Lower Bounds for Total Storage of Multiset Combinatorial Batch Codes Using Linear Programming

Yeow Meng Chee, *Senior Member, IEEE*, Han Mao Kiah, *Member, IEEE*, and Hui Zhang, *Member, IEEE*

***Abstract*—The class of multiset combinatorial batch codes (MCBCs) was introduced by Zhang *et al.* (2018) as a generalization of combinatorial batch codes (CBCs), which are replication-based batch codes. The MCBCs allow multiple users to retrieve items in parallel in a distributed storage and a fundamental objective in this study is to determine the minimum total storage given certain requirements. We formulate linear programs so that the optimal solutions provide lower bounds on the total storage of MCBCs. Borrowing techniques from linear programming, we improve known lower bounds in some cases. Furthermore, for some parameters, we showed that these lower bounds are either tight or asymptotically tight by constructing the corresponding codes.**

***Index Terms*—Distributed storage codes, combinatorial batch codes, multiset combinatorial batch codes, linear programming.**

I. Introduction

MOTIVATED by applications such as load balancing in distributed storage, private information retrieval and cryptographic protocols, Ishai *et al.* introduced the notions of *batch codes* and *multiset batch codes* [15]. Formally, a batch code encodes a database of n items into a set of m *servers* (or *buckets*), so that a user's request that comprises a *batch* or set of k database items can be retrieved by reading a restricted number of encoded symbols from each server. In the same paper, Ishai *et al.* defined *multiset batch codes* to facilitate parallel item retrieval by multiple users. In this distributed setup, we allow each user to download directly from the servers, but each server can only serve the request of at most one user. Furthermore, Ishai *et al.* assume that there are k users who each requests a data item. Hence, the total request is a *multiset* of k items, or simply, a k-tuple (see Figure 1(a)–(c)). Since the seminal work of Ishai *et al.* and recent interest in coding for distributed data storage, there has been extensive study on the fundamental limits and constructions of batch and multiset batch codes [1], [12], [15], [22], [26]–[28].

In this paper, we study *replication-based* batch codes, where each server stores a subset of data items and decoding simply means reading items from servers (see Figure 1(d) and (e)). While replication-based batch codes were introduced by Ishai *et al.* [15], these codes were systematically studied by Paterson *et al.* later in [21]. In the latter paper, these codes were called *combinatorial batch code* (CBC) and Paterson *et al.* studied the fundamental objective of minimizing storage, i.e. the total number of items (counting all replicates) stored in all the servers, for fixed values of n, k and m.

A large amount of research has been done on CBCs after the primary work of Paterson *et al.* [2], [4], [7]–[11], [16], [18], [23]–[25], [29]. In particular, one special class, called *uniform* CBC defined in [21], is a CBC in which each item is stored in exactly the same number of servers (see also [2], [4], [11], [21], [25]). Later, to address availability issues in distributed storage systems, Silberstein [23] introduced the notion of *erasure CBCs* and these erasure CBCs were also studied in [16], [18]. There are also other applications of combinatorial batch codes. For example, the connection between *fractional repetition codes* and combinatorial batch codes was established in [24], where the authors demonstrate that *fractional repetition batch codes* facilitate uncoded efficient exact repairs and load balancing in distributed storage system simultaneously. More recently, CBCs were used to construct efficient multi-point function secret sharing schemes [5] and design recovery schemes with performance guarantees in network middleboxes [19].

In the same spirit as multiset batch codes, Zhang *et al.* generalised the concept of CBCs to that of *multiset CBCs* (MCBCs) [29]. Here, we require that any multiset request can be retrieved as long as each item appears at most r times for some nonnegative integer r. Similar to CBCs, a fundamental problem in the study on MCBCs is to minimize the total storage N, given the requirements on the number of data items n, request size k, number of servers m, and the maximum multiplicity r of any request. In particular, when $r = 1$, the MCBC is simply a CBC. Readers may refer to [29] for a

Manuscript received May 20, 2019; revised September 6, 2020; accepted September 14, 2020. Date of publication October 8, 2020; date of current version December 21, 2020. The work of Han Mao Kiah was supported by the National Research Foundation, Singapore, through its Strategic Capability Research Centres Funding Initiative. The work of Hui Zhang was supported by the National Research Foundation, Singapore under Grant NRF-RSS2016-004. This article was presented in part at the 2019 IEEE International Symposium on Information Theory. *(Corresponding author: Hui Zhang.)*

Yeow Meng Chee and Hui Zhang are with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 117576 (e-mail: ymchee@nus.edu.sg; isezhui@nus.edu.sg).

Han Mao Kiah is with the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371 (e-mail: hmkiah@ntu.edu.sg).

Communicated by L. Dolecek, Associate Editor for Coding Techniques.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.3029720

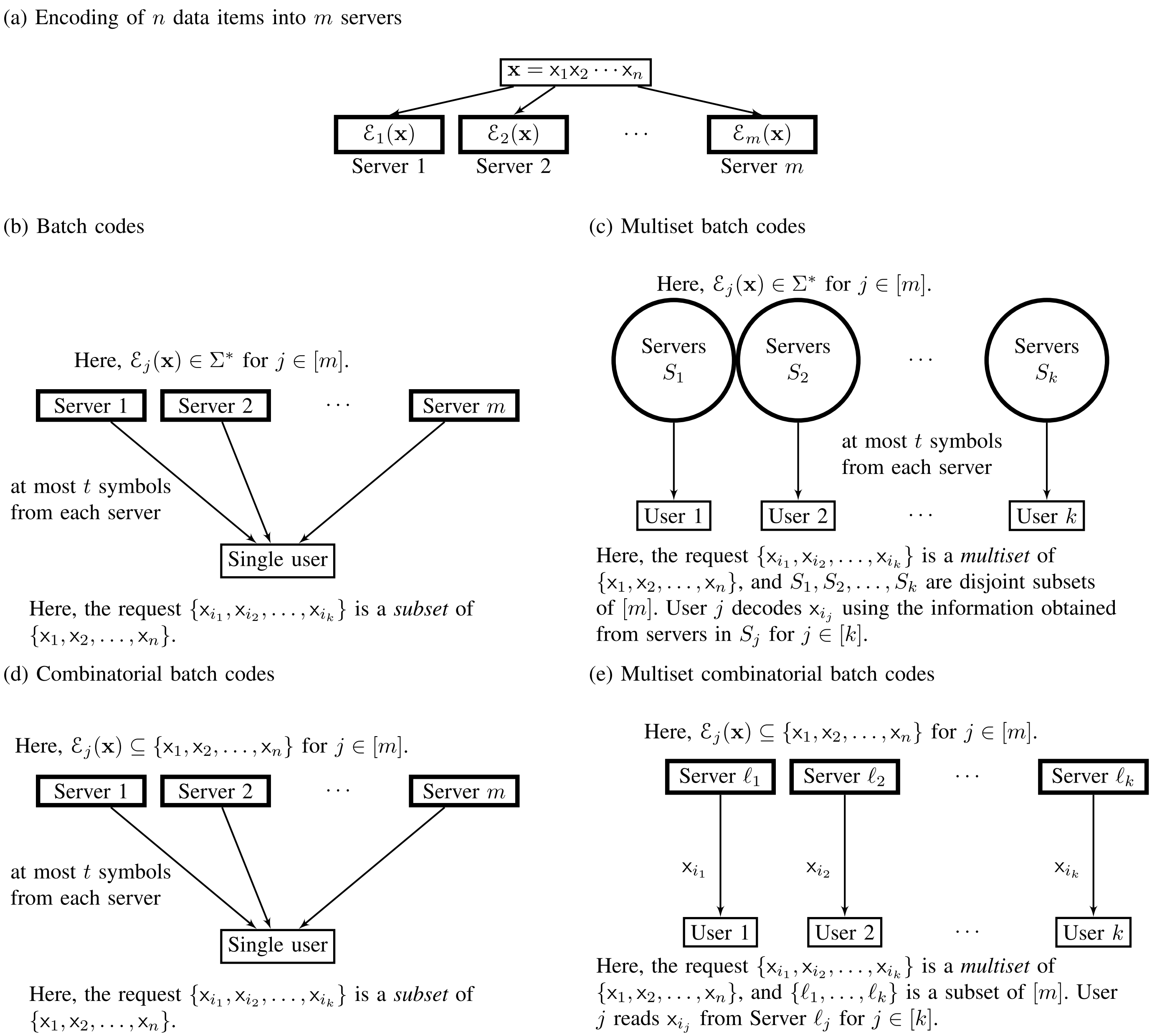
(a) Encoding of n data items into m servers
$\mathbf{x} = \mathsf{x}_1\mathsf{x}_2\cdots\mathsf{x}_n$
$\mathcal{E}_1(\mathbf{x})$
$\mathcal{E}_2(\mathbf{x})$
$\cdots$
$\mathcal{E}_m(\mathbf{x})$
Server 1
Server 2
Server m
(b) Batch codes
Here, $\mathcal{E}_j(\mathbf{x}) \in \Sigma^*$ for $j \in [m]$.
Server 1
Server 2
Server m
at most t symbols from each server
Single user
Here, the request $\{\mathsf{x}_{i_1}, \mathsf{x}_{i_2}, \ldots, \mathsf{x}_{i_k}\}$ is a *subset* of $\{\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n\}$.
(c) Multiset batch codes
Here, $\mathcal{E}_j(\mathbf{x}) \in \Sigma^*$ for $j \in [m]$.
Servers S_1
Servers S_2
Servers S_k
at most t symbols from each server
User 1
User 2
User k
Here, the request $\{\mathsf{x}_{i_1}, \mathsf{x}_{i_2}, \ldots, \mathsf{x}_{i_k}\}$ is a *multiset* of $\{\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n\}$, and $S_1, S_2, \ldots, S_k$ are disjoint subsets of $[m]$. User j decodes x_{i_j} using the information obtained from servers in S_j for $j \in [k]$.
(d) Combinatorial batch codes
Here, $\mathcal{E}_j(\mathbf{x}) \subseteq \{\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n\}$ for $j \in [m]$.
Server 1
Server 2
Server m
at most t symbols from each server
Single user
Here, the request $\{\mathsf{x}_{i_1}, \mathsf{x}_{i_2}, \ldots, \mathsf{x}_{i_k}\}$ is a *subset* of $\{\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n\}$.
(e) Multiset combinatorial batch codes
Here, $\mathcal{E}_j(\mathbf{x}) \subseteq \{\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n\}$ for $j \in [m]$.
Server ℓ_1
Server ℓ_2
Server ℓ_k
x_{i_1}
x_{i_2}
x_{i_k}
User 1
User 2
User k
Here, the request $\{\mathsf{x}_{i_1}, \mathsf{x}_{i_2}, \ldots, \mathsf{x}_{i_k}\}$ is a *multiset* of $\{\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n\}$, and $\{\ell_1, \ldots, \ell_k\}$ is a subset of $[m]$. User j reads x_{i_j} from Server ℓ_j for $j \in [k]$.

Fig. 1. Batch codes and its variants. (a) In all definitions, we are interested in designing the encoding functions $\mathcal{E}_j(\mathbf{x}) \in \Sigma^*$ for the string $\mathbf{x}$ over an alphabet Σ, so as to distribute the storage of the n data items on Server j for $j \in [m]$. (b) and (c) Original definitions of batch codes and multiset batch codes provided by Ishai *et al.* [15]. (d) Combinatorial batch codes (CBCs), or replication-based batch codes. (e) Multiset combinatorial batch codes (MCBCs). Notice that as described in Definition 2, C_j is composed of all the indices of data items stored in $\mathcal{E}_j(\mathbf{x})$ for $j \in [m]$ in (d) and (e).

summary of parameters about CBCs and MCBCs for which the exact values of minimum total storage have been determined.

In this work, we continue the investigation for MCBCs. Specifically, we introduce a new technique[1], *linear programming*, to provide new lower bounds on the minimum total storage of the MCBCs. Using certain necessary conditions, we formulate integer linear programming problems so that the optimal solutions provide lower bounds to our quantity of interest. In particular, we provide an alternative proof of the lower bounds in previous work [4], [29] by solving the linear programming problem. Later, we also introduce additional constraints which are imposed from the definitions of MCBCs. In this case, we not only improve the known lower bound, but also determine the optimal values in certain cases that cannot be determined in previous work, especially when $r \in \{1, k/2, k-2\}$. Finally, we focus on a specific set of parameters $(k, r) = (6, 3)$ and formulate a linear program as in previous cases. A stronger lower bound is obtained and constructions of codes reaching this bound exactly or asymptotically are given. To conclude, we remark that our lower bounds and code constructions rely on known results of optimal constant weight codes. The class of constant weight codes is a fundamental well-studied combinatorial object in coding theory and these codes have been widely used in digital communication systems (see [6], [14], [17] and the references therein).

[1] Our techniques are applicable to other variants of CBCs and MCBCs and we defer this investigation to future work.

II. Preliminaries

In this section, we formally introduce the background, definitions and a brief summary of known results. Given integers $m < n$, we use $[m, n]$ to denote the set of integers $\{m, m+1, \ldots, n\}$ and $[n]$ to denote the set $[1, n]$.

A. Background and Definitions

Suppose that a database is partitioned into n data items $\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n$. Batch codes and multiset batch codes were introduced in [15] as a means to represent systems that distribute these n data items on m servers. The formal definitions are as follows.

Definition 1:

(i) An (n, N, k, m, t) *batch code* over an alphabet Σ, encodes a string $\mathbf{x} \in \Sigma^n$ into an m-tuple of strings $\mathcal{E}_1(\mathbf{x}), \ldots, \mathcal{E}_m(\mathbf{x}) \in \Sigma^*$ (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 $\mathsf{x}_{i_1}, \ldots, \mathsf{x}_{i_k}$ can be decoded by reading at most t symbols from each server.

(ii) 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 servers into k subsets $S_1, \ldots, S_k \subseteq [m]$ such that each item $\mathsf{x}_{i_j}, j \in [k]$, can be retrieved by reading at most t symbols from each bucket in S_j.

As mentioned in the former section, an MCBC simply distributes the physical items of the database into the servers. For convenience, we only list the indices of data items storing in each server, which is a subset of $[n]$.

Definition 2: An $(n, N, k, m; r)$ *multiset combinatorial batch code (MCBC)* is a set system [2] $([n], \mathcal{C})$, where the following hold.

(i) $\mathcal{C}$ is a collection of m blocks C_j, $j \in X$ with $\sum_{j \in X} |C_j| = N$. We refer to each block C_j as a *server* and so, C_j comprises the indices of all items stored in the corresponding server. Here, X is the index set for the servers and $|X| = m$. In this paper, unless otherwise state, the index set X is always $[m]$.

(ii) For every multiset request $\{i_1, i_2, \ldots, i_k\}$ where each element has multiplicity at most r, we can find distinct blocks C_{ℓ_j} such that $i_j \in C_{\ell_j}$ for all $j \in [k]$.

When $r = 1$, an $(n, N, k, m; r)$-MCBC is simply a CBC. Note that in the setting of MCBC, it suffices to consider the case where each user reads or accesses *only one item* in a server. This is because in a replication-based batch code, all stored items are replicates of the original n data items. Since each user only requests one item, there is no incentive for a user to download more than one replicate. However, for CBCs, we may allow the user to read more than one symbol from each server and in this case, there are gains due to space savings. Interested readers may refer to [10] for more details.

Finally, to avoid triviality, we always assume $r \le k \le m$ throughout the paper.

Example 3: Let $n = 31$ and $m = 12$. We distribute the data items $\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_{31}$ across 12 servers according to the set system described in Figure 2. For example, Server 1 stores the replicates x_1, x_5, x_8, x_{11}, x_{14}, x_{17}, x_{20}, x_{23}, x_{26}, and x_{31}.

Here, $([31], \{C_i : i \in [12]\})$ forms a set system for a $(31, 120, 6, 12; 3)$-MCBC. In other words, for any multiset request of length six and multiplicity at most three, we are able to contact six distinct servers to retrieve the request.

For example, if six users request for items x_1, x_1, x_1, x_2, x_2, and x_2, respectively, one for each user, then they may access Servers of indices 1, 5, 9, 2, 6, and 10, respectively, to retrieve their items in *parallel*. Also, if they request for items x_5, x_5, x_5, x_8, x_8, and x_8, respectively, then they may access Servers of indices 1, 3, 4, 2, 7, and 8, respectively, to retrieve their items.

In this example, since every server stores exactly ten items, we have that the total storage N is 120. The results in this paper will demonstrate that $N = 120$ is indeed optimal. □

B. Previous Results

Following the primary work of Paterson *et al.* [21], a significant amount of work has been done to determine the minimum total storage N of an MCBC given the other parameters n, k, m and r (see Zhang *et al.* [29] for a survey).

In particular, let $N(n, k, m; r)$ denote the smallest N such that an $(n, N, k, m; r)$-MCBC exists, and set $N(n, k, m) = N(n, k, m; 1)$ as in [21]. The MCBC with $N = N(n, k, m; r)$ is said to be *optimal*. Naturally, for fixed n, k, r, we define an $(n, N, k, m; r)$-MCBC is *asymptotically optimal* if $N - N(n, k, m; r) = o(N(n, k, m; r))$ with respect to m.

Exact values of $N(n, k, m)$ have been established for the following parameters.

(i) $n \ge \binom{m}{k-2}$ (see [4], [8], [21]).
(ii) $k \in \{2, 3, 4\}$ (see [9], [21]).
(iii) $n \in \{m+1, m+2\}$ (see [7], [9], [21]).
(iv) $m = k$ (see [15], [21]).
(v) $\binom{m}{k-2} - (m-k+1)A(m, 4, k-3) \le n \le \binom{m}{k-2}$ (see [4]).

For MCBCs with $r > 1$, the values of $N(n, k, m; r)$ have been determined when $r \ge k-1$ for all n; and when $r \le k-2$ and $n \ge \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - (m-k+1)A(m, 4, k-2)$. Other adhoc results are given in [29]. Here, $A(n, d, w)$ refers to the maximum size of a binary constant-weight code of length n and minimum Hamming distance at least d, where every codeword has weight exactly w. For more details about the state-of-the-art of binary constant-weight codes, please refer to [6], [14], [17].

So far, the exact value of $N(n, k, m; r)$ remains open for most cases when $r \le k-2$ and $n < \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} - (m-k+1)A(m, 4, k-2)$. Thus, this motivates us to study the possibility of improving known lower bounds of $N(n, k, m; r)$.

The following lower bound is due to Zhang *et al.* [29].

[2] A *set system* refers to a pair $(V, \mathcal{C})$, where V is a set of *points* and $\mathcal{C}$ is a collection of subsets of V called *blocks*.

Let

$$\begin{aligned}
C_1 &= \{1,5,8,11,14,17,20,23,26,31\}, & C_2 &= \{2,5,8,11,16,19,21,25,27,29\},\\
C_3 &= \{3,5,10,12,14,19,21,24,28,31\}, & C_4 &= \{4,5,10,12,15,17,20,25,27,30\},\\
C_5 &= \{1,6,9,12,15,18,21,24,27,29\}, & C_6 &= \{2,6,9,12,14,17,22,23,28,30\},\\
C_7 &= \{3,6,8,13,15,17,22,25,26,29\}, & C_8 &= \{4,6,8,13,16,18,21,23,28,31\},\\
C_9 &= \{1,7,10,13,16,19,22,25,28,30\}, & C_{10} &= \{2,7,10,13,15,18,20,24,26,31\},\\
C_{11} &= \{3,7,9,11,16,18,20,23,27,30\}, & C_{12} &= \{4,7,9,11,14,19,22,24,26,29\}.
\end{aligned}$$

Then $\Big([31], \{C_i : i \in [12]\}\Big)$ forms a set system for a $(31,120,6,12;3)$-MCBC.

Its dual set system is $\Big([12], \{B_i : i \in [31]\}\Big)$, where

$$B_1 = \{1,5,9\},\ B_2 = \{2,6,10\},\ B_3 = \{3,7,11\},\ B_4 = \{4,8,12\},$$

and

$$\begin{aligned}
B_5 &= \{1,2,3,4\}, & B_6 &= \{5,6,7,8\}, & B_7 &= \{9,10,11,12\},\\
B_8 &= \{1,2,7,8\}, & B_9 &= \{5,6,11,12\}, & B_{10} &= \{3,4,9,10\},\\
B_{11} &= \{1,2,11,12\}, & B_{12} &= \{3,4,5,6\}, & B_{13} &= \{7,8,9,10\},\\
B_{14} &= \{1,3,6,12\}, & B_{15} &= \{4,5,7,10\}, & B_{16} &= \{2,8,9,11\},\\
B_{17} &= \{1,4,6,7\}, & B_{18} &= \{5,8,10,11\}, & B_{19} &= \{2,3,9,12\},\\
B_{20} &= \{1,4,10,11\}, & B_{21} &= \{2,3,5,8\}, & B_{22} &= \{6,7,9,12\},\\
B_{23} &= \{1,6,8,11\}, & B_{24} &= \{3,5,10,12\}, & B_{25} &= \{2,4,7,9\},\\
B_{26} &= \{1,7,10,12\}, & B_{27} &= \{2,4,5,11\}, & B_{28} &= \{3,6,8,9\},\\
B_{29} &= \{2,5,7,12\}, & B_{30} &= \{4,6,9,11\}, & B_{31} &= \{1,3,8,10\}.
\end{aligned}$$

Fig. 2. The set system of a $(31,120,6,12;3)$-MCBC and its corresponding dual set system.

Theorem 4 (Zhang et al. [29, Theorem 7]): Let $r \le k-1$. For any $r \le c \le k-1$,

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

We remark that the proof of Theorem 4 is based on Lemma 3.2 in Bhattacharya *et al.* [4], and indeed when $r = 1$, Theorem 4 reduces to the latter one. For the convenience of comparison with the proof in this paper, we give the proof sketch of Theorem 4 in the appendix. In this paper, we formulate an integer linear program whose solutions provide lower bounds on $N(n,k,m;r)$. In the following sections, we introduce related results in this area to provide our constraints for the linear program.

C. Multiset Hall's Conditions

To formulate our linear constraints, we follow Zhang *et al.* [29] and consider the dual set system of an MCBC. Specifically, given a set system $(V,\mathcal{C})$ with $V = [n]$ and $\mathcal{C} = \{C_j : j \in X\}$, its *dual set system* is $(X,\mathcal{B})$, where $\mathcal{B} = \{B_1, B_2, \dots, B_n\}$ with $B_j = \{i \in X : j \in C_i\}$. In other words, B_j consists of the indices of servers that store the j-th item. The following theorem characterizes the dual set system of an MCBC.

Theorem 5 (Zhang et al. [29]): 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 $h \in [\lceil \frac{k}{r} \rceil]$, any h distinct blocks $B_{i_1}, B_{i_2}, \dots, B_{i_h} \in \mathcal{B}$, we have $|\bigcup_{j\in[h]} B_{i_j}| \ge \min\{hr,k\}$.

Example 6: The dual set system for the $(31,120,6,12;3)$-MCBC is given in Figure 2. We check for example that the indices of servers that contain the item x_1 are 1, 5, and 9, and thus $B_1 = \{1,5,9\}$.

According to Theorem 5, any two blocks in the dual set system contain at least six distinct points. For example, $B_1 \cup B_2 = \{1,2,5,6,9,10\}$ and $B_5 \cup B_8 = \{1,2,3,4,7,8\}$. In fact, the six distinct points correspond to the six distinct servers that the users contact. □

For the dual set system $(X,\mathcal{B})$, let x_i be the number of blocks in $\mathcal{B}$ of size i for any $i > 0$. Note that for $i < r$, we must have $x_i = 0$ since every item is contained in at least r different servers in an $(n,N,k,m;r)$-MCBC [29, Lemma 1(i)]. As pointed in [21], for an optimal MCBC, we may assume that $x_i = 0$ 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. We state the following simple properties without proof.

Proposition 7: For a set system $(X,\mathcal{B})$, the multiset Hall's condition is equivalent to any of the following:

(i) $|\bigcap_{j\in[h]} \overline{B_{i_j}}| \le m - \min\{hr,k\}$ for any $B_{i_j} \in \mathcal{B}$, where $\overline{B}$ is the *complement* of B in X;

(ii) any $(m - \min\{hr,k\} + 1)$-subset of X is contained in at most $h-1$ blocks in $\overline{\mathcal{B}}$, where $\overline{\mathcal{B}} = \{\overline{B} : B \in \mathcal{B}\}$.

Thus, by counting the occurrence of $(m-\min\{hr,k\}+1)$-subsets in $\overline{\mathcal{B}}$, we have for any $h \in [\lceil \frac{k}{r} \rceil]$:

$$\sum_{i\in[r,k-1]} \binom{m-i}{m-\min\{hr,k\}+1} x_i \le (h-1)\binom{m}{m-\min\{hr,k\}+1}. \tag{2}$$

Especially, when $h = \lceil \frac{k}{r} \rceil = \lfloor \frac{k-1}{r} \rfloor + 1$, we have

$$\sum_{i\in[r,k-1]} \binom{m-i}{k-1-i} x_i \le \left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1}. \tag{3}$$

Remark 8: Inequality (3) was in [29, Lemma 3]. Here, we provide a different proof that may be more intuitive. Furthermore, as in [4, Lemma 2.1], it can be proved that in (2), the inequality for h implies the inequality for $h-1$ for any $h \in [2, \lceil \frac{k}{r} \rceil]$. In other words, it suffices to impose the constraint given by (3) in our linear program. We will show that by solving the linear programming problem with this constraint we are able to provide a new proof of Theorem 4.

D. Uniform MCBCs

The uniform CBCs were introduced as a special class of CBCs by Paterson *et al.* in [21]. In this subsection, we generalize the notion of uniform CBCs to uniform MCBCs. In Section IV, uniform MCBCs play a key role in improving Theorem 4. Specifically, uniform MCBCs impose additional constraints in our optimization problem and hence, improves the corresponding lower bounds.

An $(n, N, k, m; r)$-MCBC is a *c-uniform $(n, cn, k, m; r)$-MCBC* if each item is stored in exactly c servers. In other words, in its dual set system $(X, \mathcal{B})$, we have $|B| = c$ for any $B \in \mathcal{B}$. We denote the maximum n for which there exists a c-uniform $(n, cn, k, m; r)$-MCBC by $n_c(k, m; r)$, and abbreviate it to n_c if there is no confusion.

It follows from Proposition 7 (or proceeding in a similar manner as the proof for uniform CBCs in [21, Theorem 3.1]), we obtain the following inequality:

$$n_c(k,m;r) \le \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}}{\binom{m-c}{k-1-c}} = \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{c}}{\binom{k-1}{c}}. \tag{4}$$

We derive the following Johnson-type bound for $n_c(k, m; r)$.

Theorem 9: $n_c(k,m;r) \le \left\lfloor \frac{m}{m-c} n_c(k, m-1; r) \right\rfloor$.

Proof: Suppose the dual set system of $(X = [m], \mathcal{B})$ is a c-uniform $(n, cn, k, m; r)$-MCBC with $n = n_c(k, m; r)$. By the pigeonhole principle, there must exist some $i' \in [m]$, such that it is not contained in at least $(m-c)n/m$ blocks of $\mathcal{B}$.

Now, we define a new set system in which the points are obtained by removing i' from X, and the block set consists of all the blocks in $\mathcal{B}$ that are not containing i', that is, $(X' = [m] \setminus \{i'\}, \mathcal{B}' = \{B \in \mathcal{B} : i' \notin B\})$. By checking the multiset Hall's condition in Theorem 5, we obtain that: the dual set system of $(X', \mathcal{B}')$ is a c-uniform $(n', cn', k, m-1; r)$-MCBC for some $n' \le n$. Obviously, $n' \le n_c(k, m-1; r)$.

Finally, by the statements above, we have

$$(m-c)n_c(k,m;r)/m = (m-c)n/m \le n' \le n_c(k,m-1;r),$$

and the desired inequation holds since $n_c(k, m; r)$ should be an integer. ■

When $r = 1$, the values of $n_c(k, m; 1)$ have been studied in [2], [4], [11], [21], [25]. It was proved that $n_{k-1}(k, m; 1) = (k-1)\binom{m}{k-1}$ and $n_{k-2}(k, m; 1) = \binom{m}{k-2}$ [21]. The value of $n_2(k, m; 1)$ was determined asymptotically in [2], [21], while for general c, the asymptotic behaviour of $n_c(k, m; 1)$ was studied in [2], [4], [21]. A few constructions of optimal uniform CBCs from affine planes and transversal designs were also given in [25]. When $r > 1$, the value of $n_c(k, m; r)$ was not studied before, but Theorem 5 implies that

$$n_c(k,m;r) = A(m, 2(k-c), c) \text{ when } r \ge \frac{k}{2}. \tag{5}$$

E. Main Contributions and Organizations

In this paper, we formulate integer linear programs so that the optimal solutions provide lower bounds of $N(n, k, m; r)$. Our first set of linear constraints from multiset Hall's condition yields a new proof of Theorem 4 in Section III.

In Section IV, we derive additional constraints from the uniform MCBCs that are naturally imposed on the problem for the optimization, and obtain new lower bounds of $N(n, k, m; r)$. In particular, we demonstrate the following theorem and analyze the cases where Theorem 10 improves Theorem 4.

Theorem 10: Let $n_c(k, m; r)$ be the maximum n such that a c-uniform $(n, cn, k, m; r)$-MCBC exists as defined. Then

$$N(n,k,m;r) \ge n(r+1) - n_r(k,m;r). \tag{6}$$

Also, for $c \in [r+1, k-2]$, we have that $N(n, k, m; r) \ge nk - \lfloor Z \rfloor$, where

$$\begin{aligned} Z = {} & \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}}{\binom{m-c}{k-c}} \cdot \frac{2(m-c)(m-k+1)}{(m-c+1)(m-c)-(k-c)(k-c-1)} \\ & + n\left(k - c - \frac{(m-c+1)(m-c)+(k-c)(k-c-1)}{(m-c+1)(m-c)-(k-c)(k-c-1)}\right) \\ & + n_c(k,m;r)\frac{(m-k+1)(m-k)}{(m-c+1)(m-c)-(k-c)(k-c-1)}. \end{aligned} \tag{7}$$

Using the new set of lower bounds, we provide constructions and determine the exact values of $N(n, 2r, r(r+1); r)$ for $n \in [r+1, 2r+1]$ and $N(n, k, m; k-2)$ for $n \in [A(m, 4, k-2), \binom{m}{k-1} - (m-k+1)A(m, 4, k-2)]$ in Section IV. Thus, combining the results in [29], the value of $N(n, k, m; k-2)$ is completely determined. When $r = 1$, we also analyse the improvement on the lower bounds of $N(m(m-1)/(k-2), k, m)$.

Finally, in Section V, we focus on the special case $(k, r) = (6, 3)$, and formulate a linear program as in the former sections. A lower bound and detailed constructions of codes reaching this bound exactly or asymptotically are both given. In particular, when $m \equiv 0 \pmod 6$, we determine the value of $N(n, 6, m; 3)$ exactly or asymptotically for all n.

We conclude with some open problems in Section VI.

III. Linear Programming With the First Sets of Constraints

In this section, we formulate an integer linear program so that its optimal value yields a lower bound for $N(n, k, m; r)$.

We then solve its linear relaxation and hence, obtain a new proof of Theorem 4.

Let x_i be the number of blocks of size i in the dual set system $(X, \mathcal{B})$ of an $(n, N, k, m; r)$-MCBC. Now, observe that $\sum_{i\in[r,k]} x_i$ yields the number of n, while $\sum_{i\in[r,k]} ix_i$ yields the total storage N. In other words, we have

$$x_k = n - \sum_{i\in[r,k-1]} x_i,$$

and

$$N = \sum_{i\in[r,k]} ix_i = kn - \sum_{i\in[r,k-1]} (k-i)x_i\,.$$

Therefore, minimizing N is equivalent to maximizing the quantity $\sum_{i\in[r,k-1]}(k-i)x_i$, and we are ready to state our optimization problem.

Let $\boldsymbol{A}$ be the following $2\times(k-r)$-matrix

$$\boldsymbol{A} = \begin{bmatrix} \binom{m-r}{k-1-r} & \binom{m-r-1}{k-2-r} & \cdots & 1 \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \tag{8}$$

$\boldsymbol{c} = (k-r, k-r-1, \ldots, 1)$, $\boldsymbol{b} = (\lfloor (k-1)/r \rfloor \binom{m}{k-1}, n)^T$ and $\boldsymbol{x} = (x_r, x_{r+1}, \ldots, x_{k-1})^T$.

Let ξ be the optimal value given by

$$\xi \triangleq \max\{\boldsymbol{cx} : \boldsymbol{Ax} \le \boldsymbol{b},\ x_i \in \mathbb{Z}_{\ge 0} \text{ for } i \in [r, k-1]\}. \tag{9}$$

Following our discussion, we have that $N(n, k, m; r) \ge kn - \xi$. Nevertheless, it is not easy to solve this integer linear program. Hence, we relax the integer constraint and consider its *linear programming (LP) relaxation*. Specifically, set

$$\xi^* \triangleq \max\{\boldsymbol{cx} : \boldsymbol{Ax} \le \boldsymbol{b},\ x_i \ge 0 \text{ for } i \in [r, k-1]\}, \tag{10}$$

and it follows that $\xi \le \xi^*$. Applying the *strong duality* of linear programming (see Chvatal [13]), ξ^* can be computed via its *dual problem*.

$$\xi^* = \min\{\boldsymbol{yb} : \boldsymbol{yA} \ge \boldsymbol{c}, \boldsymbol{y} = (y_1, y_2) \ge \boldsymbol{0}\}. \tag{11}$$

Therefore, we have the following lower bound

$$N(n, k, m; r) \ge kn - \lfloor \xi^* \rfloor\,. \tag{12}$$

It is obvious that any feasible solution $\boldsymbol{y}$ of the dual problem (11) provides an upper bound $\boldsymbol{yb}$ for the primal problem (10). Actually, in the remaining of this section, we solve the dual problem (11) that involves only two variables. In fact, the feasible region, i.e. $\{\boldsymbol{y} : \boldsymbol{yA} \ge \boldsymbol{c}, \boldsymbol{y} \ge \boldsymbol{0}\}$, can be described as in the following lemma. Since the proof only relies on tedious computation, we omit its technical detail here.

Proposition 11: For $i \in [r, k-1]$, set the line $\mathsf{L}_i \triangleq \{\boldsymbol{y} : y_2 + \binom{m-i}{k-1-i} y_1 = k - i\}$. Then the following are true.

(i) The feasible region $\{\boldsymbol{y} : \boldsymbol{yA} \ge \boldsymbol{c}, \boldsymbol{y} \ge \boldsymbol{0}\}$ is bounded by L_i, $i \in [r, k-1]$ and the y_1-, y_2-axes.

(ii) The gradient of L_i is $-\binom{m-i}{k-1-i}$. The gradients are negative and increase with i.

(iii) The vertices of the region are given by $(1, 0)$, $(0, k-r)$ and $(1/\binom{m-i}{k-i}, k-i-(k-i)/(m-k+1))$, the intersection point of L_{i-1} and L_i for $i \in [r+1, k-1]$.

(iv) Furthermore, the intersection point of L_{i-1} and L_{i+1} for $i \in [r+1, k-2]$ lies below L_i when $m > k$, and on L_i when $m = k$.

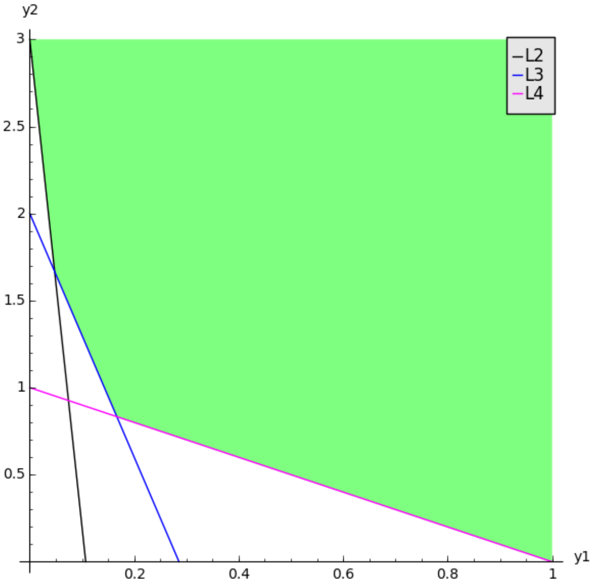

Fig. 3. Feasible region when $n = 20$, $k = 5$, $m = 10$ and $r = 2$.

Figure 3 displays a feasible region when $k = 5$ and $r = 2$.

Therefore, by analysing the gradients, we compute the optimal solutions of (11). Recall that the objective function is given by $\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1} y_1 + ny_2$. We summarize the optimal solutions and their corresponding objective values in Table I.

Therefore, since the optimal objective value ξ in (9) is at most ξ^*, and ξ is an integer, we have the following theorem.

Theorem 12: $N(n, k, m; r) \ge \Xi$, where

$$\Xi \triangleq \begin{cases} nr, \text{ when } n \le \lfloor \frac{k-1}{r} \rfloor \frac{\binom{m}{r}}{\binom{k-1}{r}}; \\ \left\lceil \left(c + \frac{k-c}{m-k+1}\right) n - \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}}{\binom{m-c}{k-c}} \right\rceil, \\ \text{ when } \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{c-1}}{\binom{k-1}{c-1}} < n \le \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{c}}{\binom{k-1}{c}},\ c \in [r+1, k-1]; \\ kn - \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}, \text{ when } n \ge \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}. \end{cases}$$

Remark 13: Notice that we recover Theorem 4 when $c \in [r+1, k-1]$, i.e., $\lfloor \frac{k-1}{r} \rfloor \binom{m}{r} / \binom{k-1}{r} < n \le \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$, while it also yields a new proof of Theorem 4. Actually, when $n \le \lfloor \frac{k-1}{r} \rfloor \binom{m}{r} / \binom{k-1}{r}$ and $n \ge \lfloor \frac{k-1}{r} \rfloor \binom{m}{k-1}$, the corresponding lower bounds can also be obtained from [29, Theorem 2 (i)] and [21, Theorem 2.9] respectively.

In the next section, we provide a new lower bound of $N(n, k, m; r)$ by adding some constraints to the optimization problem (9) and thus obtain an improvement of Theorem 12. Before that, we show an example.

Example 14: When $(n, k, m; r) = (31, 6, 12; 3)$, the integer optimization problem (9) is given by

$$\begin{array}{rrl} \max & 3x_3 + 2x_4 + x_5 & \\ \text{subject to} & 36\ x_3 + 8\ x_4 + x_5 & \le 792, \\ & x_3 + x_4 + x_5 & \le 31, \\ & x_3, x_4, x_5 & \in \mathbb{Z}_{\ge 0}. \end{array}$$

TABLE I

SOLUTION FOR THE DUAL PROBLEM (11)

gradient of objective function $-\left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1}/n$	n	optimal solution $\boldsymbol{y}^* = (y_1^*, y_2^*)$	optimal value ξ^*
at most $-\binom{m-r}{k-1-r}$	at most $\left\lfloor \frac{k-1}{r} \right\rfloor \frac{\binom{m}{r}}{\binom{k-1}{r}}$	$(0, k-r)$	$n(k-r)$
for $c \in [r+1, k-1]$, between $-\binom{m-c+1}{k-c}$ and $-\binom{m-c}{k-1-c}$ (inclusive)	between $\frac{\left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{c-1}}{\binom{k-1}{c-1}}$ and $\frac{\left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{c}}{\binom{k-1}{c}}$ (inclusive)	$\left(\frac{1}{\binom{m-c}{k-c}}, k-c-\frac{k-c}{m-k+1} \right)$	$\frac{\left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1}}{\binom{m-c}{k-c}} + \left(k-c-\frac{k-c}{m-k+1}\right) n$
at least -1	at least $\left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1}$	$(1, 0)$	$\left\lfloor \frac{k-1}{r} \right\rfloor \binom{m}{k-1}$

Solving the dual of its linear relaxation (11) yields $\xi^* \approx 81.429$ (here, the solution corresponds to $c = 4$ in Table I). Hence, we have that $N(31, 6, 12; 3) \geq 6(31) - 81 = 105$. □

IV. IMPROVED LOWER BOUNDS WITH CONSTRAINTS FROM UNIFORM MCBCs

In the former section, we formulate an integer linear program utilizing the constraints derived from the multiset Hall's condition. In this section, we derive new constraints that are imposed from definitions of the uniform MCBCs. We provide certain sufficient conditions where a solution to the new ILP problem is optimal and hence, obtain some new lower bounds and exact values of $N(n, k, m; r)$. Recall that we use n_c to abbreviate $n_c(k, m; r)$ for some fixed k, m, and r whenever there is no confusion.

A. Linear Programming With the Second Sets of Constraints

Given the dual set system $(X, \mathcal{B})$ of an $(n, N, k, m; r)$-MCBC, recall that x_i counts the number of blocks of size i. Clearly, these blocks form the dual set system of an i-uniform $(x_i, ix_i, k, m; r)$-MCBC. Therefore, we have the following additional constraints

$$x_i \leq n_i(k, m; r) \text{ for } i \in [r, k-1]. \tag{13}$$

Therefore, we reformulate our optimization program as such. Let $\boldsymbol{A}$, $\boldsymbol{x}$, $\boldsymbol{c}$ be as defined before. Set

$$\boldsymbol{A}' = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{I}_{k-r} \end{bmatrix},$$

$$\boldsymbol{b}' = \begin{bmatrix} \lfloor (k-1)/r \rfloor \binom{m}{k-1} \\ n \\ n_r \\ n_{r+1} \\ \vdots \\ n_{k-1} \end{bmatrix}.$$

Let ζ be the optimal value given by

$$\zeta \triangleq \max\{\boldsymbol{cx} : \boldsymbol{A}'\boldsymbol{x} \leq \boldsymbol{b}',\, x_i \in \mathbb{Z}_{\geq 0} \text{ for } i \in [r, k-1]\}. \tag{14}$$

As before, we consider the LP relaxation of (14) and its corresponding dual

$$\begin{aligned} \zeta^* &\triangleq \max\{\boldsymbol{cx} : \boldsymbol{A}'\boldsymbol{x} \leq \boldsymbol{b}',\, x_i \geq 0 \text{ for } i \in [r, k-1]\} && (15) \\ &= \min\{\boldsymbol{zb}' : \boldsymbol{zA}' \geq \boldsymbol{c},\, \boldsymbol{z} = (y_1, y_2, z_r, \ldots, z_{k-1}) \geq \boldsymbol{0}\}. && (16) \end{aligned}$$

We have the following chain of inequalities

$$N(n, k, m; r) \geq kn - \lfloor \zeta^* \rfloor \geq kn - \lfloor \xi^* \rfloor. \tag{17}$$

Example 15: When $(n, k, m; r) = (31, 6, 12; 3)$, the integer optimization problem (14) is given by

$$\begin{array}{lrl} \max & 3x_3 + 2x_4 + x_5 & \\ \text{subject to} & 36\ x_3 + 8\ x_4 + x_5 & \leq 792, \\ & x_3 + x_4 + x_5 & \leq 31, \\ & x_3 & \leq 4, \\ & x_4 & \leq 51, \\ & x_5 & \leq 792, \\ & x_3, x_4, x_5 & \in \mathbb{Z}_{\geq 0}. \end{array}$$

(See (20) for the values of n_i for $i \in \{3, 4, 5\}$.) Solving the dual of its linear relaxation (16) yields $\zeta^* = 66$. Hence, $N(31, 6, 12; 3) \geq 6 \times 31 - 66 = 120$. Now, since we have a $(31, 120, 6, 12; 3)$-MCBC in Figure 2, we have that $N(31, 6, 12; 3) = 120$ and that the optimal solution ζ of (14) is 66. □

Instead of providing a closed formula for ζ^*, we propose a few feasible solutions for the dual problem (16) and demonstrate that these solutions are optimal under certain conditions.

To demonstrate the optimality, we appeal to the notion of optimality certificates.

Definition 16: A pair $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ is an *optimality certificate* for (15) and (16) if the following hold. Here, $\boldsymbol{x}^*$ and $\boldsymbol{z}^*$ are some specific assignment of $\boldsymbol{x}$ and $\boldsymbol{z}$.

(i) $\boldsymbol{x}^*$ is feasible solution for the primal problem (15).
(ii) $\boldsymbol{z}^*$ is feasible solution for the dual problem (16).
(iii) The corresponding objective values are the same. In other words, $\boldsymbol{cx}^* = \boldsymbol{z}^*\boldsymbol{b}'$.

Given an optimality certificate, it is then straightforward to obtain the optimal value ζ^*.

Proposition 17 (See Chvatal [13]): If $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ is an optimality certificate for (15) and (16), then $\zeta^* = \boldsymbol{cx}^* = \boldsymbol{z}^*\boldsymbol{b}'$.

B. The First Set of Feasible Solutions

Proposition 18: Set $y_2^* = k - r - 1$, $z_r^* = 1$, and $y_1^* = z_{r+1}^* = z_{r+2}^* = \cdots = z_{k-1}^* = 0$. Then $\boldsymbol{z}^* = (y_1^*, y_2^*, z_r^*, \ldots, z_{k-1}^*)$ is a feasible solution for (16) and so, $\zeta^* \leq n(k - r - 1) + n_r$.

Furthermore, whenever

$$n_r \le n \le \min\left\{\frac{\left\lfloor\frac{k-1}{r}\right\rfloor\binom{m}{k-1}}{\binom{m-r-1}{k-r-2}} - \frac{(m-k+1)n_r}{k-r-1}, n_r + n_{r+1}\right\}, \tag{18}$$

the solution $\boldsymbol{z}^*$ is also optimal and we have that

$$\zeta^* = n(k-r-1) + n_r.$$

Proof: It is straightforward to verify that $\boldsymbol{z}^*\boldsymbol{A}' \ge \boldsymbol{c}$ and hence, $\boldsymbol{z}^*$ is feasible.

To show that $\boldsymbol{z}^*$ is optimal, we then produce an optimality certificate. Let $x_r^* = n_r$, $x_{r+1}^* = n - n_r$, and $x_j^* = 0$ for $j \in [r+2, k-1]$. Then (18) implies that $\boldsymbol{A}'\boldsymbol{x}^* \le \boldsymbol{b}'$. In other words, $\boldsymbol{x}^*$ is a feasible solution for (15). It is also straightforward to verify that $\boldsymbol{c}\boldsymbol{x}^* = \boldsymbol{z}^*\boldsymbol{b}'$.

Therefore, $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ is an optimality certificate and the value of ζ^* follows from Proposition 17. ■

The bound (6) in Theorem 10 is therefore immediate from Proposition 18. In what follows, we construct MCBCs whose size attains the lower bound (6). In other words, we show that the lower bound (6) is in fact tight in some cases.

Construction A: For any $r \ge 1$, let $k = 2r$, $m = r(r+1)$, and $r+1 \le n \le 2r+1$. In this case, we have $n_r(2r, r(r+1); r) = A(r(r+1), 2r, r)$ by (5). Since it is well known that $A(m, 2w, w) = \lfloor m/w \rfloor$, we have $n_r(2r, r(r+1); r) = r+1$. Set $X = [r+1] \times [r]$ and

$$\mathcal{B}_r \triangleq \{\{i\} \times [r] : i \in [r+1]\},$$
$$\mathcal{B}_{r+1} \triangleq \{[r+1] \times \{j\} : j \in [r]\}.$$

Observe that $\mathcal{B}_r$ is a collection of $r+1$ disjoint blocks of size r, while $\mathcal{B}_{r+1}$ is a collection of r disjoint blocks of size $r+1$. Also, any block in $\mathcal{B}_r$ intersects with any block in $\mathcal{B}_{r+1}$ at most one point. Let $\mathcal{B}$ be a collection of blocks from $\mathcal{B}_r$ and any $n-r-1$ blocks from $\mathcal{B}_{r+1}$. By Theorem 5, we have that the dual set system of $(X, \mathcal{B})$ is an $(n, (n-1)(r+1), 2r, r(r+1); r)$-MCBC. Therefore, it follows from (6) that when $n \in [r+1, 2r+1]$

$$N(n, 2r, r(r+1); r) = (n-1)(r+1).$$

Hence, Theorem 10 yields the exact value in some cases. Note that, in Construction A, the blocks of size $r+1$ need not be disjoint. Thus, the choice of more blocks of size $r+1$ is possible, which leads to larger n that reaches the bound. We will see this from the construction for $(k, r) = (6, 3)$ in Section V.

Next, we consider the case for $r = k-2$. The previous bound in Theorem 12 states that

$$N(n,k,m;k-2) \ge \begin{cases} (k-2)n, \ \text{when } n \le \frac{1}{k-1}\binom{m}{k-2}; \\ (k-1)n - \left\lfloor \frac{\binom{m}{k-1}-n}{m-k+1} \right\rfloor, \\ \quad \text{when } \frac{1}{k-1}\binom{m}{k-2} < n \le \binom{m}{k-1}; \\ kn - \binom{m}{k-1}, \ \text{when } n \ge \binom{m}{k-1}. \end{cases}$$

In [29], codes attaining this bound were constructed when $n \le A(m,4,k-2)$ and $n \ge \binom{m}{k-1} - (m-k+1)A(m,4,k-2)$. Hence, the exact value of $N(n,k,m;k-2)$ has been determined for this range and thus, it remains to consider the range $A(m,4,k-2) < n < \binom{m}{k-1} - (m-k+1)A(m,4,k-2)$.

Now, in this range, the proof of Proposition 18 states that the optimization problem (15) attains maximum when $x_{k-2}^* = n_{k-2}$ and $x_{k-1}^* = n - n_{k-2}$. Hence, this hints that if the bound in Proposition 18 is attained, then the dual set system comprises blocks of size $k-2$ and $k-1$ only. Furthermore, of these n blocks, exactly $n_{k-2} = A(m,4,k-2)$ blocks are of size $k-2$. Now, a natural choice is to pick these n_{k-2} blocks to correspond to blocks in the dual set system of an optimal $(k-2)$-uniform $(n_{k-2}, (k-2)n_{k-2}, k, m; k-2)$-MCBC. Then we proceed to augment the set system with blocks of size $k-1$ so that the multiset Hall's condition (Theorem 5) is met. It turns out that this is always possible in our range of interest.

Construction B: Assume $n_{k-2} \le n \le \binom{m}{k-1} - (m-k+1)n_{k-2}$. Set $X = [m]$ and

$$\mathcal{B}_{k-2} \triangleq \text{blocks of the dual set system of an optimal}$$
$$(k-2)\text{-uniform}\,(n_{k-2}, (k-2)n_{k-2}, k, m; k-2)\text{-MCBC},$$
$$\mathcal{B}_{k-1} \triangleq \{B \subset X : |B| = k-1,\ B \not\supset B' \text{ for } B' \in \mathcal{B}_{k-2}\}.$$

So, the size of $\mathcal{B}_{k-2}$ is n_{k-2}, while the size of $\mathcal{B}_{k-1}$ is $\binom{m}{k-1} - (m-k+2)n_{k-2}$. Moreover, the union of any block in $\mathcal{B}_{k-2}$ with any block in $\mathcal{B}_{k-1}$ contains at least k points. Let $\mathcal{B}$ be a collection of all blocks from $\mathcal{B}_{k-2}$ and any $n - n_{k-2}$ blocks from $\mathcal{B}_{k-1}$. By Theorem 5, we have that the dual set system of $(X, \mathcal{B})$ is an $(n, n(k-1) - n_{k-2}, k, m; k-2)$-MCBC. Therefore, it follows from (6) that in this case

$$N(n,k,m;k-2) = n(k-1) - n_{k-2}.$$

Thus, combining the results in [29], the exact value of $N(n,k,m;k-2)$ is determined for all values of n in Theorem 19, shown at the bottom of the next page.

C. The Second Set of Feasible Solutions

Proposition 20: For $c \in [r+1, k-2]$, consider the following 3×3-submatrix $\boldsymbol{M}$ of $\boldsymbol{A}'$,

$$\boldsymbol{M} = \begin{bmatrix} \binom{m-c+1}{k-c} & \binom{m-c}{k-c-1} & \binom{m-c-1}{k-c-2} \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Set $z_j^* = 0$ for $j \ne c$ and let y_1^*, y_2^* and z_c^* be the unique solution of

$$(y_1^*, y_2^*, z_c^*)\boldsymbol{M} = (k-c+1, k-c, k-c-1).$$

Then $\boldsymbol{z}^* = (y_1^*, y_2^*, z_r^*, \ldots, z_{k-1}^*)$ is a feasible solution for (16).

Furthermore, let x_{c-1}^*, x_c^* and x_{c+1}^* be the unique solution of

$$\boldsymbol{M}\begin{bmatrix} x_{c-1}^* \\ x_c^* \\ x_{c+1}^* \end{bmatrix} = \begin{bmatrix} \left\lfloor\frac{k-1}{r}\right\rfloor\binom{m}{k-1} \\ n \\ n_c \end{bmatrix}$$

If $0 \le x_{c-1}^* \le n_{c-1}$ and $0 \le x_{c+1}^* \le n_{c+1}$, then $\boldsymbol{z}^*$ is optimal.

Proof: We use the observations in Proposition 11 to provide a lower bound for $\boldsymbol{z}^*\boldsymbol{A}'$. By definition, the point (y_1^*, y_2^*) is the intersection of the two lines L_{c-1} and L_{c+1}. It follows from Proposition 11 that the point (y_1^*, y_2^*) lies above the line L_i for $i \in [c-2]$ and for $i \in [c+2, k-1]$, but not above the line L_c. In other words,

$$y_1^*\binom{m-i}{k-1-i} + y_2^* \ge k-i \text{ for } i \ne c,$$

and

$$y_1^*\binom{m-c}{k-1-c} + y_2^* \le k-c. \tag{19}$$

Furthermore when $i = c$, the i-th entry of $\boldsymbol{z}^*\boldsymbol{A}'$ is given by $y_1^*\binom{m-c}{k-1-c} + y_2^* + z_c^* = k-c$. Therefore, $\boldsymbol{z}^*\boldsymbol{A}' \ge \boldsymbol{c}$ holds as desired. By the property of intersection point, $y_1^* \ge 0$ and $y_2^* \ge 0$, and by (19) $z_c^* \ge 0$, and thus $\boldsymbol{z}^* \ge 0$ also holds.

When $0 \le x_{c-1}^* \le n_{c-1}$ and $0 \le x_{c+1}^* \le n_{c+1}$, we construct an optimality certificate. In addition to x_{c-1}^*, x_c^* and x_{c+1}^*, we set $x_j^* = 0$ when $j \notin [c-1, c+1]$. Then it is straightforward to verify that $\boldsymbol{A}'\boldsymbol{x}^* \le \boldsymbol{b}'$ and so, $\boldsymbol{x}^*$ is feasible.

Finally, it remains to show $\boldsymbol{c}\boldsymbol{x}^* = \boldsymbol{z}^*\boldsymbol{b}'$. Indeed, following standard manipulations,

$$\begin{aligned}\boldsymbol{z}^*\boldsymbol{b}' &= (y_1^*, y_2^*, z_c^*)\begin{bmatrix}\left\lfloor\frac{k-1}{r}\right\rfloor\binom{m}{k-1}\\ n\\ n_c\end{bmatrix}\\ &= (y_1^*, y_2^*, z_c^*)\boldsymbol{M}\begin{bmatrix}x_{c-1}^*\\ x_c^*\\ x_{c+1}^*\end{bmatrix}\\ &= (k-c+1, k-c, k-c-1)\begin{bmatrix}x_{c-1}^*\\ x_c^*\\ x_{c+1}^*\end{bmatrix} = \boldsymbol{c}\boldsymbol{x}^*.\end{aligned}$$

Hence, $\boldsymbol{z}^*$ is an optimal solution. ■

Solving for y_1^*, y_2^* and z_c^* and with some algebraic manipulations, we obtain (7) in Theorem 10. However, it is unclear whether Theorem 10 provides a better bound as compared to Theorem 4. Nevertheless, Proposition 20 provides the conditions whereby (7) yields the value ζ^*. Since $\zeta^* \le \xi^*$, we have that Theorem 10 is not worse than Theorem 4 in these cases.

In the following example, we compare the two lower bounds and show that Theorem 10 improves the lower bound by a strictly positive quantity.

Example 21: Fix $k \ge 5$ and $r = 1$. Let $n = m(m-1)/(k-2) = m^2/(k-2) + o(m^2)$ and we study the lower bound for $N(n,k,m)$ as a function of m.

In this case, the optimal value ξ^* in (11) is given by

$$\begin{aligned}\xi^* &= \frac{(k-1)\binom{m}{k-1}}{\binom{m-2}{k-2}} + \left(k-2-\frac{k-2}{m-k+1}\right)n\\ &= m(m-1) = m^2 + o(m^2).\end{aligned}$$

Therefore, Theorem 4 implies that $N(n,k,m) \ge nk - m^2 - o(m^2)$.

In contrast, we estimate the quantity given by (7). Now, Balachandran and Bhattacharya estimated $n_2(k,m;1)$ as follows [2, Theorems 4.1, 6.2, Corollary 6.4].

(i) $n_2(5,m;1) = \left\lfloor\frac{m^2}{4}\right\rfloor = \frac{m^2}{4} + o(m^2)$.

(ii) For $k \in \{6,7,8\}$, $n_2(k,m;1) = \Theta\left(m^{3/2}\right) = o(m^2)$.

(iii) For $k \ge 9$, $n_2(k,m;1) = O\left(m^{1+1/\lfloor k/4\rfloor}\right) = o(m^2)$.

Setting $c = 2$ in (7), we have

$$\begin{aligned}Z &= n(k-3) + n_2(k,m;1) + o(m^2)\\ &= \begin{cases}\frac{11}{12}m^2 + o(m^2), & \text{for } k = 5,\\ \frac{k-3}{k-2}m^2 + o(m^2), & \text{for } k \ge 6.\end{cases}\end{aligned}$$

Since $N(n,k,m) \ge nk - \lfloor Z\rfloor$, we have that Theorem 10 improves the lower bound by a quantity of $m^2/12 + o(m^2)$ when $k = 5$ and $m^2/(k-2) + o(m^2)$ when $k \ge 6$. □

Furthermore, when $k = 5$, the new lower bound can be reached asymptotically.

Construction C: Let $k = 5$ and $r = 1$. For $m \ge 5$, set $n = m(m-1)/3$ with $m \not\equiv 2 \pmod 3$. Set $X = [m]$, $X_1 = \Big[\lfloor m/2\rfloor\Big]$, $X_2 = \Big[\lfloor m/2\rfloor + 1, m\Big]$, and

$$\begin{aligned}\mathcal{B}_2 &\triangleq \{\{i,j\} : \ i \in X_1, j \in X_2\},\\ \mathcal{B}_3 &\triangleq \{|B| = 3 : B \subseteq X_1 \text{ or } B \subseteq X_2\}.\end{aligned}$$

So, the size of $\mathcal{B}_2$ is $\lfloor m^2/4\rfloor$ and in fact, the dual set system of $(X = [m], \mathcal{B}_2)$ is an optimal 2-uniform $(\lfloor m^2/4\rfloor, 2\lfloor m^2/4\rfloor, 5, m)$-CBC. Now, the size of $\mathcal{B}_3$ is $\binom{\lfloor m/2\rfloor}{3} + \binom{\lceil m/2\rceil}{3}$, which is greater than $n - \lfloor m^2/4\rfloor$. Moreover, the union of h blocks in $\mathcal{B}_2 \cup \mathcal{B}_3$ contains at least h points for $h \in [5]$. Let $\mathcal{B}$ be a collection of all blocks from $\mathcal{B}_{k-2}$ and any $n - \lfloor m^2/4\rfloor$ blocks from $\mathcal{B}_{k-1}$. By Theorem 5, we have that the dual set sytem of $(X, \mathcal{B})$ is an $(n, N, 5, m)$-CBC with $N = (3/4)m^2 + o(m^2)$.

Therefore, it follows from Example 21 that $N(m(m-1)/3, 5, m) = (3/4)m^2 + o(m^2)$.

Remark 22: To formulate the optimization program (15), we need to provide estimates for $n_c(k,m;r)$. While it is not clear on how we can constrain $n_c(k,m;r)$, we note that certain constraints do not provide better results. Applying the bound of $n_c(k,m;r)$ from (4) as a new constraint in (13) does not yield a better lower bound for $N(n,k,m;r)$. This is because (4) is implied by (3), a constraint that is in the linear program (10). Nevertheless, the Johnson-type bound and bounds from constant weight codes may be applied to obtain better bounds, as illustrated in this section and the next.

Theorem 19: Let $k \ge 4$.

$$N(n,k,m;k-2) = \begin{cases}(k-2)n, & \text{if } n \le A(m,4,k-2);\\ (k-1)n - A(m,4,k-2), & \text{if } A(m,4,k-2) \le n \le \binom{m}{k-1} - (m-k+1)A(m,4,k-2);\\ (k-1)n - \left\lfloor\frac{\binom{m}{k-1}-n}{m-k+1}\right\rfloor, & \text{if } \binom{m}{k-1} - (m-k+1)A(m,4,k-2) \le n \le \binom{m}{k-1};\\ kn - \binom{m}{k-1}, & \text{if } n \ge \binom{m}{k-1}.\end{cases}$$

V. The Case $(k, r) = (6, 3)$

In the previous section, we included constraints arising from uniform MCBCs to improve the known lower bound, and subsequently, obtain exact values in certain cases, especially when $r = k - 2$. In this section, we study the next open case where $r = k - 3$ and focus on the special instance where $(k, r) = (6, 3)$. Furthermore, for this set of parameters, we show Construction A can be improved. We formulate an LP problem as in the former sections, and also provide constructions reaching the lower bound of $N(n, 6, m; 3)$ exactly or asymptotically.

A. A Lower Bound of $N(n, 6, m; 3)$

Firstly, in this case (5) provides the following values of $n_c(6, m; 3)$ for $c \in \{3, 4, 5\}$:

$$\begin{aligned} n_3(6, m; 3) &= A(m, 6, 3) = \left\lfloor \frac{m}{3} \right\rfloor, \\ n_4(6, m; 3) &= A(m, 4, 4), \\ n_5(6, m; 3) &= A(m, 2, 5) = \binom{m}{5}, \end{aligned} \tag{20}$$

where the value of $A(m, 4, 4)$ was determined as below.

Lemma 23 (Bao and Ji [3]):

$$A(m,4,4) = \begin{cases} \frac{m(m-1)(m-2)}{24}, & \text{if } m \equiv 2, 4 \pmod 6; \\ \frac{m(m-1)(m-3)}{24}, & \text{if } m \equiv 1, 3 \pmod 6; \\ \frac{m(m^2-3m-6)}{24}, & \text{if } m \equiv 0 \pmod 6; \\ \frac{m^3-4m^2+m-6}{24}, & \text{if } m \equiv 5 \pmod{12}; \\ \frac{m^3-4m^2+m-18}{24}, & \text{if } m \equiv 11 \pmod{12}. \end{cases}$$

To obtain a tighter bound, we also derive the following constraint from the property of codes with this parameter.

Proposition 24: Let $(X, \mathcal{B})$ be the dual set system of an $(n, N, 6, m; 3)$-MCBC. For $i \in \{3, 4\}$, recall that x_i is the number of blocks in $\mathcal{B}$ of size i. Then

$$(3m - 8)x_3 + 4x_4 \le \binom{m}{3}.$$

Proof: For $i \in \{3, 4\}$, let $\mathcal{B}_i \subseteq \mathcal{B}$ comprise the blocks of size i. By the multiset Hall's condition in Theorem 5, in an $(n, N, 6, m; 3)$-MCBC, any two blocks in $\mathcal{B}_3$ are disjoint; any two blocks in $\mathcal{B}_4$ intersect at most two points; and any block in $\mathcal{B}_3$ intersect in at most one point with any block in $\mathcal{B}_4$.

For any $B \in \mathcal{B}_3$, let $\widehat{B}$ consist of all the 3-subsets of $[m]$ in which there are two elements from B and one element from $[m] \setminus B$. Hence, $|\widehat{B}| = 3(m - 3)$. Let $\mathcal{B}_4^3$ consist of all the 3-subsets of blocks in $\mathcal{B}_4$, and $|\mathcal{B}_4^3| = 4x_4$. By the properties stated in the former paragraph, we have that there exists no 3-subset of $[m]$ that appears twice in $\mathcal{B}_3 \cup \{\widehat{B} : B \in \mathcal{B}_3\} \cup \mathcal{B}_4^3$. Thus, by counting the number of triples in $[m]$, we get the desired inequality. ■

Here, we explicitly state our optimization problem obtained by taking the new constraint into account in (15).

$$\begin{array}{lrl} \max & 3x_3 + 2x_4 + x_5 & \\ \text{subject to} & \binom{m-3}{2}x_3 + (m-4)x_4 + x_5 & \le \binom{m}{5} \\ & x_3 + x_4 + x_5 & \le n \\ & x_3 & \le \left\lfloor \frac{m}{3} \right\rfloor \\ & x_4 & \le A(m,4,4) \\ & x_5 & \le \binom{m}{5} \\ & (3m-8)x_3 + 4x_4 & \le \binom{m}{3} \\ & x_3, x_4, x_5 & \ge 0 \end{array}$$

The corresponding dual problem is hence

$$\begin{array}{lrl} \min & \binom{m}{5}y_1 + ny_2 + \left\lfloor \frac{m}{3} \right\rfloor z_3 + A(m,4,4)z_4 & \\ & +\binom{m}{5}z_5 + \binom{m}{3}w & \\ \text{subject to} & \binom{m-3}{2}y_1 + y_2 + z_3 + (3m-8)w & \ge 3 \\ & (m-4)y_1 + y_2 + z_4 + 4w & \ge 2 \\ & y_1 + y_2 + z_5 & \ge 1 \\ & y_1, y_2, z_3, z_4, z_5, w & \ge 0 \end{array} \tag{21}$$

By judiciously choosing certain feasible solutions of (21), we obtain the following lower bound for $N(n, 6, m; 3)$.

Theorem 25: Let $m \ge 6$.

$$N(n,6,m;3) \ge \max \left\{ \begin{array}{c} 3n, \\ 4n - \left\lfloor \frac{m}{3} \right\rfloor, \\ \left\lceil \frac{12m-44}{3m-12}n - \binom{m}{3}\frac{1}{3m-12} \right\rceil, \\ \left\lceil 5n - \frac{3m-16}{3m-8}A(m,4,4) - \frac{2}{3m-8}\binom{m}{3} \right\rceil, \\ \left\lceil \frac{5m-24}{m-5}n - \frac{1}{m-5}\binom{m}{5} \right\rceil, \\ 6n - \binom{m}{5}. \end{array} \right\}$$

Proof: The following six specific assignments gives feasible solutions $(y_1^*, y_2^*, z_3^*, z_4^*, z_5^*, w^*)$ of (21).

- $y_2^* = 3$ and $y_1^* = z_3^* = z_4^* = z_5^* = w^* = 0$;
- $y_2^* = 2$, $z_3^* = 1$ and $y_1^* = z_4^* = z_5^* = w^* = 0$;
- $y_2^* = \frac{6m-28}{3m-12}$, $w^* = \frac{1}{3m-12}$, and $y_1^* = z_3^* = z_4^* = z_5^* = 0$;
- $y_2^* = 1$, $z_4^* = \frac{3m-16}{3m-8}$, $w^* = \frac{2}{3m-8}$, and $y_1^* = z_3^* = z_5^* = 0$;
- $y_1^* = \frac{1}{m-5}$, $y_2^* = \frac{m-6}{m-5}$, and $z_3^* = z_4^* = z_5^* = w^* = 0$;
- $y_1^* = 1$ and $y_2^* = z_3^* = z_4^* = z_5^* = w^* = 0$.

To obtain the above feasible solutions, we assign certain variables to be zero, and certain constraints to be equalities[3]. For example, assigning $y_1^* = z_3^* = z_4^* = z_5^* = 0$, and the first two constraints to be equalities, results in the third one above. As in Section IV, if W is the objective value of (21) obtained from one feasible solution, it then follows that $N(n, 6, m; 3) \ge \lceil 6n - W \rceil$, and thus the desired bound is obtained via straightforward calculations. ■

[3] In fact, these feasible solutions are known as *basic* solutions in linear programming literature. We refer the interested reader to Chvatal [13] for details.

B. Constructions of the $(n, N, 6, m; 3)$-MCBC

First, we recall that the exact value of $N(n, 6, m; 3)$ has been determined for some specific ranges of n.

Lemma 26 (Zhang et al. [29, Theorem 3]):

$$N(n,6,m;3) = \begin{cases} 3n, \text{ if } n \le \left\lfloor \frac{m}{3} \right\rfloor; \\ \left\lceil \frac{5m-24}{m-5} n - \frac{1}{m-5}\binom{m}{5} \right\rceil, \\ \ \text{if } n \in \left[\binom{m}{5} - (m-5)A(m,4,4), \binom{m}{5}\right]; \\ 6n - \binom{m}{5}, \text{ if } n \ge \binom{m}{5}. \end{cases}$$

In the remainder of this section, we provide constructions of codes reaching the bound in Theorem 25 exactly or asymptotically for the remaining range of n, that is, $n \in \left[\lfloor m/3 \rfloor, \binom{m}{5} - (m-5)A(m,4,4)\right]$. It is easy to check that $N(n, 6, 6; 3)$ has been determined for all n by Lemma 26, and therefore we only consider $m > 6$.

1) When $n \in \left[A(m,4,4), \binom{m}{5} - (m-5)A(m,4,4)\right]$: In this range, we have the following statements of $N(n, 6, m; 3)$.

Theorem 27: Let $m > 6$ and $n \in [A(m, 4, 4), \binom{m}{5} - (m-5)A(m, 4, 4)]$.

(i) $N(n, 6, m; 3) = 5n - A(m, 4, 4)$ whenever m is even and $m \notin \{12, 18, 24\}$.

(ii) $N(n, 6, m; 3) \in \{5n - A(m, 4, 4) - 1, 5n - A(m, 4, 4)\}$ when $m \in \{12, 18, 24\}$.

(iii) $N(n, 6, m; 3) = 5n - A(m, 4, 4) - O(m)$ whenever m is odd.

Proof: Similarly as in Construction B, we can construct an $(n, 5n - A(m, 4, 4), 6, m; 3)$-MCBC for this range of n, and Theorem 25 implies that $N(n, 6, m; 3) \ge \left\lceil 5n - \frac{3m-16}{3m-8} A(m, 4, 4) - \frac{2}{3m-8}\binom{m}{3} \right\rceil$. Hence, the gap between the upper bound $5n - A(m, 4, 4)$ and this lower bound of $N(n, 6, m; 3)$ is given by

$$\left\lfloor \frac{m(m-1)(m-2) - 24A(m,4,4)}{3(3m-8)} \right\rfloor \tag{22}$$

When $m \equiv 2, 4 \pmod 6$, the gap (22) is zero and hence, the bound is tight. When $m \equiv 0 \pmod 6$, the gap (22) reduces to $\left\lfloor \frac{8m}{3(3m-8)} \right\rfloor$. This value is zero when $m \ge 30$ and is one when $m \in \{12, 18, 24\}$. Therefore, we obtain the first two statements. Finally, when m is odd, the gap (22) reduces to $\left\lfloor \frac{m^2+O(m)}{3(3m-8)} \right\rfloor = O(m)$. This yields the last statement. ■

2) When $\lfloor m/3 \rfloor < n < A(m,4,4)$: To simplify our exposition, we focus on the case when $m \equiv 0 \pmod 6$, and the cases for other congruence classes will be left for future research. To this end, we introduce the following combinatorial structure from design theory, which has close connection with the optimal constant weight code of weight 4 and minimum distance 4.

Definition 28: An $\mathrm{H}(g^u)$-*design* is a triple $(X, \mathcal{G}, \mathcal{B}_H)$ such that:

(i) X is a set of gu points;

(ii) $\mathcal{G}$ is partition of X into u subsets of size g, called *groups*;

(iii) $\mathcal{B}_H$ is a collection of 4-subsets of X, called *blocks*, such that each block B in $\mathcal{B}_H$ intersects any group $G \in \mathcal{G}$ in at most one point;

(iv) every 3-subset of X that intersects every $G \in \mathcal{G}$ in at most one point is contained in exactly one block $B \in \mathcal{B}_H$.

It was proved that [20]: an $\mathrm{H}(3^{m/3})$ design exists if and only if $m \equiv 0 \pmod 6$. Let $(X, \mathcal{G}, \mathcal{B}_H)$ be an $\mathrm{H}(3^{m/3})$ design with $\delta_m \triangleq m(m-3)(m-6)/24$ blocks. Notice that the dual set system of $(X, \mathcal{B}_H)$ is a 4-uniform $(n, 4n, 6, m; 3)$-MCBC with $n = \delta_m$, but it is not optimal with respect to n because $\delta_m < n_4(6, m; 3) = A(m, 4, 4)$. Now, we provide a construction of an $(n, N, 6, m; 3)$-MCBC when $m \equiv 0 \pmod 6$ utilizing the H-designs.

Construction D: Let $m \equiv 0 \pmod 6$ and suppose that $n \in [m/3, m/3 + \delta_m]$. Set $X = [m]$ and let $(X, \mathcal{G}, \mathcal{B}_H)$ be an $\mathrm{H}(3^{m/3})$ design of size δ_m. We take the collection of groups of size three in $\mathcal{G}$ and any $n - m/3$ blocks of $\mathcal{B}_H$ as $\mathcal{B}$. Then the dual set system of $(X, \mathcal{B})$ is an $(n, 4n - m/3, 6, m; 3)$-MCBC by Theorem 5, and hence by Theorem 25 we have

$$N(n,6,m;3) = 4n - \frac{m}{3} \text{ when } n \in \left[\frac{m}{3}, \frac{m}{3} + \delta_m\right].$$

Theorem 29: Let $m \equiv 0 \pmod 6$ and $m > 6$.

$$N(n,6,m;3) = \begin{cases} 4n - \frac{m}{3}, \text{ if } \frac{m}{3} \le n \le \frac{m^3-9m^2+26m}{24}; \\ 4n - O(m), \\ \text{if } \frac{m^3-9m^2+26m+24}{24} \le n \le \frac{m^3-3m^2-16m+72}{24}; \\ 4n, \text{ if } \frac{m^3-3m^2-16m}{24} + 4 \le n \le A(m,4,4); \end{cases}$$

Proof: When $n \in [m/3, (m^3 - 9m^2 + 26m)/24]$, the value of $N(n, 6, m; 3)$ is from Construction D. For $n \in [m/3 + \delta_m + 1, A(m, 4, 4)]$, (5) implies that there exists a 4-uniform $(n, 4n, 6, m; 3)$-MCBC, and therefore $N(n, 6, m; 3) \le 4n$. On the other hand, we have the lower bound $N(n, 6, m; 3) \ge \left\lceil \frac{12m-44}{3m-12} n - \binom{m}{3}\frac{1}{3m-12} \right\rceil$ from Theorem 25. Hence, the gap between the upper and lower bounds is $\lfloor (\binom{m}{3} - 4n)/(3m - 12) \rfloor$. We analyze this gap for two ranges of n.

- When $(m^3 - 3m^2 - 16m)/24 + 4 \le n \le A(m, 4, 4)$,

$$\left\lfloor \frac{\binom{m}{3} - 4n}{3m-12} \right\rfloor < 1.$$

Hence, the gap is zero and upper bound is tight.

- When $m/3 + \delta_m + 1 \le n \le (m^3 - 3m^2 - 16m)/24 + 3$,

$$\left\lfloor \frac{\binom{m}{3} - 4n}{3m-12} \right\rfloor \le \left\lfloor \frac{m^2 - 4m - 4}{3m-12} \right\rfloor = O(m).$$

Hence, the gap is $O(m)$ while the upper bound is $4n = \Omega(m^3)$. In other words, the bound is asymptotically tight. ■

Now, when $m/3 + \delta_m + 1 \le n \le (m^3 - 3m^2 - 16m)/24 + 3$, we may improve the upper bound $4n$ of $N(n, 6, m; 3)$ for certain parameters. We show this by some examples.

Example 30: When $n = m/3 + \delta_m + 1 = (m^3 - 9m^2 + 26m)/24 + 1$, we let $\mathcal{B}$ be the union of $\mathcal{G}$, $\mathcal{B}_H$ and one 5-subset of $[m]$, which does not contain any block in $\mathcal{G} \cup \mathcal{B}_H$. By checking the multiset Hall's condition in Theorem 5, the dual set system of $(X, \mathcal{B})$ is an $(n, N, 6, m; 3)$-MCBC with $N = m + 4\delta_m + 5 = (m^3 - 9m^2)/6 + 4m + 5$. We can check

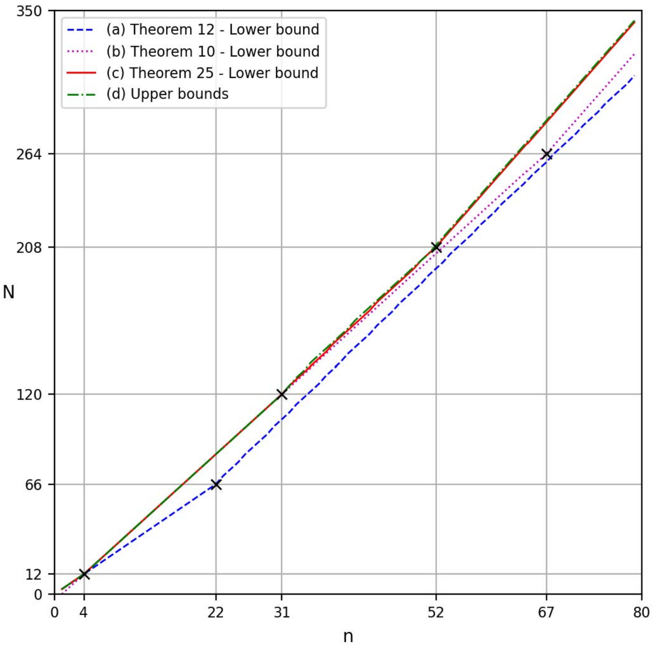

Fig. 4. Comparison of the upper and lower bounds of $N(n,6,12;3)$ for $n<80$. (a) In the ranges of $n \le 22$ and $n > 22$, the segments represent the lower bounds $3n$ and $\lceil \frac{30n}{7} - \frac{198}{7} \rceil$, which can also be obtained from [29, Theorem 2 (i)] and Theorem 4 ([29, Theorem 7]) respectively, as stated in Remark 13. (b) In the ranges of $n \le 67$ and $n > 67$, the segments represent the lower bounds $4n-4$ and $\lceil \frac{177}{35}n - \frac{531}{7} \rceil$ respectively. (c) When n ranges in the intervals $[1,4]$, $[4,31]$, $[31,52]$, and $[52,79]$, the segments represent the lower bounds $3n$, $4n-4$, $\lceil \frac{25}{6}n - \frac{55}{6} \rceil$, and $\lceil 5n - \frac{365}{7} \rceil$ respectively. (d) When $n \le 4$, the upper bounds are obtained from Lemma 26 ([29, Theorem 3]), and when $n > 4$, the upper bounds are obtained from the constructions in Section V-B.

that it reaches the lower bound $\left\lceil \frac{12m-44}{3m-12}n - \binom{m}{3}\frac{1}{3m-12} \right\rceil$ by calculation. □

For other values of n, the construction depends on the H$(3^{m/3})$ design, and we illustrate via an example for $m=12$.

Example 31: Let $m = 12$, and $(X, \mathcal{G}, \mathcal{B}_H)$ be an H(3^4) design with $X = \{i_j : i \in [4], j \in [3]\}$ and $\mathcal{G} = \big\{\{i_1, i_2, i_3\} : i \in [4]\big\}$. For the following values of n, we choose the block collection $\mathcal{B}$ as follows such that the dual set system of $(X, \mathcal{B})$ forms an $(n, N, 6, 12; 3)$-MCBC.

Now, we set $\mathcal{B} = \mathcal{G} \cup \mathcal{B}_H$ and the current size of $\mathcal{B}$ is 31. In fact, if we relabel the points in X according to the following rule:

$$1_1 \mapsto 1,\quad 1_2 \mapsto 5,\quad 1_3 \mapsto 9,\quad 2_1 \mapsto 2,\quad 2_2 \mapsto 6,\quad 2_3 \mapsto 10,$$
$$3_1 \mapsto 3,\quad 3_2 \mapsto 7,\quad 3_3 \mapsto 11,\quad 4_1 \mapsto 4,\quad 4_2 \mapsto 8,\quad 4_3 \mapsto 12,$$

then we recover the dual system of the $(31, 120, 6, 12; 3)$-MCBC given in Figure 2.

When we increase the number of data items n, we can make the following modifications.

(i) When $n \in \{33, 34\}$, remove the two blocks $\big\{\{i_1, i_2, i_3\} : i \in [2]\big\}$ from $\mathcal{B}$, and add any 4 or 5 blocks from below to obtain the desired set of blocks.

$$\{1_j, 1_{j'}, 2_j, 2_{j'}\} \text{ for } 1 \le j < j' \le 3,$$
$$\text{and } \{1_1, 1_2, 1_3, 4_1\}, \{2_1, 2_2, 2_3, 4_2\}.$$

(ii) When $n \in [35, 40]$, remove the three blocks $\big\{\{i_1, i_2, i_3\} : i \in [3]\big\}$ from $\mathcal{B}$, and add any $[7, 12]$ blocks from below to obtain the desired set of blocks.

$$\{i_j, i_{j'}, i'_j, i'_{j'}\},\ \text{for } 1 \le i < i' \le 3, 1 \le j < j' \le 3,$$
$$\text{and } \{1_1, 1_2, 1_3, 4_1\}, \{2_1, 2_2, 2_3, 4_2\}, \{3_1, 3_2, 3_3, 4_3\}.$$

We can check that for each $n \in [33, 40]$, the gap between the resulting upper bound and the lower bound $\lceil \frac{25}{6}n - \frac{55}{6} \rceil$ is at most two. In Figure 4, for $n < 80$, we compare the lower bounds of $N(n, 6, 12; 3)$ from Theorems 10, 12, 25. In addition, we also summarize the upper bounds resulting from the constructions presented in this section. □

VI. Conclusion and Open Problems

We studied the lower bounds on the minimum total storage $N(n, k, m; r)$ of MCBCs. To this end, we formulated optimization programs whose optimal solutions yielded lower bounds for $N(n, k, m; r)$. We improved known lower bounds of $N(n, k, m; r)$ and determined the exact values in some cases. To conclude, we discuss some open problems.

(i) Critical to our lower bound derivations are the linear constraints obtained from variations of Hall's conditions. It is conceivable that these techniques are applicable to CBCs where more than one item can be read from each server (see [10]), and also other variants of CBCs like erasure CBCs (see [23]).

(ii) In addition to the constraints imposed by Hall's conditions, we utilized constraints implied from uniform MCBCs to improve the lower bounds (see Sections IV and V). Informed by the choice of certain feasible solutions, we again utilized uniform MCBCs to build optimal MCBCs with varying block sizes. Hence, it appears promising to study uniform CBCs and MCBCs in depth to discover the connections with optimal codes.

(iii) Notice that our constructions of the $(n, N, 6, m; 3)$-MCBC when $m \equiv 0 \pmod 6$ in Construction D and Example 31 rely on the existence of H designs. For other congruence classes of m, or other parameters, it will be interesting to find other similar combinatorial structures to construct the corresponding codes.

Appendix
Proof Sketch of Theorem 4

Proof Sketch [4, Lemma 3.2]: Dividing both sides of (3) by $\binom{m-c}{k-c-1}$ and then subtracting $\sum_{i\in[r,k]} x_i = n$ on both sides respectively, we get

$$\sum_{i\in[r,k-1]} \left(\frac{\binom{m-i}{k-i-1}}{\binom{m-c}{k-c-1}} - 1 \right) x_i - x_k \le U_{m,k,c;r} - n,$$

where $U_{m,k,c;r} = \frac{\lfloor \frac{k-1}{r} \rfloor \binom{m}{c}}{\binom{k-1}{c}}$. Since $\frac{\binom{m-i}{k-i-1}}{\binom{m-c}{k-c-1}} - 1 \ge \frac{(m-k+1)(c-i)}{k-c}$ for any $c \in [k-1]$ (see [4, Lemma 3.1]), we have

$$\sum_{i\in[r,k-1]} (c-i)x_i \le \frac{(k-c)(U_{m,k,c;r} - n + x_k)}{m-k+1},$$

and therefore

$$N(n,k,m;r)=\sum_{i\in[r,k]} ix_i = nc-\sum_{i\in[r,k]}(c-i)x_i$$
$$\geq nc-\frac{(k-c)(U_{m,k,c;r}-n)}{m-k+1}+\frac{(k-c)(m-k)}{m-k+1}x_k.$$

Since $x_k \geq 0$ and $N(n,k,m;r)$ is an integer, we get (1). ■

Acknowledgment

The authors would like to thank the Associate Editor and the two anonymous reviewers for their constructive comments and suggestions which significantly improve the presentation of this work.

References

[1] H. Asi and E. Yaakobi, "Nearly optimal constructions of PIR and batch codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 947–964, Feb. 2019.

[2] N. Balachandran and S. Bhattacharya, "On an extremal hypergraph problem related to combinatorial batch codes," *Discrete Appl. Math.*, vol. 162, pp. 373–380, Jan. 2014.

[3] J. Bao and L. Ji, "The completion determination of optimal $(3,4)$-packings," *Des., Codes Cryptogr.*, vol. 77, no. 1, pp. 217–229, 2015.

[4] S. Bhattacharya, S. Ruj, and B. Roy, "Combinatorial batch codes: A lower bound and optimal constructions," *Adv. Math. Commun.*, vol. 6, no. 2, pp. 165–174, 2012.

[5] E. Boyle, G. Couteau, N. Gilboa, and Y. Ishai, "Compressing vector OLE," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Toronto, ON, Canada, Jan. 2018, pp. 896–912.

[6] A. E. Brouwer, J. B. Shearer, N. J. A. Sloane, and W. D. Smith, "A new table of constant weight codes," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1334–1380, Nov. 1990.

[7] R. A. Brualdi, K. P. Kiernan, S. A. Meyer, and M. W. Schroeder, "Combinatorial batch codes and transversal matroids," *Adv. Math. Commun.*, vol. 4, no. 3, pp. 419–431, 2010.

[8] Z. Tuza and C. Bujtás, "Optimal batch codes: Many items or low retrieval requirement," *Adv. Math. Commun.*, vol. 5, no. 3, pp. 529–541, Aug. 2011.

[9] C. Bujtás and Z. Tuza, "Optimal combinatorial batch codes derived from dual systems," *Miskolc Math. Notes*, vol. 12, no. 1, pp. 11–23, 2011.

[10] C. Bujtás and Z. Tuza, "Relaxations of Hall's condition: Optimal batch codes with multiple queries," *Applicable Anal. Discrete Math.*, vol. 6, no. 1, pp. 72–81, 2012.

[11] C. Bujtás and Z. Tuza, "Turán numbers and batch codes," *Discrete Appl. Math.*, vol. 186, pp. 45–55, May 2015.

[12] S. Buzaglo, Y. Cassuto, P. H. Siegel, and E. Yaakobi, "Consecutive switch codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2485–2498, Apr. 2018.

[13] V. Chvatal, *Linear Programming*. New York, NY, USA: Macmillan, 1983.

[14] F.-W. Fu and S.-T. Xia, "Binary constant-weight codes for error detection," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1294–1299, May 1998.

[15] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Batch codes and their applications," in *Proc. 36th Annu. ACM Symp. Theory Comput. (STOC)*, 2004, pp. 262–271.

[16] D. Jia, S. Zhang, and G. Zhang, "Erasure combinatorial batch codes based on nonadaptive group testing," *Des., Codes Cryptogr.*, vol. 87, no. 7, pp. 1647–1656, Jul. 2019.

[17] S. Johnson, "A new upper bound for error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-8, no. 3, pp. 203–207, Apr. 1962.

[18] J. Jung, C. Mummert, E. Niese, and M. Schroeder, "On erasure combinatorial batch codes," *Adv. Math. Commun.*, vol. 12, no. 1, pp. 49–65, 2018.

[19] Y. Kanizo, O. Rottenstreich, I. Segall, and J. Yallouz, "Designing optimal middlebox recovery schemes with performance guarantees," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 10, pp. 2373–2383, Oct. 2018.

[20] W. H. Mills, "On the existence of H designs," *Congr. Numer.*, vol. 79, pp. 129–141, 1990.

[21] M. B. Paterson, D. R. Stinson, and R. Wei, "Combinatorial batch codes," *Adv. Math. Commun.*, vol. 3, no. 1, pp. 13–27, 2009.

[22] A. S. Rawat, Z. Song, A. G. Dimakis, and A. Gál, "Batch codes through dense graphs without short cycles," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1592–1604, Apr. 2016.

[23] N. Silberstein, "Fractional repetition and erasure batch codes," in *Coding Theory and Applications* (CIM Series in Mathematical Sciences), vol. 3, R. Pinto, P. R. Malonek, and P. Vettori, Eds. Cham, Switzerland: Springer, 2015, pp. 335–343.

[24] N. Silberstein and T. Etzion, "Optimal fractional repetition codes based on graphs and designs," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4164–4180, Aug. 2015.

[25] N. Silberstein and A. Gál, "Optimal combinatorial batch codes based on block designs," *Des., Codes Cryptogr.*, vol. 78, no. 2, pp. 409–424, Feb. 2016.

[26] A. Vardy and E. Yaakobi, "Constructions of batch codes with near-optimal redundancy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 1197–1201.

[27] Z. Wang, H. M. Kiah, Y. Cassuto, and J. Bruck, "Switch codes: Codes for fully parallel reconstruction," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2061–2075, Apr. 2017.

[28] H. Zhang and V. Skachek, "Bounds for batch codes with restricted query size," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 1192–1196.

[29] H. Zhang, E. Yaakobi, and N. Silberstein, "Multiset combinatorial batch codes," *Des., Codes Cryptogr.*, vol. 86, no. 11, pp. 2645–2660, Nov. 2018.

Yeow Meng Chee (Senior Member, IEEE) received the B.Math., M.Math., and Ph.D. degrees in computer science from the University of Waterloo in 1988, 1989, and 1996, respectively.

He held senior positions in public service, including the Head of Security (information infrastructure) and the Assistant Director of internationalization at the National Computer Board, the Deputy Director of Strategic Programs with the Infocomm Development Authority (IDA), and the Program Director of Interactive Digital Media Research and Development with the Media Development Authority. He deployed South East Asia's First Certification Authority Netrust in 1997, and also founded the Singapore Computer Emergency Response Team (SingCERT). He was the Head of the Division of Mathematical Sciences, Nanyang Technological University, from 2008 to 2010, the Chair of the School of Physical and Mathematical Sciences, Nanyang Technological University, from 2011 to 2017, and the Interim Dean of the College of Science, Nanyang Technological University, from 2018 to 2019. He is currently a Professor of industrial systems engineering and management, and an Associate Vice President for innovation and enterprise, with the National University of Singapore (NUS). His research interests include interplay between combinatorics and computer science, especially coding theory, extremal set systems, and their applications.

Dr. Chee is a Fellow and a Council Member of the Institute of Combinatorics and its Applications.

Han Mao Kiah (Member, IEEE) received the Ph.D. degree in mathematics from Nanyang Technological University (NTU), Singapore, in 2014. From 2014 to 2015, he was a Post-Doctoral Research Associate with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 2015 to 2018, he was a Lecturer with the School of Physical and Mathematical Sciences (SPMS), NTU, Singapore. He is currently an Assistant Professor with SPMS, NTU. His research interests include DNA-based data storage, coding theory, enumerative combinatorics, and combinatorial design theory.

Hui Zhang (Member, IEEE) received the Ph.D. degree in applied mathematics from Zhejiang University, Hangzhou, Zhejiang, China, in 2013. From 2012 to 2015 and 2018 to 2019, she used to work as a Project Officer and a Research Fellow at the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. From 2015 to 2017, she was a Post-Doctoral Researcher with the Department of Computer Science, Technion–Israel Institute of Technology. She is currently a Research Fellow with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore. Her research interests include combinatorial theory, coding theory and cryptography, and their intersections.