ge \min\{hr,k\}$. Therefore, by Theorem 6, $(V,\mathcal{C})$ is an $(n,rN,k,m;r)$-MCBC, and $N(n,k,m;r) \le rN(n, \lceil \frac{k}{r} \rceil, \lfloor \frac{m}{r} \rfloor)$. □

Let $(V,\mathcal{C})$ be a set system of an $(n,N,k,m;r)$-MCBC and let $(X,\mathcal{B})$ be its dual set system. For $i \ge 0$, we denote by A_i the number of subsets in $\mathcal{B}$ of size i. Note that for $i<r$, $A_i=0$ since every item is contained in at least r different servers. As pointed in [17], $A_i=0$ for $i \ge k+1$ since for any block of size larger than k, we can reduce the block to k points and the multiset Hall's condition is still satisfied. The following bound is a generalization of the results in [3,7,17].

Lemma 3 *If $(X,\mathcal{B})^*$ is an $(n,N,k,m;r)$-MCBC with $r \le k-1$, and A_i for $i \in [k-1]$ is defined as above, then*

$$\sum_{i=r}^{k-1} \binom{m-i}{k-1-i} A_i \le \left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1}.$$

Proof Let M_{k-1} be the $\binom{m}{k-1} \times n$ matrix, whose rows are labeled by all the $(k-1)$-subsets of X, and the columns are labeled by the blocks in $\mathcal{B}$ that contain less than k points. The (i,j)-th entry of M_{k-1} is 1 if the j-th block B_j is contained in the i-th $(k-1)$-subset of X, and otherwise it is 0.

Each row in M_{k-1} has at most $\lfloor \frac{k-1}{r} \rfloor$ ones. In order to verify this property, assume in the contrary that there exist $\lfloor \frac{k-1}{r} \rfloor + 1$ blocks, and without loss of generality let them be the blocks $B_1, B_2, \ldots, B_{\lfloor\frac{k-1}{r}\rfloor+1}$, which are all subsets of the same $(k-1)$-subset. Therefore, $|\cup_{i=1}^{\lfloor\frac{k-1}{r}\rfloor+1} B_i| \leq k-1$, and the multiset Hall's condition is not satisfied, since $\lfloor\frac{k-1}{r}\rfloor+1 = \lceil k/r\rceil$ and $\min\{(\lfloor\frac{k-1}{r}\rfloor+1)r, k\} = k$. Every column which corresponds to a block of size $i < k$ has exactly $\binom{m-i}{k-1-i}$ ones, since the number of $(k-1)$-subsets of X containing a fixed i-subset is exactly $\binom{m-i}{k-1-i}$. Therefore, by counting the number of ones in M_{k-1} by rows and columns separately, we get that $\sum_{i=r}^{k-1}\binom{m-i}{k-1-i}A_i \leq \lfloor\frac{k-1}{r}\rfloor\binom{m}{k-1}$. □

According to Lemma 3, we derive the next theorem.

Theorem 7 *Let $r \leq k-1$. For any $c \in [r, k-1]$,*

$$N(n,k,m;r) \geq nc - \left\lfloor \frac{k-c}{m-k+1}\left\lceil \frac{\lfloor\frac{k-1}{r}\rfloor\binom{m}{k-1}}{\binom{m-c}{k-1-c}} - n\right\rceil\right\rfloor.$$

Proof The proof is similar to the one given in Lemma 3.2 [3], and hence we omit it here. □

4 Constructions of MCBCs

In this section we present several constructions of MCBCs. Constructions 1 and 3 are generalizations of the equivalent ones in [3,7,17] which determine the value of $N(n,k,m)$ in Theorem 1(ii) and (iii). Construction 4 is a generalization of that in [17] which determines the value of $N(n,k,k)$ in Theorem 1(i).

4.1 A construction by replication

Our first construction uses simple replication which is a generalization of the one in [3,7,17].

Construction 1 *Let n, k, m, r be positive integers such that $n \geq \lfloor\frac{k-1}{r}\rfloor\binom{m}{k-1}$ with $r < k$. We construct an $(n, N, k, m; r)$-MCBC with $N = kn - \lfloor\frac{k-1}{r}\rfloor\binom{m}{k-1}$, by explicitly constructing its dual set system $(X = [m], \mathcal{B} = \{B_1, \ldots, B_n\})$ as follows:*

1. *The first $\lfloor\frac{k-1}{r}\rfloor\binom{m}{k-1}$ blocks of $\mathcal{B}$ consist of $\lfloor\frac{k-1}{r}\rfloor$ copies of all different $(k-1)$-subsets of $[m]$.*
2. *Each remaining block of $\mathcal{B}$ is taken to be any k-subset of $[m]$.*

Thus, the value of N is given by

$$N = \left\lfloor\frac{k-1}{r}\right\rfloor\binom{m}{k-1}(k-1) + \left(n - \left\lfloor\frac{k-1}{r}\right\rfloor\binom{m}{k-1}\right)k = kn - \left\lfloor\frac{k-1}{r}\right\rfloor\binom{m}{k-1}.$$

The correctness of this construction is proved in the next theorem.

Theorem 8 *The code $(X, \mathcal{B})^*$ from Construction 1 is an $(n, N, k, m; r)$-MCBC with $n \geq \lfloor\frac{k-1}{r}\rfloor\binom{m}{k-1}$, $r < k$ and $N = kn - \lfloor\frac{k-1}{r}\rfloor\binom{m}{k-1}$.*

Proof We only need to check that $(X, \mathcal{B})$ satisfies the multiset Hall's condition. For $1 \leq h \leq \lceil\frac{k}{r}\rceil$, let $B_{i_1}, B_{i_2}, \ldots, B_{i_h} \in \mathcal{B}$ be some h different blocks. If there exists a block of size k or there exist two distinct blocks of size $k-1$, then $|\sum_{j=1}^{h} B_{i_j}| \geq k$; otherwise, we have $h \leq \lfloor\frac{k-1}{r}\rfloor$ by construction, and $|\sum_{j=1}^{h} B_{i_j}| \geq k-1 \geq \min\{hr, k\}$. □

Before we show that this construction is optimal, let us recall a useful lemma from [3,7].

Lemma 4 [3,7] *Let* $1 \le k \le m$ *and* $0 \le i \le k-1$. *Then* $\binom{m-i}{k-1-i} - 1 \ge (m-k+1)(k-1-i)$.

We can now deduce that Construction 1 is optimal.

Corollary 1 *For any* $n \ge \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$ *with* $r < k$, $N(n,k,m;r) = kn - \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$.

Proof For any $(n, N, k, m; r)$-MCBC, let A_i for $i \in [k]$ be the number of blocks in the dual set system of size i. By Lemma 4, for $i \le k-1$, $\binom{m-i}{k-1-i} \ge (m-k+1)(k-1-i)+1 \ge k-i$, then

$$\sum_{i=r}^{k-1}(k-i)A_i \le \sum_{i=r}^{k-1}\binom{m-i}{k-1-i}A_i \le \left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1},$$

and the last inequality holds according to Lemma 3. Therefore, we get that

$$\begin{aligned} N &= \sum_{i=r}^{k} iA_i = \sum_{i=r}^{k}(k-(k-i))A_i = \sum_{i=r}^{k} kA_i - \sum_{i=r}^{k}(k-i)A_i \\ &= kn - \sum_{i=r}^{k-1}(k-i)A_i \ge kn - \left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1}. \end{aligned}$$

Hence, we conclude that $N(n,k,m;r) = kn - \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$ when $n \ge \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$, since the codes from Construction 1 achieve this bound. □

As a special case when $r = k-1$ we get the following corollary.

Corollary 2 $N(n,k,m;k-1) = \begin{cases} kn - \binom{m}{k-1} & \text{if } n \ge \binom{m}{k-1}, \\ (k-1)n & \text{if } n < \binom{m}{k-1}. \end{cases}$

Proof For $n \ge \binom{m}{k-1}$, according to Corollary 1 for $r = k-1$, we get that $N(n,k,m;k-1) = kn - \binom{m}{k-1}$. For $n < \binom{m}{k-1}$, we slightly modify the code from Construction 1 such that the n blocks in $\mathcal{B}$ are some different $(k-1)$-subsets of $[m]$. It is readily verifies that the multiset Hall's condition holds for this modified construction and thus it provides an $(n, N = n(k-1), k, m; k-1)$-MCBC. Finally, according to Lemma 1(i), this construction is optimal. □

4.2 Constructions based on constant weight codes

Next, we give constructions based upon constant weight codes. Let (m, d, w)-code denote a binary constant weight code of length m, weight w and minimum Hamming distance d, and let $A(m, d, w)$ denote the maximum number of codewords of an (m, d, w)-code.

Construction 2 *Let* $X = [m]$ *and* $\mathcal{C}$ *be an* $(m, 2(k-w), w)$*-code with* n *codewords for some* $w \in [r, k-1]$. *Let* $\mathcal{B} = \{B_1, \ldots, B_n\}$ *be the support sets of all the codewords in* $\mathcal{C}$.

Theorem 9 *The code* $(X, \mathcal{B})^*$ *from Construction 2 is an* $(n, wn, k, m; r)$*-MCBC.*

Proof We only need to check that $(X, \mathcal{B})$ satisfies the multiset Hall's condition. It is satisfied as the size of each block in $\mathcal{B}$ is $w \ge r$ and since the minimum distance of $\mathcal{C}$ is $2(k-w)$, we get that the union of any two blocks in $\mathcal{B}$ is at least k. □

If we take $w = r$, we get the following family of optimal codes.

Corollary 3 *For any* $n \leq A(m, 2(k-r), r)$, $N(n, k, m; r) = rn$.

Proof By Construction 2, if there exists an $(m, 2(k-r), r)$-code with $A(m, 2(k-r), r)$ codewords, we get an $(n, rn, k, m; r)$-MCBC for any $n \leq A(m, 2(k-r), r)$, and it is optimal by Lemma 1(i). □

Constant weight codes are used in [3] to prove Theorem 1(iv). Now, we give a similar construction for MCBCs.

Construction 3 *Let* $X = [m]$, $r \leq k-2$. *Let* $\mathcal{C}$ *be an* $(m, 4, k-2)$*-code with* α *codewords with* $\alpha \leq A(m, 4, k-2)$. *First, let* $\mathcal{B}_0$ *be a set of* $\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$ *blocks, in which each* $(k-1)$*-subset of* $[m]$ *appears* $\lfloor \frac{k-1}{r} \rfloor$ *times. Let* $\mathcal{S}$ *consist of the support sets of the codewords in* $\mathcal{C}$. *Then, for any block in* $\mathcal{S}$, *add it to* $\mathcal{B}_0$, *and remove one copy of each of its* $m-k+2$ *supersets*[3] *of size* $k-1$ *in* $\mathcal{B}_0$. *Let the resulting block set be* $\mathcal{B}$.

Theorem 10 *The code* $(X, \mathcal{B})^*$ *from Construction 3 is an* $(n, N, k, m; r)$*-MCBC with*

$$n = \left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1} - \alpha(m-k+1) \text{ and } N = n(k-1) - \alpha,$$

where $\alpha \leq A(m, 4, k-2)$.

Proof Since the code has minimum distance four, for any two blocks in $\mathcal{S}$, their supersets of size $k-1$ are different. Therefore, each $(k-1)$-subset of $[m]$ is removed at most once. During the process, we add α blocks and remove $\alpha(m-k+2)$ blocks. Hence, we get that $n = \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - \alpha(m-k+1)$. Finally, since only α of all n blocks are of size $k-2$, we get that $N = n(k-1) - \alpha$.

Next we show that the multiset Hall's condition holds. For $1 \leq h \leq \lceil \frac{k}{r} \rceil$, let $B_{i_1}, B_{i_2}, \ldots, B_{i_h} \in \mathcal{B}$ be some h different blocks. In case $h = 1$, the size of each block in $\mathcal{B}$ is at least $k-2 \geq r$ so the condition holds and thus we assume that $h \geq 2$. If there exist two blocks $B_{i_a}, B_{i_b} \in \mathcal{B} \setminus \mathcal{S}$ which are different $(k-1)$-subsets, then $|B_{i_a} \cup B_{i_b}| \geq k$; if there exists two blocks $B_{i_a}, B_{i_b} \in \mathcal{S}$, then $|B_{i_a} \cup B_{i_b}| \geq k$ because of the minimum Hamming distance of the code $\mathcal{C}$ is four.

Therefore, we only need to check the case when there is one block from $\mathcal{S}$, and the other $h-1$ blocks are the same $(k-1)$-subset from $\mathcal{B} \setminus \mathcal{S}$. For example, $B_{i_1} \in \mathcal{S}$, and $B_{i_2}, \ldots, B_{i_h} \in \mathcal{B} \setminus \mathcal{S}$ are the same $(k-1)$-subset. If B_{i_1} is not a subset of B_{i_2}, then $|B_{i_1} \cup B_{i_2}| \geq k$; if B_{i_1} is a subset of B_{i_2}, then by the construction $h \leq \lfloor \frac{k-1}{r} \rfloor$ and $|\cup_{j=1}^{h} B_{i_j}| = k-1 \geq hr = \min\{hr, k\}$. Therefore, the multiset Hall's condition holds. □

The following lower bound of $A(n, 4, w)$ is known.

Lemma 5 [12] $A(n, 4, w) \geq \frac{1}{n}\binom{n}{w}$.

Next, we apply Construction 3 to get a family optimal codes.

Corollary 4 *For any* $\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - (m-k+1)A(m, 4, k-2) \leq n \leq \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$, $r \leq k-2$,

$$N(n, k, m; r) = n(k-1) - \left\lfloor \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - n}{m-k+1} \right\rfloor.$$

[3] For a block $S \in \mathcal{S}$ of size $k-2$, the *supersets* are the $(k-1)$-subsets of $[m]$ that contain S.

Proof When $r \leq k-2$, taking $c = k-1$ in Theorem 7, we have $N \geq n(k-1) - \left\lfloor \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - n}{m-k+1} \right\rfloor$. For any positive integers n, m, k, r such that $r+2 \leq k \leq m$ and

$$\left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1} - (m-k+1)A(m,4,k-2) \leq n \leq \left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1},$$

we have that $0 \leq \left\lfloor \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - n}{m-k+1} \right\rfloor \leq A(m,4,k-2)$. Let $\alpha = \left\lfloor \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - n}{m-k+1} \right\rfloor$. By Construction 3, there exists an $(n', N', k, m; r)$-MCBC with

$$n' = \left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1} - \alpha(m-k+1) \text{ and } N' = n'(k-1) - \alpha.$$

Removing any $n' - n$ blocks of size $k-1$ from its dual set system, we get an optimal $(n, N, k, m; r)$-MCBC with $N = N' - (k-1)(n'-n) = n(k-1) - \alpha$. □

4.3 A construction for $m = k$

In the following, we give a construction of $(n, N, k, k; r)$-MCBC and determine the value of $N(n, k, k; r)$ for $1 \leq r \leq k$.

Construction 4 *Let $m = k$ and $X = [k]$, $\mathcal{B} = \{B_1, \ldots, B_n\}$ and $k = \alpha r + \beta$, where $\alpha \geq 1$ and $0 \leq \beta \leq r-1$ such that the following holds.*

(i) *When $\beta = 0$, for any $n \geq \alpha$, let $B_i = [(i-1)r+1, (i-1)r+r]$ for $i \in [\alpha]$, and $B_i = [k]$ for any $i \in [\alpha+1, n]$.*
(ii) *When $\beta > 0$, for any $n \geq \alpha + r$, let $B_i = [(i-1)r+1, (i-1)r+r]$ for $i \in [\alpha]$, $B_i = [k] \setminus \{i-\alpha, i-\alpha+r, i-\alpha+2r, \ldots, i-\alpha+(\alpha-1)r\}$ for $i \in [\alpha+1, \alpha+r]$, and $B_i = [k]$ for any $i \in [\alpha+r+1, n]$.*

Theorem 11 *The code $(X, \mathcal{B})^*$ from Construction 4 is an $(n, N, k, k; r)$-MCBC with $N = kn - \lfloor \frac{k-1}{r} \rfloor k$.*

Proof We show that in both cases, i.e. $\beta > 0$ and $\beta = 0$, the multiset Hall's condition holds.

(i) $\beta = 0$. For any $h \geq 1$ different blocks $B_{i_1}, B_{i_2}, \ldots, B_{i_h} \in \mathcal{B}$, if there exists some block $B_{i_j} = [k]$, then $|\cup_{j=1}^h B_{i_j}| = k \geq r$; otherwise, $i_j \in [\alpha]$ for all $j \in [h]$, and then $|\cup_{j=1}^h B_{i_j}| = hr$ since by the construction the blocks $B_1, \ldots, B_\alpha$ are mutually disjoint. Thus, the multiset Hall's condition holds, and it is an $(n, N, k, k; r)$-MCBC with

$$N = \alpha r + (n-\alpha)k = kn - \frac{k}{r}(k-r) = kn - \left\lfloor \frac{k-1}{r} \right\rfloor k.$$

(ii) $\beta > 0$. First, note that for any $i \in [\alpha+1, \alpha+r]$, $|B_i| = k - \alpha \geq r$. This holds since $k - \alpha - r = \alpha r + \beta - \alpha - r = (\alpha-1)(r-1) + (\beta-1) \geq 0$. For any h different blocks $B_{i_1}, B_{i_2}, \ldots, B_{i_h} \in \mathcal{B}$ with $h \geq 2$, if there exists some $j \in [h]$ such that $B_{i_j} = [k]$ or for all $j \in [h]$, $i_j \in [\alpha]$, then the proof is similar as in case (i). If there exist two blocks B_{i_a}, B_{i_b} such that $i_a, i_b \in [\alpha+1, \alpha+r]$, then $|B_{i_a} \cup B_{i_b}| = k$. Therefore, the remaining case to check is when only one block is from the set $\{B_i : i \in [\alpha+1, \alpha+r]\}$, and the other blocks are from the set $\{B_i : i \in [\alpha]\}$. Without loss of generality assume that

$i_1, i_2, \ldots, i_{h-1} \in [\alpha]$ and $i_h \in [\alpha+1, \alpha+r]$. Since $|B_{i_h}| = k-\alpha$, when $1 \leq h \leq \alpha-1$, by the construction we get that $|\cup_{j=1}^{h} B_{i_j}| \geq k-\alpha+h-1$. Since

$$k-\alpha+h-1-hr = \alpha r+\beta-\alpha+h-1-hr = (\alpha-h)(r-1)+(\beta-1) \geq 0,$$

we conclude that $|\cup_{j=1}^{h} B_{i_j}| \geq hr$. If $h=\alpha$, by the construction we get that $|\cup_{j=1}^{h} B_{i_j}| = k-1 = \alpha r+\beta-1 \geq \alpha r$. Lastly, if $h = \alpha+1$, then $|\cup_{j=1}^{h} B_{i_j}| = k$. Therefore, the multiset Hall's condition is satisfied for any $1 \leq h \leq \lceil \frac{k}{r} \rceil$, and the code is an $(n, N, k, k; r)$-MCBC with

$$N = kn-\alpha(k-r)-\alpha r = kn-\alpha k = kn-\left\lfloor \frac{k-1}{r} \right\rfloor k.$$

□

The next corollary summarizes the construction and results in this section.

Corollary 5 $N(n,k,k;r) = kn - \lfloor \frac{k-1}{r} \rfloor k$ *if* $r \mid k$, $n \geq \frac{k}{r}$ *or* $r \nmid k$, $n \geq \lfloor \frac{k}{r} \rfloor + r$.

Proof Taking $m = k$ in Lemma 3, we have that $\sum_{i=r}^{k-1}(k-i)A_i \leq \lfloor \frac{k-1}{r} \rfloor k$. Similarly to the proof of Corollary 1, we get that

$$\begin{aligned} N(n,k,k;r) &= \sum_{i=r}^{k} iA_i = \sum_{i=r}^{k}(k-(k-i))A_i = \sum_{i=r}^{k} kA_i - \sum_{i=r}^{k}(k-i)A_i \\ &= kn - \sum_{i=r}^{k-1}(k-i)A_i \geq kn - \left\lfloor \frac{k-1}{r} \right\rfloor k. \end{aligned}$$

Hence we conclude that $N(n,k,k;r) = kn - \lfloor \frac{k-1}{r} \rfloor k$ since Construction 4 gives optimal codes that reach this bound. □

4.4 A construction from Steiner systems

In the following we construct a class of MCBCs based upon Steiner systems, which is a generalization of Example 2.

A *Steiner system* $S(2,\ell,m)$ is a set system $(X, \mathcal{B})$, where X is a set of m points, $\mathcal{B}$ is a collection of ℓ-subsets (blocks) of X, such that each pair of points in X occurs together in exactly one block of $\mathcal{B}$. By the well known Fisher's inequality [14], for an $S(2,\ell,m)$ with $m > \ell \geq 2$, $|\mathcal{B}| \geq m$. For the existence of Steiner systems, we refer the reader to [11].

Construction 5 *Let* $X = [m]$ *and* $(X, \mathcal{B})$ *be an* $S(2,\ell,m)$ *for some* $\ell < m$.

Theorem 12 *The code* $(X, \mathcal{B})^*$ *from Construction 5 is a* $(|\mathcal{B}|, \ell|\mathcal{B}|, k, m; r)$*-MCBC for any* $\lfloor \frac{\ell}{2} \rfloor + 1 \leq r \leq \ell$ *and* $k \leq (\ell-r+1)(2r-1)$.

Proof By Remark 1(ii), we only need to check that $(X, \mathcal{B})^*$ is a $(|\mathcal{B}|, \ell|\mathcal{B}|, k, m; r)$-MCBC for any $\lfloor \frac{\ell}{2} \rfloor + 1 \leq r \leq \ell$ and $k = (\ell-r+1)(2r-1)$.

Let us first determine the number of points in the union of any h blocks in $\mathcal{B}$. Since any two blocks intersect in at most one point, for any $i \in [2,h]$, if the first $i-1$ blocks are chosen, then the i-th block can contribute at least $\ell-(i-1)$ new points. Therefore, if $h \leq \ell+1 \leq |\mathcal{B}|$, the union of any h blocks contains at least

$$\ell + (\ell-1) + \cdots + (\ell-(h-1)) = h\ell - \binom{h}{2}$$

points.

Let us consider some h blocks $B_{i_1}, B_{i_2}, \ldots, B_{i_h}$ with

$$1 \le h \le \left\lceil \frac{k}{r} \right\rceil = \left\lceil \frac{(\ell - r + 1)(2r - 1)}{r} \right\rceil \le \left\lceil 2(\ell - r + 1) - \frac{l+1}{r} + 1 \right\rceil \le 2(\ell - r) + 2,$$

where the last inequality holds since $r \ge \lfloor \frac{\ell}{2} \rfloor + 1$. We see $2(\ell - r) + 2 \le \ell + 1$ for $r \ge \lfloor \frac{\ell}{2} \rfloor + 1$.

If $h \in [2(\ell - r) + 1]$, then $r \le \ell - \frac{h-1}{2}$, and

$$| \cup_{j=1}^{h} B_{i_j} | \ge h\ell - \binom{h}{2} = h\left(\ell - \frac{h-1}{2}\right) \ge hr = \min\{hr, k\}.$$

If $h = 2(\ell - r) + 2$, then $| \cup_{j=1}^{h} B_{i_j} | \ge \ell h - \binom{h}{2} = (\ell - r + 1)(2r - 1) = k$. Therefore, the multiset Hall's condition holds for any $1 \le h \le \lceil \frac{k}{r} \rceil$. □

An *affine plane* of order q is an $S(2, q, q^2)$. It has q^2 points and $q^2 + q$ blocks. It is well known that an affine plane exists for any prime power q [11]. The next result of CBCs based upon affine planes was given in [20].

Theorem 13 [20] *Let q be a prime power and $(X, \mathcal{B})$ be an affine plane of order q. Then $(X, \mathcal{B})^*$ is an optimal uniform $(q^2 + q, q^3 + q^2, q^2, q^2)$-CBC.*

The code in Theorem 13 is also an optimal CBC, since for $k = m$ by Theorem 1(i), $N(q^2 + q, q^2, q^2) = q^3 + q^2$. However, note that it is a different code from the optimal (n, N, k, k)-CBC in [17] which is constructed as follows: Let $X = [k]$, and $\mathcal{B} = \{B_1, \ldots, B_n\}$, which are given by $B_i = \{i\}$ for $i \in [k]$, and $B_i = [k]$ for $i \in [k + 1, n]$. By Theorem 12, we can see the code in Theorem 13 is also $(q^2 + q, q^3 + q^2, k, q^2; r)$-MCBCs for different pair-values of k and r.

Corollary 6 *Let q be a prime power. Then there exists a $(q^2 + q, q^3 + q^2, k, q^2; r)$-MCBC for any $\lfloor \frac{q}{2} \rfloor + 1 \le r \le q$ and $k \le (q - r + 1)(2r - 1)$.*

Note that since there are constructions of constant weight codes based on Steiner systems (see [4]), Constructions 2 and 5 may generate the same code in some cases. For example, when $r = q$, we receive a $(q^2 + q, q^3 + q^2, 2q - 1, q^2; q)$-MCBC from Corollary 6 (it is optimal because it reaches the bound in Lemma 1(i) with total storage $N = rn = q(q^2 + q) = q^3 + q^2$). We can also consider that this code is obtained by Construction 2 using $(q^2, 2q - 2, q)$-codes, whose existence follows from affine planes as below: for any block $B \in \mathcal{B}$, we get a codeword u of length q^2 in which the value of each coordinate u_i for $i \in [q^2]$ is 1 if and only if $i \in B$. Since any two blocks intersect in at most one point, the distance between every two distinct codewords is at least $2(q-1)$. However, when $r < q$, by applying Construction 2 to the $(q^2, 2q-2, q)$-codes, we can only get a $(q^2+q, q^3+q^2, 2q-1, q^2; r)$-MCBC. Meanwhile, by analyzing the property of Steiner systems, we get wider range of k from Construction 5.

Lastly, we note that it is also possible to improve the value of k in Theorem 12 when $r \le \lfloor \frac{q}{2} \rfloor$. We demonstrate this in the following example.

Example 4 Let $q = 4$. From Corollary 6, we obtain a $(20, 80, k, 16; r)$-MCBC for $(k, r) \in \{(10, 3), (7, 4)\}$. In addition, one can check that for the dual set system $(X, \mathcal{B})$ of the $(20, 80, 16, 16)$-CBC in Example 2, the lower bounds on the size of the union of any h blocks, $1 \le h \le 6$, are as shown in the following table:

h	Size of union
1	4
2	7
3	9
4	10
5	10
6	11

Therefore, we get also a (20, 80, 11, 16; 2)-MCBC.

5 Regular MCBCs

In this section, we study regular MCBCs, and give a construction for such codes. Given n, m, k, r, let $\mu(n, k, m; r)$ denote the smallest number of items stored in each server in a regular MCBC. The following lemma presents a simple lower bound on the value of $\mu(n, k, m; r)$.

Lemma 6 $\mu(n, k, m; r) \geq \left\lceil \frac{N(n,k,m;r)}{m} \right\rceil$.

Proof This property holds since a regular $(n, N, k, m; r)$-MCBC is also an $(n, N, k, m; r)$-MCBC, and therefore $m\mu(n, k, m; r) \geq N(n, k, m; r)$. □

Remark 2 It is easy to check that the constructions of MCBCs in Sect. 4 also give regular MCBCs for some specific parameters. Moreover, when the MCBCs are optimal, the bound in Lemma 6 holds with equality. However, not all of these constructions are regular. For example, the property in which the codes in Lemma 1(iii) and Construction 1 are regular depends on the choice of the k-subsets.

According to Remark 2, we give one choice of the k-subsets in Lemma 1(iii) to make the construction regular, and therefore determine when regular MCBCs with $r = k$ and minimum storage $N = kn$ exist.

Construction 6 *Let* $n = \frac{m}{\gcd(m,k)}$ *and* $k \leq m$*, then we have* $\frac{nk}{m} = \frac{k}{\gcd(m,k)}$ *and* $m|nk$*. Let* $I = [0, nk-1] \subseteq \mathbb{Z}$*, and for each* $i \in [n]$*,* $I^{(i)} = [(i-1)k, ik-1]$*. Then let* $X = [m]$ *and* $\mathcal{B} = \{B_1, \ldots, B_n\}$*, where* $B_i = \{j \pmod m + 1 : j \in I^{(i)}\}$ *for* $i \in [n]$.

Theorem 14 *The code* $(X, \mathcal{B})^*$ *from Construction 6 is a regular* $(n, k, m; k)$*-MCBC.*

Proof For any $a, b \in I, a \neq b$, if $a \pmod m + 1 = b \pmod m + 1$, then $m \mid (a - b)$. Since each $I^{(i)}$ consists of k consecutive integers and $k \leq m$, then $|B_i| = k$ for any $i \in [n]$. We only need to prove that the code is regular. This holds since each $i \in [m]$ appears in exactly $\frac{nk}{m} = \frac{k}{\gcd(m,k)}$ blocks in $\mathcal{B}$. □

Corollary 7 $\mu(n, k, m; k) = \frac{kn}{m}$ *if and only if* $n = c \cdot \frac{m}{\gcd(m,k)}$ *for some integer* $c \geq 0$.

Proof If $\mu(n, k, m; k) = \frac{kn}{m}$, then the value of n satisfies $n = c \cdot \frac{m}{\gcd(m,k)}$ for some integer $c \geq 0$ so that $\mu(n, k, m; k)$ is an integer. For any such n, Construction 6 gives a code with the desired parameters as follows. Assume that $n = c \cdot \frac{m}{\gcd(m,k)}$, and let $X = [m]$ and $\hat{\mathcal{B}}$ consist of c copies of all the blocks of $\mathcal{B}$ in Construction 6. Then, $(X, \hat{\mathcal{B}})^*$ is a regular $(n, k, m; k)$-MCBC. □

6 Conclusion

In this paper, we generalized combinatorial batch codes to multiset combinatorial batch codes and regular multiset combinatorial batch codes. Several bounds and constructions of optimal codes were obtained. To conclude, when $n < \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - (m-k+1)A(m,4,k-2)$ and $r \leq k-2$, the determination of the exact value of minimum storage $N(n,k,m;r)$ for MCBCs is almost still open.

Acknowledgements The authors would like to thank Prof. Tuvi Etzion for valuable discussions. The authors would also like to express their gratitude to the anonymous reviewers for their detailed and constructive comments. The work of Hui Zhang was supported in part at the Technion by a fellowship of the Israel Council of Higher Education.

References

1. Aharoni R., Kessler O.: On a possible extension of Hall's theorem to bipartite hypergraphs. Discret. Math. **84**, 309–313 (1990).
2. Balachandran N., Bhattacharya S.: On an extremal hypergraph problem related to combinatorial batch codes. Discret. Appl. Math. **162**, 373–380 (2014).
3. Bhattacharya S., Ruj S., Roy B.: Combinatorial batch codes: a lower bound and optimal constructions. Adv. Math. Commun. **6**, 165–174 (2012).
4. Brouwer A.E., Shearer J.B., Sloane N.J.A., Smith W.D.: A new table of constant weight codes. IEEE Trans. Inf. Theory **36**(6), 1334–1380 (1990).
5. Brualdi R.A., Kiernan K.P., Meyer S.A., Schroeder M.W.: Combinatorial batch codes and transversal matroids. Adv. Math. Commun. **4**, 419–431 (2010).
6. Bujtás C., Tuza Z.: Combinatorial batch codes: extremal problems under Hall-type conditions. Electr. Notes in Discret. Math. **38**, 201–206 (2011).
7. Bujtás C., Tuza Z.: Optimal batch codes: many items or low retrieval requirement. Adv. Math. Commun. **5**, 529–541 (2011).
8. Bujtás C., Tuza Z.: Optimal combinatorial batch codes derived from dual systems. Miskolc Math. Notes **12**(1), 11–23 (2011).
9. Bujtás C., Tuza Z.: Relaxations of Hall's condition: optimal batch codes with multiple queries. Appl. Anal. Discret. Math. **6**(1), 72–81 (2012).
10. Bujtás C., Tuza Z.: Turán numbers and batch codes. Discret. Appl. Math. **186**, 45–55 (2015).
11. Colbourn C.J., Mathon R.: Steiner systems. In: Colbourn C.J., Dinitz J.H. (eds.) Handbook of Combinatorial Designs, 2nd edn, pp. 102–110. Chapman & Hall/CRC, Boca Raton (2007).
12. Graham R.L., Sloane N.J.A.: Lower bounds for constant weight codes. IEEE Trans. Inf. Theory **26**, 37–43 (1980).
13. Ishai Y., Kushilevitz E., Ostrovsky R., Sahai A.: Batch codes and their applications. In: Proceedings of the 36-Sixth Annual ACM Symposium on Theory of Computing STOC '04, pp. 262–271 (2004).
14. Mathon M., Rosa R.: 2-(v, k, λ) designs of small order. In: Colbourn C.J., Dinitz J.H. (eds.) Handbook of Combinatorial Designs, 2nd edn, pp. 25–58. Chapman & Hall/CRC, Boca Raton (2007).
15. Mirsky L.: Systems of representatives with repetition. Proc. Camb. Philos. Soc. **63**, 1135–1140 (1967).
16. Mirsky L., Perfect H.: Systems of representatives. J. Math. Anal. Appl. **15**, 520–568 (1966).
17. Paterson M.B., Stinson D.R., Wei R.: Combinatorial batch codes. Adv. Math. Commun. **3**, 13–27 (2009).
18. Rawat A.S., Papailiopoulos D.S., Dimakis A.G., Vishwanath S.: Locality and availability in distributed storage. IEEE Trans. Inf. Theory **62**, 4481–4493 (2016).
19. Silberstein N.: Fractional repetition and erasure batch codes. In: Proceedings of the 4th International Castle Meeting on Coding Theory and Applications, Palmela, Portugal (2014).
20. Silberstein N., Gál A.: Optimal combinatorial batch codes based on block designs. Des. Codes Cryptogr. **78**, 1–16 (2014).
21. Zeh A., Yaakobi E.: Bounds and constructions of codes with multiple localities. In: Proceedings of the IEEE International Symposium Information Theory (ISIT), Barcelona, Spain, pp. 640–644 (2016).
22. Zhang H., Yaakobi E., Silberstein N.: Multiset combinatorial batch codes. In: Proceedings of the IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, pp. 2183–2187 (2017).