An Update-based Maximum Column Distance Coding Scheme for Index Coding

Arman Sharififar, Neda Aboutorab, Parastoo Sadeghi

School of Engineering and Information Technology,

University of New South Wales, Australia

Email:{a.sharififar, n.aboutorab, p.sadeghi}@unsw.edu.au

Abstract—In this paper, we propose a new scalar linear coding scheme for the index coding problem called update-based maximum column distance (UMCD) coding scheme. The central idea in each transmission is to code messages such that one of the receivers with the minimum size of side information is instantaneously eliminated from unsatisfied receivers. One main contribution of the paper is to prove that the other satisfied receivers can be identified after each transmission, using a polynomial-time algorithm solving the well-known maximum cardinality matching problem in graph theory. This leads to determining the total number of transmissions without knowing the coding coefficients. Once this number and what messages to transmit in each round are found, we then propose a method to determine all coding coefficients from a sufficiently large finite field. We provide concrete instances where the proposed UMCD coding scheme has a better broadcast performance compared to the most efficient existing coding schemes, including the recursive scheme (Arbabjolfaei and Kim, 2014) and the interlinked-cycle cover (ICC) scheme (Thapa *et al.*, 2017). We prove that the proposed UMCD coding scheme performs at least as well as the MDS coding scheme in terms of broadcast rate. By characterizing two classes of index coding instances, we show that the gap between the broadcast rates of the recursive and ICC schemes and the UMCD scheme grows linearly with the number of messages. Then, we extend the UMCD coding scheme to its vector version by applying it as a basic coding block to solve the subinstances.[1]

Index Terms—Index coding, MDS codes, update-based index coding scheme, broadcast with side information.

I. INTRODUCTION

INDEX coding problem (introduced by Birk and Kol [2]) models an efficient communication system where a single server broadcasts a set of m messages via a noiseless channel to multiple receivers, each demanding a specific message while they may know some other messages a priori as their side information. Exploiting the side information of the receivers, the server can reduce the number of transmissions to satisfy all the receivers by sending coded messages rather than uncoded transmissions. Take a simple instance of index coding problem, depicted in Figure 1 as an example, where the server wishes to satisfy the three receivers. While a trivial solution is to send each message one-by-one (uncoded scheme) in a total of three transmissions, by taking advantage of the receivers' side information and sending two coded messages $x_1 + x_2$ and $x_1 + x_3$, each receiver is able to decode its desired message. For this instance, one transmission is saved thanks to both the receivers' side information and encoding the messages at the
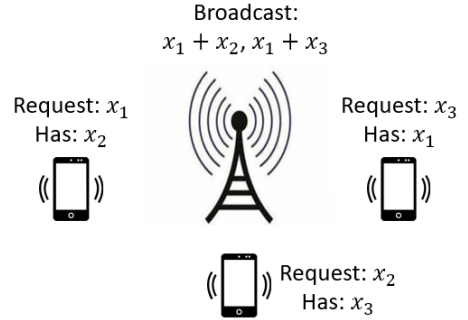

Figure 1: A simple index coding instance which shows that encoding the messages can take advantage of the receivers' side information to save one transmission.

server. The main objective of index coding problem is to design efficient coding schemes for any arbitrary index-coding instance so as to minimize the overall number of transmissions, which is still an open problem.

Index coding problem has so far been extensively studied in the literature and various coding schemes have been proposed. However, to the best of our knowledge, the proposed coding scheme in this paper is the only scheme that considers updating the problem at each step of transmission for the index coding problem. This brings about several advantages. First, it can lead to a lower broadcast rate as will be illustrated through several instances in this paper. Second, this could reduce the average decoding delay due to both the lower broadcast rate and satisfying at least one receiver instantaneously at each step of transmission. Third, in a dynamic system where some receivers may be added to or removed from the system, the update-based scheme can be adapted so that it can deal with the updated system.

The index codes can broadly be categorized into linear and nonlinear codes. Although it has been shown that optimal linear coding can be outperformed by nonlinear codes [3]–[6], owing to its simple and straightforward encoding and decoding processes, it has attracted considerable attention in the literature. While in scalar linear coding each specific message is considered as one variable, so that it is encoded using only one function, in vector linear coding each message can be decomposed into submessages (multi-variables), where each submessage can be encoded by different functions. This can lead to a lower broadcast rate for many index coding instances, but at the cost of increasing the computational complexity.

---

[1]Preliminary results of this paper are presented, in part, in [1].

The existing structured coding schemes can be classified into the following five categories.

- Minrank coding scheme [7]: The minrank scheme is a combinatorial optimization problem, where its solution can give the optimal linear code over a predetermined finite field size for any instance of index coding problem. However, this comes with two main drawbacks. First, this predetermined field size can lead to a linear code where its rate is far from the optimal linear rate, which is achievable over another field size. In fact, it has been shown in [8] that for any field size, there is an explicit way of constructing index coding instances, where the minrank can have a much better performance over another selected field size. Second, the computational complexity of the minrank scheme, especially for the vector linear coding is considerably high and intractable. This is why various other methods have so far been introduced for designing lower-complexity index coding schemes.

- Composite coding scheme [9]–[11]: Inspired by the random coding idea in information theory, in the composite coding scheme, every subset of the message set is randomly mapped to a new composite index set. While the scheme can be used to characterize an achievable rate region, it cannot be used to construct any code due to its random coding nature.

- Interference alignment coding scheme [7]: Inspired by the methods for solving the interference problem in wireless channels, the interference alignment coding scheme was proposed in the context of index coding. This scheme aims to compress the linear space spanned by the interference messages of each receiver as far as all the desired messages are still decodable. This resulted in two main techniques, namely one-to-one alignment and subspace alignment only for index coding instances with certain structures. However, a systematic method is not available for solving any arbitrary instance of the index coding problem.

- Graph-based coding schemes: Since any instance of the index coding problem can be represented as a directed graph, well-known techniques in graph theory have been employed to provide different coding schemes, including the cycle cover [12] and the clique cover [13] schemes. The interlinked-cycle cover (ICC) scheme was proposed in [14], which includes the cycle and clique cover schemes as a special case.

- MDS-based coding schemes: In the partial clique cover (PCC) scheme [2], the index coding instance is first partitioned into subinstances, then each subinstance is solved using the maximum distance separable (MDS) scheme. The vector version of the (PCC) scheme, namely the fractional partial clique cover (FPCC) scheme [15], can be achieved by time-sharing over the solution of each subinstance, which results in a lower broadcast rate for many instances. The recursive scheme [16] is an extension of the FPCC algorithm, in which the local rate of the MDS code is recursively calculated for each subinstance at each stage, which strictly improves upon the FPCC scheme.

In this paper, the index coding problem is approached from a different perspective. In the beginning, the receivers are sorted based on the size of their side information. Then, in each transmission, a linear combination of the messages is designed to instantaneously satisfy one of the receivers with the minimum size of side information. Then, the problem is updated by eliminating all receivers who are able to decode their requested message from the coded messages received so far along with the messages in their side information. This process is repeated until all receivers can successfully decode their requested message.

To design an update-based scheme, the following two questions must be addressed at each step of the transmission. How to design the coding coefficients of the messages? And how to determine whether a receiver can decode its requested message from the information available to it?

Similar to the MDS-based index coding schemes, the proposed UMCD code can be employed as a modular code for solving subinstances. We know that in the encoding matrix of the MDS coding scheme with size $k \times m$ where $k \leq m$, each square submatrix of size $k \times k$ is full-rank. Inspired by this idea, for the former question, the coefficients of the messages in the UMCD coding scheme are designed so that in the encoding matrix, each column is linearly independent from the space spanned by any other columns to the extent possible. For the latter question, this linear independence property is used to prove that the problem of identifying the receivers who are able to decode their requested message at each stage of transmission is equivalent to a well-known problem in graph theory called the maximum cardinality matching (MCM) problem, which can be solved in polynomial-time using the Hopcroft-Karp algorithm [17]. This leads to determining the broadcast rate of the proposed UMCD coding scheme independent of knowing the exact coefficients of the encoding matrix. Once the broadcast rate of the UMCD coding scheme is found by the MCM algorithm in the first phase, in the next phase, the maximum column distance (MCD) algorithm is proposed to design the coefficients of the encoding matrix from a sufficiently large finite field, such that each subset of its columns will be linearly independent as much as possible (the complexity of the MCD algorithm is in general exponential). That is why our proposed coding scheme is called update-based maximum column distance (UMCD) scheme.

The MCM algorithm is significantly more efficient than the MCD algorithm in terms of computational complexity. Thus, separating the UMCD coding scheme into two phases, (i) finding its broadcast rate using the MCM algorithm and (ii) generating its encoding matrix using the MCD algorithm, reduces the complexity of the UMCD coding scheme. This complexity reduction will be more notable for the vector version of the UMCD coding scheme. This is because, first, the optimal solution of the vector version of the UMCD scheme will be achieved only by the broadcast rate (not the encoding matrix) of the UMCD coding scheme for subinstances, and second, the broadcast rate of the UMCD coding scheme is obtained independently of using the MCD algorithm (see also Remark 6).

A. Our Contributions

1) We propose a new linear coding scheme, namely the UMCD coding scheme in Algorithm 1. The UMCD coding scheme consists of two parts:
   - First, its (achievable) broadcast rate is determined using the MCM algorithm in polynomial-time.

- Second, its encoding matrix (code) is constructed using the proposed MCD algorithm, where its complexity is in general exponential.

We provide concrete instances where the proposed UMCD coding scheme outperforms the recursive and ICC coding schemes.

2) We show the satisfied receivers in each transmission can be identified using a polynomial-time algorithm solving maximum cardinality matching (MCM) problem without any knowledge of the coding coefficients. This requires each column of the encoding matrix to be linearly independent of the space spanned by other columns as much as possible.

3) In Algorithm 2, which we call the maximum column distance (MCD) algorithm, we propose a new deterministic method to generate the elements of the encoding matrix so that it meets the aforementioned linear independence requirement.

4) We prove that the broadcast rate of the proposed UMCD coding scheme is no larger than the MDS coding scheme.

5) We characterize a class of index coding instances for which the gap between the broadcast rates of the recursive coding scheme and the proposed UMCD coding scheme grows linearly with the number of messages.

6) We characterize a class of index coding instances for which the gap between the broadcast rates of the ICC coding scheme and the proposed UMCD coding scheme grows linearly with the number of messages.

7) We extend the UMCD coding scheme to its partial and fractional versions by applying time sharing over the subinstances, where each subinstance is solved using the UMCD coding scheme. This brings about the partial UMCD and fractional partial UMCD coding schemes, which strictly improve upon the PCC and FPCC coding schemes, respectively. The fractional partial UMCD coding scheme is optimal for all index coding instances with up to and including five messages.

B. Organization of the Paper

The rest of this paper is organized as follows. Section II provides a brief overview of the system model, relevant background and definitions. In Section III, three index coding instances are provided to describe the motivation of this paper. In Section IV, the UMCD coding scheme is proposed. Section V first establishes the relation between finding the satisfied receivers and the MCM problem, and then the MCD algorithm is proposed. In Section VI, first we prove that the UMCD performs at least as well as the MDS code in terms of the broadcast rate. Then, we provide two classes of index coding instances for which the gap between the broadcast rates of the recursive and ICC coding schemes and the UMCD coding schemes grows linearly with the number of messages. In Section VII, the UMCD coding scheme is extended to its vector version. Finally, Section VIII concludes the paper.

II. SYSTEM MODEL AND BACKGROUND

A. Notation

Small letters such as n denote an integer number where $[n] \triangleq \{1, ..., n\}$ and $[n : m] \triangleq \{n, n+1, \ldots m\}$ for $n < m$. Capital letters such as L denote a set whose cardinality is denoted by $|L|$ and power set is denoted by $\mathcal{P}(A)$. Symbols in bold face such as $\boldsymbol{l}$ and $\boldsymbol{L}$ denote a vector and a matrix, respectively, with $\boldsymbol{L}^T$ denoting the transpose of matrix $\boldsymbol{L}$. A calligraphic symbol such as $\mathcal{L}$ is used to denote a set whose elements are sets. We use $\mathbb{F}_q$ to denote a finite field of size q and write $\mathbb{F}_q^{n \times m}$ to denote the vector space of all $n \times m$ matrices over the field $\mathbb{F}_q$. Given a matrix $\boldsymbol{L} \in \mathbb{F}_q^{n \times m}$ with elements $l_{i,j} \in \mathbb{F}_q$, we use $\boldsymbol{L}_S, S \subseteq [n]$ to represent the $|S| \times m$ submatrix of $\boldsymbol{L}$ comprised of the rows of $\boldsymbol{L}$ indexed by S.

B. System Model

Consider a broadcast communication system in which a server transmits a set of mt messages $X = \{x_i^j, \ i \in [m], \ j \in [t]\}$, $x_i^j \in \mathbb{F}_q$, to a number of receivers $U = \{u_i, \ i \in [m]\}$ through a noiseless broadcast channel. Each receiver u_i wishes to receive a message of length t, $X_i = \{x_i^j, \ j \in [t]\}$ and may have a priori knowledge of a subset of the messages $S_i := \{x_l^j, \ l \in A_i, \ j \in [t]\}$, $A_i \subseteq [m]\backslash\{i\}$, which is referred to as its side information set. The main objective is to minimize the number of coded messages which is required to be broadcast so as to enable each receiver to decode its requested message. An instance of the index coding problem $\mathcal{I}$ is characterized by the side information set of all receivers and can be represented as $\mathcal{I} = \{(i|A_i), i \in [m]\}$.

C. General Index Code

Definition 1 (Index Code). Given an instance of the index coding problem $\mathcal{I} = \{(i|A_i), i \in [m]\}$, a (t, r) index code is defined as $\mathcal{C}_\mathcal{I} = (\phi_\mathcal{I}, \{\psi_\mathcal{I}^i\})$, where

- $\phi_\mathcal{I} : \mathbb{F}_q^{mt} \to \mathbb{F}_q^r$ is the encoding function which maps the mt message symbols $x_i^j \in \mathbb{F}_q$ to the r coded messages as $Y = \{y_1, \ldots, y_r\}$, where $y_k \in \mathbb{F}_q, k \in [r]$.
- $\psi_\mathcal{I}^i$: represents the decoder function, where for each receiver $u_i, i \in [m]$, the decoder $\psi_\mathcal{I}^i : \mathbb{F}_q^r \times \mathbb{F}_q^{|A_i|t} \to \mathbb{F}_q^t$ maps the received r coded messages $y_k \in Y, k \in [r]$ and the $|A_i|t$ messages $x_l^j \in S_i$ in the side information to the t decoded symbols $\psi_\mathcal{I}^i(Y, S_i) = \{\hat{x}_i^j, j \in [t]\}$, where $\hat{x}_i^j$ is an estimate of x_i^j.

Definition 2 ($\beta(\mathcal{C}_\mathcal{I})$: Broadcast Rate of $\mathcal{C}_\mathcal{I}$). Given an instance of the index coding problem $\mathcal{I}$, the broadcast rate of a (t, r) index code $\mathcal{C}_\mathcal{I}$ is defined as $\beta(\mathcal{C}_\mathcal{I}) = \frac{r}{t}$.

Definition 3 ($\beta(\mathcal{I})$: Broadcast Rate of $\mathcal{I}$). Given an instance of the index coding problem $\mathcal{I}$, the broadcast rate $\beta(\mathcal{I})$ is defined as

$$\beta(\mathcal{I}) = \inf_t \inf_{\mathcal{C}_\mathcal{I}} \beta(\mathcal{C}_\mathcal{I}). \tag{1}$$

Thus, the broadcast rate of any index code $\mathcal{C}_\mathcal{I}$ provides an upper bound on the broadcast rate of $\mathcal{I}$, i.e., $\beta(\mathcal{I}) \leq \beta(\mathcal{C}_\mathcal{I})$.

Definition 4 (Index Coding Subinstance). Given an index coding instance $\mathcal{I} = \{(i|A_i), i \in [m]\}$, for any subset $M \subseteq [m]$, we define a subinstance as $\mathcal{I}_M = \{(i|A_i \cap M), i \in M\}$. Since each subinstance $\mathcal{I}_M$ is completely characterized by M, for brevity, we call it subinstance M in the rest of this paper.

Definition 5 (Partitioning an index coding instance into subinstances). We say that the index coding instance $\mathcal{I}$ is

partitioned into subinstances $M_1, \ldots, M_n$ for some $n \leq m$, if the following conditions are met.

$$\begin{cases} M_1 \cup \ldots \cup M_n = [m], \\ M_{j_1} \cap M_{j_2} = \emptyset, \quad \text{if } j_1 \neq j_2 \text{ for } j_1, j_2 \in [n]. \end{cases} \quad (2)$$

Note, unless we explicitly use the word 'partition', the subsets $M_1, \ldots, M_n$ may overlap.

D. Linear Index Code

Let $\boldsymbol{x} = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_m^T]^T \in \mathbb{F}_q^{mt \times 1}$ denote the message vector, where $\boldsymbol{x}_i = [x_i^1, \ldots, x_i^t]^T \in \mathbb{F}_q^{t \times 1}$ is the requested message vector by receiver $u_i, i \in [m]$.

Definition 6 (Linear Index Code). Given an instance of the index coding problem $\mathcal{I} = \{(i|A_i), i \in [m]\}$, a (t, r) linear index code is defined as $\mathcal{C}_{\mathcal{I}} = (\boldsymbol{H}, \{\psi_{\mathcal{I}}^i\})$, where

- $\boldsymbol{H} : \mathbb{F}_q^{mt \times 1} \rightarrow \mathbb{F}_q^{r \times 1}$ is the $r \times mt$ encoding matrix which maps the message vector $\boldsymbol{x} \in \mathbb{F}_q^{mt \times 1}$ to a coded message vector $\boldsymbol{y} = [y_1, \ldots, y_r]^T \in \mathbb{F}_q^{r \times 1}$ as follows

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} = \sum_{i \in [m]} \boldsymbol{H}^{\{i\}} \boldsymbol{x}_i.$$

Here $\boldsymbol{H}^{\{i\}} \in \mathbb{F}_q^{r \times t}$ is the local encoding matrix of the i-th message vector $\boldsymbol{x}_i$ such that $\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}^{\{1\}} & | & \ldots & | & \boldsymbol{H}^{\{m\}} \end{bmatrix} \in \mathbb{F}_q^{r \times mt}$.

- $\psi_{\mathcal{I}}^i$ represents the linear decoder function for receiver $u_i, i \in [m]$, where $\psi_{\mathcal{I}}^i(\boldsymbol{y}, S_i)$ maps the received coded message $\boldsymbol{y}$ and its side information messages S_i to $\hat{\boldsymbol{x}}_i$, which is an estimate of the requested message vector $\boldsymbol{x}_i$.

Proposition 1. *The necessary and sufficient condition for linear decoder $\psi_{\mathcal{I}}^i, \forall i \in [m]$ to correctly decode the requested message vector $\boldsymbol{x}_i$ is*

$$\text{rank } \boldsymbol{H}^{\{i\} \cup B_i} = \text{rank } \boldsymbol{H}^{B_i} + t, \quad (3)$$

where $B_i = [m] \backslash (A_i \cup \{i\})$ represents the interfering message set of receiver $u_i, i \in [m]$, and $\boldsymbol{H}^L$ denotes the matrix $\begin{bmatrix} \boldsymbol{H}^{\{l_1\}} & | & \ldots & | & \boldsymbol{H}^{\{l_{|L|}\}} \end{bmatrix}$ for the given set $L = \{l_1, \ldots, l_{|L|}\}$.

Proof. Refer to Appendix A. $\square$

Definition 7 (Scalar and Vector Linear Index Code [18]). The linear index code $\mathcal{C}_{\mathcal{I}}$ is said to be scalar if $t = 1$. Otherwise, it is called a vector (or fractional) code. For scalar codes, we use $x_i = x_i^1, i \in [m]$, for simplicity.

Since the proposed UMCD coding scheme is a scalar scheme, throughout the paper until Section VII, where we extend the UMCD to its vector version, we assume that $t = 1$. Thus, for any matrix such as $\boldsymbol{H} \in \mathbb{F}_q^{r \times m}, r \leq m$, we use $\boldsymbol{H}^L, L \subseteq [m]$ to denote the $r \times |L|$ submatrix of $\boldsymbol{H}$ comprised of the columns indexed by L.

Note, by setting $t = 1$, we have $\beta(\mathcal{C}_{\mathcal{I}}) = r$. Since sending the messages uncoded with rate $r = m$ is a linear index code for any index coding instance, we always have $r \leq m$.

Definition 8 ($\beta_{\text{MDS}}(\mathcal{I})$ MDS Broadcast Rate for $\mathcal{I}$). Given an index coding instance $\mathcal{I} = \{(i|A_i), i \in [m]\}$, let $|A|_{\min} = \min_{i \in [m]} A_i$ denote the minimum size of side information. Then, the broadcast rate of the MDS coding scheme is $\beta_{\text{MDS}}(\mathcal{I}) = m - |A|_{\min}$.

A brief overview of the PCC, FPCC, recursive and ICC coding schemes is provided in Appendix B. Throughout the paper, $\beta_{\text{PCC}}(\mathcal{I})$, $\beta_{\text{FPCC}}(\mathcal{I})$, $\beta_{\text{R}}(\mathcal{I})$ and $\beta_{\text{ICC}}(\mathcal{I})$, respectively, denote the broadcast rate of the PCC, FPCC, recursive and ICC schemes. The broadcast rate of the proposed UMCD coding scheme in this paper is denoted by $\beta_{\text{UMCD}}(\mathcal{I})$.

E. Graph Definitions

Definition 9 ($\mathcal{G}_{\mathcal{I}}$: Graph Representation of $\mathcal{I}$). The index coding instance $\mathcal{I} = \{(i|A_i), i \in [m]\}$ can be represented as a directed graph $\mathcal{G}_{\mathcal{I}} = (V, E)$, where $V = [m]$ and $E \subseteq [m] \times [m]$, respectively, denote the vertex and edge sets such that $(i, j) \in E$ if and only if (iff) $i \in A_j$, for all $(i, j) \in [m] \times [m]$. Moreover, any subinstance $M \subseteq [m]$ can be represented by the induced subgraph $\mathcal{G}_M = (M, E_M)$, where $E_M \subseteq M \times M$ such that $(i, j) \in E_M$ iff $i \in A_j \cap M$. Since, subgraph G_M is completely characterized by subset M, for brevity, we call it subgraph M in the rest of this paper.

Throughout this paper, if there is a pairwise clique between two vertices i and j (i.e., $i \in A_j$ and $j \in A_i$), then they will be connected by a solid bidirectional arrow. Otherwise, we use a dashed unidirectional arrow.

Definition 10 (Maximum Acyclic Induced Subgraph (MAIS) of $\mathcal{I}$). Given an index coding instance $\mathcal{I}$, let $\mathcal{M}$ be the set of all acyclic vertex-induced subgraphs of $\mathcal{G}_{\mathcal{I}}$. A subgraph $M \in \mathcal{M}$ with maximum size $|M|$ is said to be the MAIS set of $\mathcal{I}$, and $\beta_{\text{MAIS}}(\mathcal{I}) = |M|$ is called the MAIS bound on $\mathcal{I}$.

Proposition 2 (Bar-Yossef *et all.* [19]). *Given the index coding instance $\mathcal{I}$, we have $\beta(\mathcal{I}) \geq \beta_{\text{MAIS}}(\mathcal{I})$. Thus, the MAIS bound imposes a lower bound on the broadcast rate.*

III. MOTIVATING INDEX CODING INSTANCES

In this section, three index coding instances are provided to illustrate the motivation behind the proposed UMCD coding scheme for the index coding problem. Note that since the MCM and the MCD algorithms have not yet been discussed, for these instances, we employ neither the MCM algorithm for finding the satisfied receivers, nor the MCD algorithm for determining the elements of the encoding matrix. However, the elements of the encoding matrix are designed to meet the UMCD coding scheme requirement (where each column is designed to be linearly independent of the space spanned by other columns to the extent possible).

Example 1 (The UMCD versus the MDS Coding Scheme). Consider the index coding instance $\mathcal{I}_1 = \{(1|-), (2|3, 4), (3|2, 4), (4, |2, 3)\}$, shown in Figure 2a. For this instance, the broadcast rate of the MDS code is $\beta_{\text{MDS}}(\mathcal{I}_1) = m - |A|_{\min} = 4$, which cannot perform better than the uncoded transmission. The main demerit of the MDS code is that its broadcast rate is determined only by the minimum size of side information. This means the MDS code cannot benefit from the side information set of other receivers. Now, consider the following transmission technique. First, the server aims to satisfy receiver u_1 which has the minimum size of side information. So, it transmits x_1, and then removes this receiver from the instance. Now, the server observes three receivers u_1, u_2 and u_3, which form a clique with each other. Thus, by

(a) Index coding instance $\mathcal{I}_1$ **(b)** Index coding instance $\mathcal{I}_2$

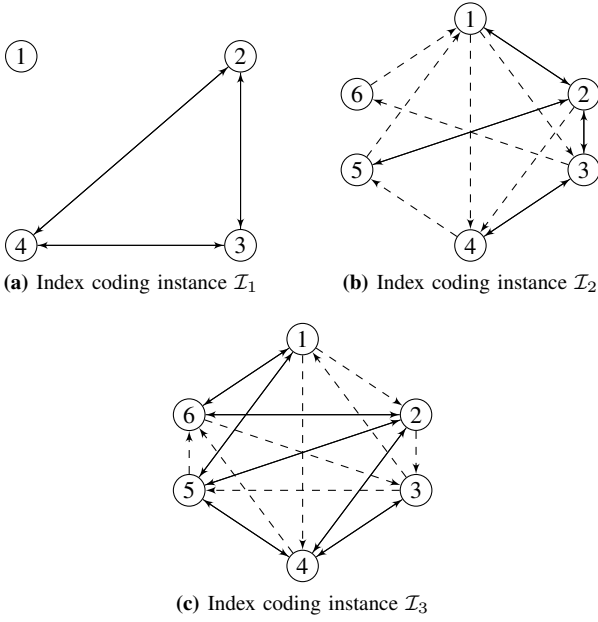**(c)** Index coding instance $\mathcal{I}_3$

Figure 2: Motivating index coding instances.

sending $x_2 + x_3 + x_4$ all the receivers will be satisfied. This shows how updating the problem by removing the satisfied receivers after each transmission can provide a more efficient broadcast rate compared to the MDS code.

Table I: COMPARISON BETWEEN THE BROADCAST RATE OF DIFFERENT CODING SCHEMES FOR $\mathcal{I}_2$

Schemes	Broadcast rate
Maximum distance separable (MDS) [18]	5
Scalar clique cover [13]	4
Scalar cycle cover [12]	4
Partial clique cover (PCC) [2]	4
Scalar Interlinked-cycle cover (ICC) [14]	4
Scalar recursive [16]	4
Fractional clique cover [2]	4
Fractional cycle cover [12]	3.5
Fractional partial clique cover (FPCC) [15]	3.5
Interlinked-cycle cover (ICC) [14]	3.5
Recursive [16]	3.5
MAIS bound	3
Proposed UMCD	3

Example 2 (The UMCD versus the MD-S-based and Graph-based Coding Schemes). Consider the index coding instance $\mathcal{I}_2 = \{(1|2,5,6),(2|1,3,5),(3|1,2,4),(4|1,2,3),(5|2,4), (6|3)\}$, depicted in Figure 2b. For the broadcast rate of the MDS code, we have $\beta_{\text{MDS}}(\mathcal{I}_2) = m - |A|_{\min} = 5$. It can be verified that for the MDS-based and graph-based schemes, partitioning the instance $\mathcal{I}_2$ can improve the broadcast rate by saving two transmissions. The broadcast rate can even be reduced to 3.5 by the vector version of the schemes, which comes at the expense of higher computational complexity. However, none can still achieve the MAIS bound with $\beta_{\text{MAIS}}(\mathcal{I}_2) = 3$. Now, assume that the server first targets receiver u_6, which has the minimum size of side information by transmitting a linear combination of messages x_6 (requested message) and

$x_i, i \in A_6$ as follows

$$y_1 = [0\ 0\ h_{1,3}\ 0\ 0\ h_{1,6}]\ \boldsymbol{x},$$

where $h_{1,i} \in \mathbb{F}_q$ and $\boldsymbol{x} = [x_1\ x_2\ x_3\ x_4\ x_5\ x_6]^T$. This transmission satisfies only receiver u_6. Then, the server targets receiver u_5 (which has the minimum size of side information among the unsatisfied receivers) by sending a linear combination of messages x_5 (requested message) and $x_i, i \in A_5$ as follows

$$y_2 = [0\ h_{2,2}\ 0\ h_{2,4}\ h_{2,5}\ 0]\ \boldsymbol{x}.$$

It can be checked that only receiver u_5 is satisfied by receiving the coded messages y_1 and y_2. Now, the remaining unsatisfied receivers $u_i, i = [4]$ have the same size of side information. Assume receiver u_4 is chosen randomly. To satisfy receiver u_4, a linear combination of messages x_4 (requested message) and $x_i, i \in A_4$ is transmitted as follows

$$y_3 = [h_{3,1}\ h_{3,2}\ h_{3,3}\ h_{3,4}\ 0\ 0]\ \boldsymbol{x}.$$

By setting the field size $q = 3$ and fixing the nonzero coefficients $h_{1,3} = h_{1,6} = h_{2,2} = h_{2,4} = h_{2,5} = h_{3,1} = h_{3,2} = h_{3,3} = 1$ and $h_{3,4} = 2$, it can be seen that each subset of columns are linearly independent as much as possible, at each step of transmission. One can verify that all the remaining receivers u_1, u_2, u_3 and u_4 are able to decode their requested message upon receiving the coded messages y_1, y_2 and y_3. This index code achieves the MAIS bound and, so it is optimal for $\mathcal{I}_2$. This simple instance illustrates how the proposed UMCD coding scheme works, which can outperform the existing linear coding schemes, including the recursive and ICC coding schemes, as presented in Table I.

Example 3 (The UMCD versus the Scalar Binary Minrank Coding Scheme). Please refer to Appendix C to see how the proposed UMCD coding scheme outperforms the scalar binary minrank coding scheme for the index coding instance $\mathcal{I}_3$, depicted in Figure 2c.

IV. MAIN RESULTS

A. A Brief Discussion of the MCM Problem and the MCD Algorithm

Consider a binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}, r \leq m$, which can be characterized by either set $G = \{(k,i) \in [r] \times [m], g_{k,i} = 1\}$ or its support sets as $G_k = \{i \in [m], g_{k,i} = 1\}, k \in [r]$. Now, we say that matrix $\boldsymbol{H} \in \mathbb{F}_q^{r \times m}$ fits the binary matrix $\boldsymbol{G}$ if for any $q \geq 2$, we have $h_{k,i} = 0, \forall (k,i) \in ([r] \times [m]) \backslash G$.

Now, we consider the following optimization problem

$$\max_{\substack{h_{k,i} \in \mathbb{F}_q \\ (k,i) \in G}} \quad \text{rank } \boldsymbol{H},$$

$$\text{subject to} \quad \boldsymbol{H} \text{ fits } \boldsymbol{G}, \tag{4}$$

which gives the maximum rank of $\boldsymbol{H}$ over all possible values from $\mathbb{F}_q$ for its elements $h_{k,i}, (k,i) \in G$ such that $\boldsymbol{H}$ fits $\boldsymbol{G}$. In Subsection V-A, we prove that for any field size $q \geq 2$, the solution of (4) is equal to the MCM value of the bipartite graph associated with the binary matrix $\boldsymbol{G}$, denoted by $\text{mcm}(\boldsymbol{G})$. In other words, subject to the condition that $\boldsymbol{H}$ fits $\boldsymbol{G}$, we have

$$\text{mcm}(\boldsymbol{G}) = \max_{\substack{h_{k,i} \in \mathbb{F}_q \\ (k,i) \in G}} \text{rank } \boldsymbol{H}.$$

The MCM problem can be solved by the polynomial-time Hopcroft-Karp algorithm where its complexity for $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ is $\mathcal{O}(r^2 m)$.[2]

In Subsection V-B, given a binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$, we propose the MCD algorithm to design a matrix $\boldsymbol{H} \in \mathbb{F}_q^{r \times m}$ over a sufficiently large field such that $\boldsymbol{H}$ fits $\boldsymbol{G}$ and $\boldsymbol{H}_{[k]}^L$ will reach its maximum rank in (4) for all $k \in [r], L \subseteq [m]$. The complexity of the MCD algorithm is in general exponential. Assuming that the encoding matrix $\boldsymbol{H}$ is designed by the MCD algorithm, for any $k \in [r]$ and $L \subseteq [m]$, we have

$$\text{rank}(\boldsymbol{H}_{[k]}^L) = \text{mcm}(\boldsymbol{G}_{[k]}^L). \tag{5}$$

Hence, the decoding condition in (3) will be equivalent to

$$\text{mcm}(\boldsymbol{G}_{[k]}^{\{i\} \cup B_i}) = \text{mcm}(\boldsymbol{G}_{[k]}^{B_i}) + 1, \tag{6}$$

which means that, the satisfied receivers in each transmission can be determined just by having access to the binary matrix $\boldsymbol{G}$. This leads to achieving the broadcast rate of the UMCD scheme without knowing the exact value of the elements in $\boldsymbol{H}$. This results in reducing the computational complexity, especially when the UMCD is extended as to its vector version and used as a basic code for solving index coding subinstances.

B. The Proposed UMCD Coding Scheme

In this section, we describe the proposed UMCD coding scheme, which is a scalar linear code and is based on updating the problem instance after each transmission. In this scheme, first, in transmission k, the UMCD coding scheme characterizes the support set G_k such that $g_{k,i} = 1$ if message x_i is included in the linear coded message y_k, and $g_{k,i} = 0$ otherwise. So, for the first k transmissions, we have a binary matrix $\boldsymbol{G}_{[k]}$. Then, using this binary matrix, the other satisfied receivers are identified and removed from the instance. After all the receivers are satisfied at the transmission $k = r = \beta_{\text{UMCD}}(\mathcal{I})$, the encoding matrix $\boldsymbol{H}$ which fits $\boldsymbol{G}_{[r]}$ will be designed. More specifically, the UMCD coding scheme is described as follows.

- In transmission k, the UMCD coding scheme aims at satisfying one of the receivers with the minimum size of side information by sending a linear combination of its desired message and the messages in its side information. This will give the support set G_k which will, in turn, determine the k-th row of the binary matrix $\boldsymbol{G}_{[k]}$.
- Assuming that the encoding matrix $\boldsymbol{H}$ will be designed by the MCD algorithm, the satisfied receivers can be determined by checking the condition in (6) using the polynomial-time Hopcroft-Karp algorithm which solves the MCM problem. This will eventually determine the broadcast rate $\beta_{\text{UMCD}}(\mathcal{I})$ as soon as all receivers are satisfied.
- Finally, given the binary matrix $\boldsymbol{G}_{[r]}$, the encoding matrix $\boldsymbol{H}$ will be determined by the proposed MCD algorithm from a sufficiently large field.

Remark 1. In the UMCD coding scheme, selecting the receiver with the minimum size of side information rather than the receivers with the non-minimum size of side information, in each transmission, can intuitively bring about two main advantages.

[2]It is worth noting that although the minrank problem is NP-hard, the maxrank problem in (4) can be solved in polynomial-time.

First, it can lead to a lower broadcast rate as it will be discussed in Example 6. Second, since the number of ones at each row of the binary matrix $\boldsymbol{G}$ is determined by the side information of such a receiver (with the minimum size of side information), it will lead to a sparser binary matrix $\boldsymbol{G}$. This, in turn, can result in a lower complexity and smaller field size for the MCD algorithm, which will be discussed in Subsection V-B.

C. Description of the UMCD Algorithm

In the UMCD algorithm, which is provided in Algorithm 1, N denotes the set of unsatisfied receivers which is updated after each transmission. In the beginning, all the receivers are considered to be unsatisfied and $N = [m]$. In transmission k, let $W = \text{argmin}_{i \in N} |A_i|$ represent the indices of the unsatisfied receivers with the minimum size of side information. Then, one element, denoted by w is chosen randomly from W. Now, to satisfy receiver u_w, UMCD designs a linear combination of the messages indexed by $G_k = \{w\} \cup A_w$, which characterizes the k-th row of the binary matrix $\boldsymbol{G}$ such that $g_{k,i} = 1$ if $i \in G_k$, and $g_{k,i} = 0$ otherwise. Now, receiver u_w is guaranteed to be removed from N, since G_k only contains what u_w wants to decode and what it knows. The instant decodability for one of the remaining receivers with the minimum size of side information is a key to reducing the overall broadcast rate in an adaptive manner. Then, using the Hopcroft-Karp algorithm, the other receivers satisfying the condition in (6) are identified and removed from N. This process continues until all the receivers are satisfied, which gives the broadcast rate $\beta_{\text{UMCD}}(\mathcal{I}) = r$. Finally, having obtained the binary matrix $\boldsymbol{G}_{[r]}$, the elements of the encoding matrix $\boldsymbol{H}$ are determined using the deterministic MCD algorithm from a field of size $q \geq q_{min}$, where

$$q_{min} = \binom{m}{p}, \ p = \min\{\lfloor \frac{m}{2} \rfloor, r\}. \tag{7}$$

It is worth noting that the broadcast rate $\beta_{\text{UMCD}}(\mathcal{I})$ is obtained independently of knowing the exact value of the elements of the encoding matrix $\boldsymbol{H}$. This helps to reduce the computational complexity of the UMCD scheme by not needing to invoke the MCD algorithm to find satisfied receivers at each transmission.

D. The UMCD Scheme Can Outperform the Recursive and ICC Coding Schemes for Some Index Coding Instances with Five Messages

Example 4. Consider the instance of index coding problem $\mathcal{I}_4 = \{(1|2,5), (2|1,4), (3|2), (4|5), (5|1,3)\}$, depicted in Figure 3a. In the first round $k = 1$, UMCD begins with one of the receivers indexed by $W = \{3, 4\}$ which have the minimum size of side information. Let $w = 3 \in W$ be chosen at random. Then, the first transmission is a linear combination of the messages indexed by $G_1 = \{3\} \cup A_3 = \{2, 3\}$, satisfying receiver u_3. So, $N = [5] \backslash \{3\}$. Using the Hopcroft-Karp algorithm, (6) does not hold for the remaining receivers in N.

Now, for the second round $k = 2$, we have $W = \{4\}$, $w = 4$, $G_2 = \{4\} \cup A_4 = \{4, 5\}$, and $N = [5] \backslash \{3, 4\}$. Using the Hopcroft-Karp algorithm, (6) does not hold for the remaining receivers in N.

Algorithm 1: UMCD Coding Scheme

Input: $\mathcal{I} = \{(i|A_i), i \in [m]\}$
Output: $\beta_{\text{UMCD}}(\mathcal{I})$ and $\boldsymbol{H}$
initialization;
$k = 0$;
$N = [m]$;
set $\boldsymbol{G}$ as a full-zero matrix of size $m \times m$;
while $N \neq \emptyset$ **do**
   $k \leftarrow k + 1$;
   $W = \arg\min_{i \in N} |A_i|$;
   choose an element $w \in W$ at random;
   set $G_k = \{w\} \cup A_w$;
   set $g_{k,i} = 1, \forall i \in G_k$;
   $N \leftarrow N \backslash \{w\}$;
   for $i \in N$ **do**
      if $\text{mcm}\left(\boldsymbol{G}_{[k]}^{\{i\} \cup B_i}\right) = \text{mcm}\left(\boldsymbol{G}_{[k]}^{B_i}\right) + 1$ **then**
         $N \leftarrow N \backslash \{i\}$;
      end
   end
end
$\beta_{\text{UMCD}}(\mathcal{I}) = r = k$: Broadcast rate;
set the field size q such that $q \geq q_{min}$;
$\boldsymbol{H} = \text{MCD}\left(\boldsymbol{G}_{[r]}, q\right)$: Encoding matrix;

(a) Index coding instance $\mathcal{I}_4$

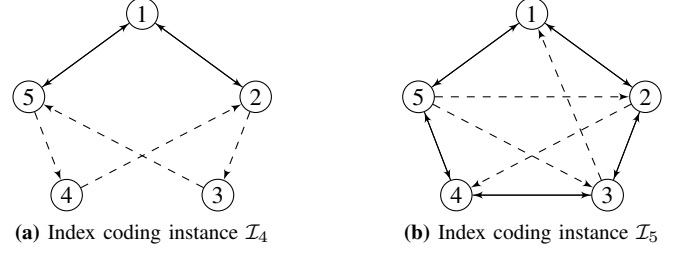**(b)** Index coding instance $\mathcal{I}_5$

Figure 3: (a) The broadcast rate of the UMCD coding scheme is 3, while it is 3.5 for the recursive coding scheme. (b) The broadcast rate of the UMCD coding scheme is 2, while it is 2.5 for the ICC coding scheme. For both instances, the UMCD coding scheme is optimal.

$w = 2$, $G_2 = \{2\} \cup A_2 = \{1, 2, 3, 5\}$, and $N = [5] \backslash \{2, 5\}$. Thus, the binary matrix $\boldsymbol{G}_{[2]}$ will be

$$\boldsymbol{G}_{[2]} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}. \tag{9}$$

Using the Hopcroft-Karp algorithm, (6) holds for all the remaining receivers, which leads to $\beta_{\text{UMCD}}(\mathcal{I}_5) = 2 < \beta_{\text{ICC}}(\mathcal{I}_5) = 2.5$. Finally, the encoding matrix will be determined by the MCD algorithm over a field of size $q \geq q_{min} = \binom{5}{2} = 10$, which will be[4]

$$\boldsymbol{H} = \text{MCD}(\boldsymbol{G}_{[2]}, q) = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 2 \end{bmatrix}.$$

Note that for the other possible random selections of w in the second transmission, it can be checked that the broadcast rate will be the same.

E. Intuition behind the selection of receivers with the minimum size of side information

Here we discuss the intuition for satisfying receivers with the minimum size of side information. Assume that in the UMCD coding scheme, in each transmission, instead of receiver u_{w_1} with the minimum size of side information, receiver u_{w_2} with a larger size of side information is chosen. Now, since $|A_{w_2}| \geq |A_{w_1}|$, then $|G_k| = |\{w_2\} \cup A_{w_2}|$ will be greater than $|\{w_1\} \cup A_{w_1}|$, which places more ones in the k-th row of $\boldsymbol{G}$. This, in turn, is expected to increase the value of $\text{mcm}(\boldsymbol{G}_{[k]}^{B_i})$ for receivers with the smaller size of side information (or larger size of interfering message set). In other words, satisfying a receiver with the larger size of side information can increase the dimension of the space spanned by the messages inside the interference message set B_i for receivers with the smaller size of side information. Hence, it is expected that the decoding condition (6) is not met for receivers with the smaller size of side information. Thus, separate transmissions can still be required to satisfy receivers with the smaller size of side information, which increases the broadcast rate. In the following example, it is shown that selecting receivers with the non-minimum size of side information will lead to suboptimal codes for $\mathcal{I}_4$ and $\mathcal{I}_5$.

Example 6. Assume that in the UMCD coding scheme for $\mathcal{I}_5$, first we satisfy receiver u_1. So, $G_1 = \{1\} \cup A_1$. Then, using the Hopcroft-Karp algorithm, (6) holds for receivers u_1 and u_2.

For the third round $k = 3$, we have $W = \{1, 2, 5\}$. Let $w = 1 \in W$, $G_3 = \{1\} \cup A_1 = \{1, 2, 5\}$, and $N = [5] \backslash \{1, 3, 4\}$. Thus, the binary matrix $\boldsymbol{G}_{[3]}$ will be

$$\boldsymbol{G}_{[3]} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}. \tag{8}$$

Using the Hopcroft-Karp algorithm, it can be verified that (6) holds for all the remaining receivers, leading to $\beta_{\text{UMCD}}(\mathcal{I}_4) = 3 < \beta_{\text{R}}(\mathcal{I}_4) = 3.5$. Finally, the encoding matrix will be determined by the MCD algorithm over a field of size $q \geq q_{min} = \binom{5}{2} = 10$, which will be[3]

$$\boldsymbol{H} = \text{MCD}(\boldsymbol{G}_{[3]}, q) = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Note that for other possible random selections of w at each transmission, it can be checked that the broadcast rate will be the same.

Example 5. Consider the index coding instance $\mathcal{I}_5 = \{(1|2, 3, 5), (2|1, 3, 5), (3|2, 4, 5), (4|2, 3, 5), (5|1, 4)\}$, depicted in Figure 3b. In the first round, UMCD begins with the receiver indexed by $W = \{5\}$ which has the minimum size of side information. So, the UMCD coding scheme sets $w = 5$, $G_1 = \{5\} \cup A_5 = \{1, 4, 5\}$, satisfying receiver u_5. So, $N = [5] \backslash \{5\}$. Using the Hopcroft-Karp algorithm, (6) does not hold for the remaining receivers in N.
Now, for the second round $k = 2$, we have $W = \{1, 2, 3, 4\}$,

---

[3]For this case, however, we show in Example 14 that the MCD algorithm over any field of size $q \geq 2$ will satisfy (5).

[4]For this case, however, we show in Example 14 that the MCD algorithm over any field of size $q \geq 3$ will satisfy (5).

Then, assume we target receiver u_3 by setting $G_2 = \{3\} \cup A_3$. Then, using the Hopcroft-Karp algorithm, (6) holds for receivers u_3 and u_4. Thus, we need one more transmission for satisfying receiver u_5, which will result in a suboptimal code. Similarly, for the index coding instance $\mathcal{I}_4$, if one of the receivers u_3 and u_4 is not selected for the first three transmissions, then the broadcast rate will be 4, leading to a suboptimal code.

V. THE MAXIMUM CARDINALITY MATCHING (MCM) PROBLEM AND THE MAXIMUM COLUMN DISTANCE (MCD) ALGORITHM

In this section, first we prove that the solution of the optimization problem in (4) will be equal to the MCM value of the bipartite graph associated with binary matrix $\boldsymbol{G}$. This shows that the satisfied receivers in the UMCD coding scheme can be identified by finding $\mathrm{mcm}(\boldsymbol{G})$. Second, we propose the MCD algorithm to design the encoding matrix $\boldsymbol{H}$ such that it fits $\boldsymbol{G}$ and its submatrices will reach their maximum possible rank as desired in (5).

A. The Equivalence of Identifying Satisfied Receivers and Maximum Cardinality Matching

In this subsection, we formalize the relation between finding the satisfied receivers and the MCM problem which will be based on Theorem 1. To prove Theorem 1, first, we provide Lemma 1 below.

Lemma 1. *For a given matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times r}$, if $g_{k_1,r} = 1$ for some $k_1 \in [r]$, and there exists a permutation of columns in $\boldsymbol{F} = \boldsymbol{G}_{[r] \setminus \{k_1\}}^{[r-1]}$ which results in $f_{k,k} = 1, \forall k \in [r-1]$, then there will also exist a permutation of columns in $\boldsymbol{G}$ which will lead to $g_{k,k} = 1, \forall k \in [r]$.*

Proof. It can be seen the relation between the elements of $\boldsymbol{F}$ and $\boldsymbol{G}$ is as follows

$$\begin{cases} f_{k,j} = g_{k,j}, & \forall \ k < k_1, \ \forall j \in [r-1], \\ f_{k,j} = g_{k,j+1}, & \forall \ k \geq k_1, \ \forall j \in [r-1], \end{cases}$$

So, if there exists a permutation of columns in $\boldsymbol{F}$ to make $f_{k,k} = 1, \forall k \in [r-1]$, the same permutation of columns in $\boldsymbol{G}$ will result in $g_{k,k} = 1, \forall k < k_1$ and $g_{k+1,k} = 1, \forall k > k_1$. Since $g_{k_1,r} = 1$, then we make a left circular shift of the columns $\boldsymbol{G}^{\{k_1\}}, \ldots, \boldsymbol{G}^{\{r-1\}}, \boldsymbol{G}^{\{r\}}$ by one position so that they will be placed in a new position as columns $\boldsymbol{G}^{\{k_1+1\}}, \ldots, \boldsymbol{G}^{\{r\}}, \boldsymbol{G}^{\{k_1\}}$, respectively. Then, it can be observed that we will have $g_{k,k} = 1, \forall k \in [r]$. $\square$

Theorem 1. *Assume matrix $\boldsymbol{H}$ fits the binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times r}$. Now, $\boldsymbol{H}$ can be designed to be full-rank over any field size $q \geq 2$, iff there exists a permutation of its columns which results in $g_{k,k} = 1, \forall k \in [r]$.*

Proof. For the *if* condition, assume that there is a permutation of columns in $\boldsymbol{G}$ which results in $g_{k,k} = 1$ for all $k \in [r]$, then we do such a permutation for $\boldsymbol{H}$, and set $h_{k,k} = 1$ for all $k \in [r]$, and its other elements to zero. Since rearranging the columns does not alter the matrix rank, we have $\mathrm{rank}(\boldsymbol{H}) = r$ over any field of size $q \geq 2$.

Conversely, having Lemma 1, we use induction for proving the converse. For $r = 1$, it is obvious that $\mathrm{rank}(\boldsymbol{H} = [h_{1,1}]) = 1$, only if $h_{1,1} \neq 0$, which gives $g_{1,1} = 1$. For the induction

hypothesis, we assume that the necessary condition holds for $\boldsymbol{H}$ of size $(r-1) \times (r-1)$. Now, we need to prove that the necessary condition must also hold for $\boldsymbol{H}$ of size $r \times r$. Let $\mathrm{rank}(\boldsymbol{H}) = r$. The determinant of $\boldsymbol{H}$ can be obtained using the Laplace expansion along the last column r, as follows

$$|\boldsymbol{H}| = \sum_{k \in [r]} (-1)^{k+r} \ h_{k,r} \ \left| \boldsymbol{H}_{[r] \setminus \{k\}}^{[r-1]} \right|. \tag{10}$$

Thus, for having $|\boldsymbol{H}| \neq 0$, there must exist at least one $k_1 \in [r]$ such that $h_{k_1,r} \ |\boldsymbol{H}_{[r] \setminus \{k_1\}}^{[r-1]}| \neq 0$. So, we must have $h_{k_1,r} \neq 0$, which requires $g_{k_1,r} = 1$, and $|\boldsymbol{H}_{[r] \setminus \{k_1\}}^{[r-1]}| \neq 0$. This is guaranteed by the induction hypothesis that for $\boldsymbol{F} = \boldsymbol{G}_{[r] \setminus \{k_1\}}^{[r-1]}$ there exists a permutation of its columns which leads to $f_{k,k} = 1, \forall k \in [r]$, which completes the proof according to Lemma 1. $\square$

Remark 2. It can be easily concluded from Theorem 1 that the maximum rank of $\boldsymbol{H} \in \mathbb{F}_q^{r \times m}$ which fits $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ in (4) will be equal to the maximum number of ones that appear on the main diagonal of $\boldsymbol{G}$ after performing an appropriate permutation of its columns. Thus, this will be equal to the maximum number of elements with value of 1, which are positioned in distinct rows and columns of $\boldsymbol{G}$.

1) The MCM Problem: A binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ can be represented as a bipartite graph $\mathcal{G}(P, C, E)$, where $P = \{p_k, k \in [r]\}$ and $C = \{c_i, i \in [m]\}$ are two disjoint sets of vertices, called, the row and column vertex set, respectively. $E \subseteq P \times C$ denotes the set of edges such that $(p_k, c_i) \in E, (k, i) \in [r] \times [m]$, connects p_k to c_i only if $g_{k,i} = 1$.

Definition 11 (Maximum Cardinality Matching of $\mathcal{G}$ [17]). For the bipartite graph $\mathcal{G}(P, C, E)$, let $E' \subseteq E$ denote the set of edges such that no two edges share a common vertex (i.e., if $(p_{k_1}, c_{i_1}) \in E'$ and $(p_{k_2}, c_{i_2}) \in E'$, then $k_1 \neq k_2$ and $i_1 \neq i_2$). Such a set E' with maximum size is called the maximum cardinality matching (MCM) of $\mathcal{G}$.

Definition 12 ($\mathrm{mcm}(\boldsymbol{G})$). Consider a binary matrix $\boldsymbol{G}$ represented as a graph $\mathcal{G}$. Let E' be an MCM of $\mathcal{G}$. Then, the maximum cardinality matching of $\boldsymbol{G}$ is defined as $\mathrm{mcm}(\boldsymbol{G}) \triangleq |E'|$.

Remark 3. As mentioned above, for each two edges (p_{k_1}, c_{i_1}) and (p_{k_2}, c_{i_2}) in E', we must have $k_1 \neq k_2$ and $i_1 \neq i_2$. Now, consider the binary matrix $\boldsymbol{G}'$ associated with subgraph $\mathcal{G}(P, C, E')$. It can be easily observed that all the elements with value of one are positioned in distinct rows and columns. This implies that $\mathrm{mcm}(\boldsymbol{G})$ gives the maximum number of elements with value of one, which can be placed on the main diagonal after permuting the columns in a specific way. Thus, if $\boldsymbol{H}$ fits $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$, then for any field size $q \geq 2$, we have

$$\mathrm{mcm}(\boldsymbol{G}) = \max_{\substack{h_{k,i} \in \mathbb{F}_q \\ (k,i) \in G}} \mathrm{rank} \ \boldsymbol{H}.$$

Example 7. An example of the MCM problem is illustrated in Figure 4.

B. The Maximum Column Distance (MCD) Algorithm

In this subsection, given a binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ with set $G = \{(k, i) \in [r] \times [m], g_{k,i} = 1\}$, the MCD algorithm is proposed to design a matrix $\boldsymbol{H} \in \mathbb{F}_q^{r \times m}$ such that $\boldsymbol{H}$ fits

(a) Binary matrix $\boldsymbol{G}_{6\times 6}$

(b) Bipartite graph $\mathcal{G}(P,C,E)$

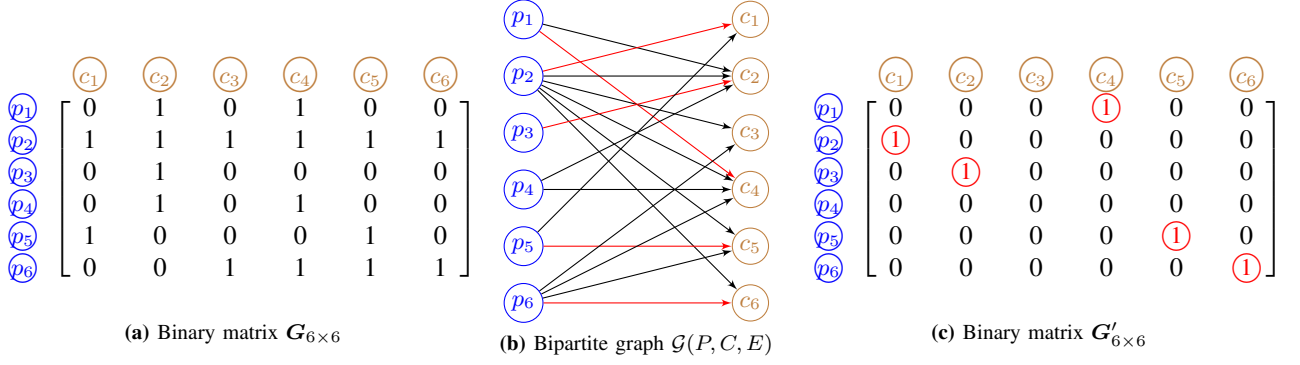**(c)** Binary matrix $\boldsymbol{G}'_{6\times 6}$

Figure 4: Consider the binary matrix $\boldsymbol{G}_{6\times 6}$ in (a) which is represented as a bipartite graph $\mathcal{G}(P,C,E)$ in (b), where $P = \{p_k, k \in [6]\}$, $C = \{c_i, i \in [6]\}$, and $E = \{(k,i) \in [6] \times [6] : g_{k,i} = 1\}$. It can be verified that $E' = \{(p_1,c_4),(p_2,c_1),(p_3,c_2),(p_5,c_5),(p_6,c_6)\}$ (in red) is an MCM of graph $\mathcal{G}$. So, we have $\text{mcm}(\boldsymbol{G}) = |E'| = 5$. The binary matrix $\boldsymbol{G}'$ in (c), associated with the subgraph $\mathcal{G}(P,C,E')$, illustrates how the elements with value of one are placed in distinct rows and columns, so by a proper permutation of its columns, all the ones can be positioned on the main diagonal.

$\boldsymbol{G}$, and each submatrix $\boldsymbol{H}_{[k]}^L, k \in [r], L \subseteq [m]$ will achieve its maximum possible rank.

First, it is trivial that if we set $\boldsymbol{H}_{[1]} = \boldsymbol{G}_{[1]}$, then each $\boldsymbol{H}_{[1]}^L, L \subseteq [m]$ will reach its maximum possible rank. Thus, the MCD algorithm fixes the first row as $\boldsymbol{H}_{[1]} = \boldsymbol{G}_{[1]}$, and in the following, we discuss designing the other rows of $\boldsymbol{H}_{[k]}, k \in [r]\backslash\{1\}$. First, we begin with the concepts of basis and circuit set in the matroid theory [20]. Then, for each element $h_{k,i}, (k,i) \in G$, we define a veto set to characterize the values which must be vetoed by the MCD algorithm so that each submatrix $\boldsymbol{H}_{[k]}^L$ will achieve its maximum possible rank. This will be followed by a description of the MCD algorithm which is presented as Algorithm 2.

1) Prerequisite Material for the MCD Algorithm: In this part, it is assumed that $k \in [r]\backslash\{1\}$, $i \in L \subseteq [m]$, and all elements of $\boldsymbol{H}_{[k]}^L$ are prefixed except element $h_{k,i}$. We also assume that $\text{rank}(\boldsymbol{H}_{[k]}^\emptyset) = \text{rank}([\ ]) = 0$.

Goal: It is aimed to assign a proper value from $\mathbb{F}_q$ to element $h_{k,i}$ such that column $\boldsymbol{H}_{[k]}^{\{i\}}$ will become as much linearly independent of the space spanned by any other columns in $\boldsymbol{H}_{[k]}^{L\backslash\{i\}}$ as possible.

Definition 13 (Basis and Circuit Set of H [20]). We say that set L is an independent set of $\boldsymbol{H}$ if $\text{rank}\,\boldsymbol{H}^L = |L|$. Otherwise, L is considered to be a dependent set. A maximal independent set L is referred to as a basis set. A minimal dependent set L is referred to as a circuit set. Let sets $\mathcal{B}_k$ and $\mathcal{C}_k$, respectively, denote the set of all basis and circuit sets of $\boldsymbol{H}_{[k]}$. It can be shown that

$$\text{rank}(\boldsymbol{H}_{[k]}) = \text{rank}(\boldsymbol{H}_{[k]}^L) = |L|, \qquad \forall L \in \mathcal{B}_k,$$
$$\text{rank}(\boldsymbol{H}_{[k]}^L) = \text{rank}(\boldsymbol{H}_{[k]}^{L\backslash\{i\}}) = |L| - 1, \quad \forall i \in L, \ \forall L \in \mathcal{C}_k.$$
$$(11)$$

Proposition 3. *Let $L \in \mathcal{C}_{k-1}$. Then, there exists only one value for $h_{k,i}$, denoted by $h_{k,i}(L)$, such that if $h_{k,i} \in \mathbb{F}_q\backslash\{h_{k,i}(L)\}$, then L will become an independent set of $\boldsymbol{H}_{[k]}$.*

Proof. Refer to Appendix D-A. $\qquad\square$

Definition 14 (Veto Value). Let $L \in \mathcal{C}_{k-1}$. Since the unique value $h_{k,i}(L)$ in Proposition 3 is undesirable, it is referred to as the veto value of set L for element $h_{k,i}$.

Remark 4. Let $L = \{i\}$. Then, $\{i\} \in \mathcal{C}_{k-1}$ means that column $\boldsymbol{H}_{[k-1]}^{\{i\}}$ is a full-zero column. Thus, by setting $h_{k,i} \in \mathbb{F}_q\backslash\{0\}$, set $\{i\}$ will be an independent set of $\boldsymbol{H}_{[k]}$. This means that if $L = \{i\}$, then $h_{k,i}(\{i\}) = 0$.

Definition 15 (Veto Set). Let $\mathcal{C}_{k-1,i}^L$ denote the set of all $L' \subseteq L\backslash\{i\}$ such that $L' \cup \{i\} \in \mathcal{C}_{k-1}$. We refer to the set $Z_{k,i}^L = \{h_{k,i}(L' \cup \{i\}) \in \mathbb{F}_q : L' \in \mathcal{C}_{k-1,i}^L\}$ as the veto set of L for $h_{k,i}$.

Example 8. Consider the following matrix $\boldsymbol{H} \in \mathbb{F}_5^{3\times 6}$, where $\mathbb{F}_5 = GF(5)$,

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 0 & h_{3,6} \end{bmatrix}. \qquad (12)$$

We want to find the veto set of $L = [6]$ for $h_{3,6}$. We set $k = 3$, $i = 6$. Now, it can be seen that $\mathcal{C}_{2,6}^{[6]} = \{L'_j \subseteq L\backslash\{i\}, j \in [4]\}$, where $L'_1 = \{1,4\}$, $L'_2 = \{2\}$, $L'_3 = \{3,4\}$, and $L'_4 = \{5\}$, which means that each $L'_j \cup \{6\}, j \in [4]$ is a circuit set of $\boldsymbol{H}_{[2]}$. It can also be checked that the corresponding veto value of each $L'_j \cup \{6\}$ for $h_{3,6}$ is as follows: $h_{3,6}(L'_1 \cup \{6\}) = 2$, $h_{3,6}(L'_2 \cup \{6\}) = h_{3,6}(L'_3 \cup \{6\}) = 1$, and $h_{3,6}(L'_4 \cup \{6\}) = 0$. Thus, $Z_{3,6}^{[6]} = \{0,1,2\}$.

Proposition 4. *Assume that $L\backslash\{i\} \in \mathcal{B}_{k-1}$ and L is a dependent set of $\boldsymbol{H}_{[k-1]}$. Now, if $h_{k,i} \in \mathbb{F}_q\backslash Z_{k,i}^L$, then $L \in \mathcal{B}_k$ (we also have $|Z_{k,i}^L| = 1$).*

Proof. Refer to Appendix D-B. $\qquad\square$

Theorem 2. *If $h_{k,i} \in \mathbb{F}_q\backslash Z_{k,i}^L$, then column $\boldsymbol{H}_{[k]}^{\{i\}}$ will be linearly independent of the space spanned by any other columns in $\boldsymbol{H}_{[k]}^{L\backslash\{i\}}$ to the extent possible.*

Proof. We assume that the column $\boldsymbol{H}_{[k-1]}^{\{i\}}$ is linearly dependent on columns in $\boldsymbol{H}_{[k-1]}^{L\backslash\{i\}}$, since otherwise the statement always holds and $|Z_{k,i}^L| = 0$.

Now, let $\mathcal{B}_{k-1,i}^L$ denote the set of all $L' \subseteq L\backslash\{i\}$ such that $L' \in \mathcal{B}_{k-1}$. Since column $\boldsymbol{H}_{[k-1]}^{\{i\}}$ is linearly dependent on columns in $\boldsymbol{H}_{[k-1]}^{L\backslash\{i\}}$, for each $L' \in \mathcal{B}_{k-1,i}^L$, set $L' \cup \{i\}$ will be a dependent set of $\boldsymbol{H}_{[k-1]}$. Now, according to Proposition 4, if

$h_{k,i} \in \mathbb{F}_q \backslash Z_{k,i}^{L' \cup \{i\}}$ (where $|Z_{k,i}^{L' \cup \{i\}}| = 1$) for each $L' \in \mathcal{B}_{k-1,i}^L$, then we will have $L' \cup \{i\} \in \mathcal{B}_k$.

Thus, by setting $h_{k,i} \in \mathbb{F}_q \backslash Z_{k,i}^L$, where

$$Z_{k,i}^L = \cup_{L' \in \mathcal{B}_{k-1,i}^L} Z_{k,i}^{L' \cup \{i\}}, \tag{13}$$

column $\boldsymbol{H}_{[k]}^{\{i\}}$ will be linearly independent of the space spanned by any other columns in $\boldsymbol{H}_{[k]}^{L \cup \{i\}}$ to the extent possible. $\qquad \square$

2) Description of the MCD Algorithm: Assume a binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ with support sets $G_k = \{i \in [m], g_{k,i} = 1\}, k \in [r]$ and a field of size $q \geq q_{min}$ are given, where $q_{min} = \binom{m}{p}$, $p = \min\{\lfloor \frac{m}{2} \rfloor, r\}$. The MCD algorithm, presented as Algorithm 2, in transmission k, assigns a proper value to the elements $h_{k,i}, i \in G_k$ such that each submatrix $\boldsymbol{H}_{[k]}^L, k \in [r], L \subseteq [m]$ will reach its maximum possible rank. First, to guarantee that $\boldsymbol{H}$ will always fit $\boldsymbol{G}$, the MCD algorithm sets $h_{k,i} = 0, \forall(k,i) \in ([r] \times [m]) \backslash G$ where $G = \{(k,i) \in [r] \times [m], g_{k,i} = 1\}$. Then, for the first row $k = 1$, we fix $h_{1,i} = 1, \forall i \in G_1$, leading to $\boldsymbol{H}_{[1]} = \boldsymbol{G}_{[1]}$.

In the k-th row, the MCD algorithm determines a value for the element $h_{k,i}$ where i is chosen randomly from the support set G_k. Then, element i is removed from set G_k as $h_{k,i}$ is going to be fixed in this round. Now, all elements of $\boldsymbol{H}_{[k]}^L$ where $L = [m] \backslash G_k$ are fixed except $h_{k,i}$, which is aimed to be designed so that column $\boldsymbol{H}_{[k]}^{\{i\}}$ will be as much linearly independent of any other columns in $\boldsymbol{H}_{[k]}^{L \backslash \{i\}}$ as possible. Thus, the MCD algorithm computes the veto set $Z_{k,i}^L$, and assigns a randomly chosen value from $\mathbb{F}_q \backslash Z_{k,i}^L$ to $h_{k,i}$. We repeat this process for the remaining elements in G_k (this process is repeated for all rows of $\boldsymbol{G}$) until all the elements $h_{k,i}, (k,i) \in G$ are assigned a value outside of their veto set.

Algorithm 2: MCD Algorithm

Input: $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ and $q \geq q_{min}$
Output: $\boldsymbol{H} = \mathrm{MCD}(\boldsymbol{G}, q)$
initialization;
set $h_{k,i} = 0, \forall(k,i) \in ([r] \times [m]) \backslash G$;
$h_{1,i} = 1, \forall i \in G_1$;
$k = 2$;
while $k \leq r$ **do**
    while $G_k \neq \emptyset$ **do**
        choose an element $i \in G_k$ at random;
        $G_k = G_k \backslash \{i\}$;
        $L = [m] \backslash G_k$;
        find the veto set $Z_{k,i}^L$;
        assign a value to $h_{k,i}$ from $\mathbb{F}_q \backslash Z_{k,i}^L$ at random;
    end
    $k \leftarrow k + 1$
end

Proposition 5. *Let $\boldsymbol{H} = \mathrm{MCD}(\boldsymbol{G}, q)$ where $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ and $q \geq q_{min}$. Then, each submatrix $\boldsymbol{H}_{[k]}^L, k \in [r], L \subseteq [m]$ will reach its maximum possible rank.*

Proof. We note that the statement holds for the first row as $\boldsymbol{H}_{[1]} = \boldsymbol{G}_{[1]}$. Now, we prove the statement for all $k \in [r] \backslash \{1\}, L \subseteq [m]$ by induction. First, for $k = [2] \backslash \{1\}$ and

$L = \{i\} \subseteq [m]$, the statement is correct according to Remark 4. Let $k_1 \in [r], i_1 \in L_1 \subseteq [m]$. Now, we assume that except element h_{k_1,i_1} all the elements of $\boldsymbol{H}_{[k_1]}^{L_1}$ are already fixed such that each submatrix $\boldsymbol{H}_{[k]}^L, k \in [k_1] \backslash \{1\}, L \subseteq L_1$ except $\boldsymbol{H}_{[k_1]}^{L_1}$ reaches its maximum possible rank. Now, based on Theorem 2, if $h_{k_1,i_1} \in \mathbb{F}_q \backslash Z_{k_1,i_1}^{L_1}$, then the column $\boldsymbol{H}_{[k_1]}^{\{i_1\}}$ will be as much linearly independent of the space spanned by any other columns in $\boldsymbol{H}_{[k_1]}^{L_1 \backslash \{i_1\}}$ as possible. Thus, now all the submatrices $\boldsymbol{H}_{[k]}^L, k \in [k_1] \backslash \{1\}, L \subseteq L_1$ including $\boldsymbol{H}_{[k_1]}^{L_1}$ reach their maximum possible rank, which completes the proof. $\qquad \square$

3) On the Required Field Size and Complexity of the MCD Algorithm:

Remark 5. Based on Definition 15, we have $|Z_{k,i}^L| \leq |\mathcal{C}_{k-1,i}^L|$, where the equality holds if all the veto values inside $Z_{k,i}^L$ are distinct. Now, if column $\boldsymbol{H}_{[k-1]}^{\{i\}}$ is already linearly independent of columns in $\boldsymbol{H}_{[k-1]}^{L \backslash \{i\}}$, then $|\mathcal{C}_{k-1,i}^L| = 0, i \in L$. Otherwise, based on (13), we have $|\mathcal{C}_{k-1,i}^L| = |\mathcal{B}_{k-1,i}^L|$. For this case, for each $k \in [r] \backslash \{1\}, i \in L \subseteq [m]$, we have

$$|Z_{k,i}^L| \leq |\mathcal{B}_{k-1,i}^L| \leq \max_{l \in \min\{k-1, |L|-1\}} \binom{|L|-1}{l}. \tag{14}$$

Proposition 6. *Given the binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$, there always exists a value outside of the veto set for all the elements $h_{k,i}, i \in G_k, k \in [r]$ if the field size is chosen as $q \geq q_{min} = \binom{m}{p}$, where $p = \min\{\lfloor \frac{m}{2} \rfloor, r\}$.*

Proof. Based on (14), for the maximum size of $Z_{k,i}^L, k \in [r] \backslash \{1\}, i \in L \subseteq [m]$, we have

$$\max_{\substack{k \in [r] \backslash \{1\} \\ i \in L \subseteq [m]}} |Z_{k,i}^L| \leq \max_{\substack{k \in [r] \backslash \{1\} \\ L \subseteq [m], L \neq \emptyset}} \max_{l \in \min\{k-1, |L|-1\}} \binom{|L|-1}{l}$$
$$< \binom{m}{p} = q_{min}$$

Thus, if $q \geq q_{min}$, we have $|\mathbb{F}_q \backslash Z_{k,i}^L| > 0$ for all $k \in [r] \backslash \{1\}, i \in L \subseteq [m]$. $\qquad \square$

Proposition 7. *Given $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$, the complexity of the MCD algorithm in the worst case scales as $\mathcal{O}(2^m)$.*

Proof. First note that for each circuit set we need to run the rref (reduced row echelon form) to achieve the corresponding veto value for each nonzero element $h_{k,i}, (k,i) \in G$. Now, the worst case scenario for the complexity occurs when all the elements in $\boldsymbol{G}$ are nonzero. In this case, each submatrix $\boldsymbol{H}_{[k]}^L$ will become full-rank designed by the MCD algorithm. Thus, if $k > |L|$, then $|\mathcal{C}_{k-1,i}^L| = 0$ for each $i \in L$. If $k \leq |L|$, we have $|\mathcal{C}_{k-1,i}^L| = |\mathcal{B}_{k-1,i}^L| = \binom{|L|-1}{k-1}$. Thus, the maximum number of times the rref needs to run will be equal to

$$\sum_{\substack{k \in [r] \backslash \{1\} \\ i \in L \subseteq [m]}} |\mathcal{C}_{k-1,i}^L| = \sum_{\substack{k \in [n] \backslash \{1\} \\ n \in [m]}} \binom{n-1}{k-1} = \sum_{n \in [m]} (2^{n-1} - 1)$$
$$= 2^m - (m+1), \tag{15}$$

where $n = |L| \in [m]$. Thus, the complexity scales as $\mathcal{O}(2^m)$. $\qquad \square$

Appendix E provides more discussion on the required field size and complexity of the MCD algorithm for the index coding problem.

Example 9. Consider the following binary matrix

$$\boldsymbol{G} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

Now, given a finite field of size $q \geq q_{min} = 10$, we run the MCD algorithm to generate the encoding matrix $\boldsymbol{H}$. Let $q = 11$ and $\mathbb{F}_{11} = GF(11)$.

First, the MCD algorithm sets $h_{k,i} = 0$ whenever $g_{k,i} = 0, (k,i) \in [3] \times [5]$, and fixes $h_{1,i} = 1, \forall i \in G_1 = \{2\}$. Now, we move to the second row $k = 2$, where $G_2 = \{2,3,4,5\}$. Let $i = 4 \in G_2$ be chosen randomly. Then, $G_2 \leftarrow G_2 \backslash \{4\} = \{2,3,5\}$, $L = ([5] \backslash G_2) = \{1,4\}$ and $Z_{2,4}^L = \{0\}$. Let $h_{2,4} = 1 \in \mathbb{F}_{11} \backslash \{0\}$. Now, let $i = 5 \in G_2$ be chosen randomly, then $G_2 \leftarrow G_2 \backslash \{5\} = \{2,3\}$, $L = \{1,4,5\}$, and $Z_{2,5}^L = \{0\}$. Let $h_{2,5} = 2 \in \mathbb{F}_{11} \backslash \{0\}$. Now, let $i = 2 \in G_2$. Then, $G_2 \leftarrow G_2 \backslash \{2\} = \{3\}$, $L = \{1,2,4,5\}$, and $Z_{2,2}^L = \{0\}$. Let $h_{2,2} = 1 \in \mathbb{F}_{11} \backslash \{0\}$. Now, we have $i = 3 \in G_2$, $G_2 \leftarrow G_2 \backslash \{3\} = \emptyset$, $L = [5]$, $Z_{2,3}^L = \{0\}$. Let $h_{2,3} = 3 \in \mathbb{F}_{11} \backslash \{0\}$. Now, since $G_2 = \emptyset$, we move to the third row $k = 3$, where $G_3 = \{1,3,4,5\}$. Let $i = 3 \in G_3$ be chosen randomly. Then, $G_3 \leftarrow G_3 \backslash \{3\} = \{1,4,5\}$, $L = \{2,3\}$, $Z_{3,3}^L = \{0\}$. Let $h_{3,3} = 4 \in \mathbb{F}_{11} \backslash \{0\}$. Let $i = 4 \in G_3$ be chosen at random. Then, $G_3 \leftarrow G_3 \backslash \{4\} = \{1,5\}$, $L = \{2,3,4\}$, $Z_{3,4}^L = \{5\}$. Let $h_{3,4} = 3 \in \mathbb{F}_{11} \backslash \{5\}$. Now, let $i = 5 \in G_3$ be chosen randomly. Then, $G_3 \leftarrow G_3 \backslash \{5\} = \{1\}$, $L = \{2,3,4,5\}$, $Z_{3,5}^L = \{6,10\}$. Let $h_{3,5} = 2 \in \mathbb{F}_{11} \backslash \{6,10\}$. Now, $i = 1 \in G_3$, $G_3 \leftarrow G_3 \backslash \{1\} = \emptyset$, $L = [5]$, and $Z_{3,1}^L = \{0\}$. Let $h_{3,1} = 5 \in \mathbb{F}_{11} \backslash \{0\}$.

Thus, the encoding matrix $\boldsymbol{H} \in \mathbb{F}_{11}^{3 \times 5}$ will be as follows

$$\boldsymbol{H} = \mathrm{MCD}(\boldsymbol{G}, q) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 3 & 1 & 2 \\ 5 & 0 & 4 & 3 & 2 \end{bmatrix}.$$

It can be verified each submatrix $\boldsymbol{H}_{[k]}^L, k \in [3], L \subseteq [5]$ reaches its maximum possible rank.

VI. THE PROPOSED UMCD VERSUS THE MDS, RECURSIVE, AND ICC CODING SCHEMES

In this section, first we prove that the broadcast rate of the UMCD scheme is always at least as low as the broadcast rate of the MDS scheme. Then, two classes of index coding instances are characterized to show that the gap between the broadcast rates of the recursive and ICC coding schemes and the proposed UMCD coding scheme can grow linearly with the number of messages.

A. The UMCD versus the MDS Coding Scheme

First, we show that the binary matrix $\boldsymbol{G}$ of the UMCD scheme meets the linear code condition with a specific distance. Then, we prove that for any index coding instance $\mathcal{I}$, we have $\beta_{\mathrm{UMCD}}(\mathcal{I}) \leq \beta_{\mathrm{MDS}}(\mathcal{I})$.

Definition 16 (Linear Code Condition [21]). Assume the binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ with support sets $G_k, k \in [r]$ represents the generator matrix of an (m, r, d) linear code, where $d \leq m - r + 1$. Then, the linear code condition is defined as follows

$$|F_K| \geq d - 1 + |K|, \quad \forall K \subseteq [r], \tag{16}$$

where $F_K = \cup_k \in KG_k$. Note, F_K characterizes the indices of nonzero columns in matrix $\boldsymbol{G}_K$.

Lemma 2. Let $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$. Suppose that the binary matrix $\boldsymbol{G}_{[k]}, k \in [r]$ meets the condition in (16). Now, if for $i \in [m]$ the following three conditions are satisfied:

$$(1) \; |F_{[k]}| = d - 1 + k, \quad (2) \; \{i\} \cup A_i \subseteq F_{[k]}, \quad (3) \; |A_i| \geq d - 1,$$

then, we will have $\mathrm{mcm}(\boldsymbol{G}_{[k]}^{\{i\} \cup B_i}) = 1 + \mathrm{mcm}(\boldsymbol{G}_{[k]}^{B_i})$.

Proof. Refer to Appendix F-A. □

Lemma 3. The binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$, obtained at the end of the UMCD coding scheme, satisfies the linear code condition in (16) with $d = |A|_{min} + 1$.

Proof. Refer to Appendix F-B. □

Theorem 3. For an index coding instance $\mathcal{I}$, we have $\beta_{\mathrm{UMCD}}(\mathcal{I}) \leq \beta_{\mathrm{MDS}}(\mathcal{I})$.

Proof. Let matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ be the binary matrix of the UMCD scheme. Then, based on Lemma 3, matrix $\boldsymbol{G}$ meets the linear code condition in (16) with $d = |A|_{min} + 1$. Now, we show that the UMCD scheme guarantees that all the receivers will be satisfied at transmission $r \leq m - |A|_{min}$. If we set $r = m - |A|_{min}$, then due to (16), we have $|F_{[r]}| = m$. So, $F_{[r]} = [m]$.

Now, since $\{i\} \cup A_i \subseteq F_{[r]} = [m]$ and $|A_i| \geq |A|_{min} = d - 1, \forall i \in [m]$, the three conditions of Lemma 2 are met for $r = m - |A|_{min}$ and $d = |A|_{min} + 1$. So, we have

$$\mathrm{mcm}(\boldsymbol{G}_{[r]}^{\{i\} \cup B_i}) = 1 + \mathrm{mcm}(\boldsymbol{G}_{[r]}^{B_i}), \quad \forall i \in [m], \tag{17}$$

which means that the decoding condition in (6) is met for each receiver $u_i, i \in [m]$. So, we have $r \leq m - |A|_{min}$, which gives $\beta_{\mathrm{UMCD}}(\mathcal{I}) \leq \beta_{\mathrm{MDS}}(\mathcal{I})$. □

B. The UMCD versus the Recursive Coding Scheme

In this subsection, a class of index coding instances are characterized to show that the gap between the broadcast rates of the recursive and the proposed UMCD coding schemes can grow linearly with the number of messages.

Definition 17 ($\mathcal{I}_6(l)$: Class-$\mathcal{I}_6$ Index Coding Instances). Class-$\mathcal{I}_6$ index coding instances are defined as $\mathcal{I}_6(l) = \{(i|A_i), i \in [4l + 1]\}$, where the side information sets are as follows

$$\begin{cases} A_{2i-1} &= \{2j : j \in [2l] \backslash \{i\}\} \cup \{4l + 1\}, & \forall i \in [2l], \\ A_{2i} &= \{2i - 1\}, & \forall i \in [2l], \\ A_{4l+1} &= \{2i - 1 : i \in [2l]\}. \end{cases}$$

Note, the instance $\mathcal{I}_4$ is a special case of the class-$\mathcal{I}_6$ instances with $l = 1$, i.e., $\mathcal{I}_6(1) = \mathcal{I}_4$.

Theorem 4. For the class-$\mathcal{I}_6$ index coding instances, we have

$$\beta_R(\mathcal{I}_6(l)) - \beta_{\mathrm{UMCD}}(\mathcal{I}_6(l)) = l - \frac{1}{2}, \tag{18}$$

which scales as $\mathcal{O}(m)$. This means that the gap between the broadcast rates of the recursive coding scheme and our proposed UMCD grows linearly with the number of messages.

Proof can directly be derived from Propositions 8 and 9.

Proposition 8. *The broadcast rate of the proposed UMCD coding scheme for the class-$\mathcal{I}_6$ index coding instances is $\beta_{UMCD}(\mathcal{I}_6(l)) = 2l + 1$.*

Proof. First, note that $|A_{2i}| = 1$, $|A_{2i-1}| = 2l$, $\forall i \in [2l]$, and $|A_{4l+1}| = 2l$. Now, the UMCD coding scheme begins with one of the receivers indexed by $W = \{2i, i \in [2l]\}$ $u_{2i}, \forall i \in [2l]$. It can checked that for the first $2l$ transmissions, we have $w = \{2i\}$, $G_i = \{2i\} \cup A_{2i} = \{2i-1, 2i\}, \forall i \in [2l]$. So, the first $2l$ rows of the binary matrix will be set as follows

$$G_{[2l]} = \begin{bmatrix} 1 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & 1 & 0 \end{bmatrix}, \quad (19)$$

which satisfies all the receivers $u_{2i}, \forall i \in [2l]$. Now, for the round $2l+1$, we have $W = \{2i-1, i \in [2l+1]\}$. Let $w = 2i-1$, for some randomly chosen $i \in [2l]$. To determine whether other receivers $u_{2j-1}, j \neq i \in [2l]$ can also decode their message, we study the mcm of the following matrix:

$$G_{[2l+1]}^{\{2j-1\} \cup B_{2j-1}} = \begin{bmatrix} 1 & 0 & \ldots & 0 & 1 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix}. \quad (20)$$

It can be seen that the elements of the main diagonal are all equal to 1. Thus, the decoding condition in (3) holds for all receiver $u_{2i}, i \in [2l]$. Now, for receiver u_{4l+1}, we have

$$G_{[2l+1]}^{\{4l+1\} \cup B_{4l+1}} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix}. \quad (21)$$

Similarly, it can be seen that the diagonal elements are all set to 1. Thus, the decoding condition in (3) holds for receiver u_{4l+1}, which completes the proof. $\qquad\square$

Proposition 9. *For the class-$\mathcal{I}_6$ index coding instances, we have $\beta_R(\mathcal{I}_6(l)) = 3l + \frac{1}{2}$.*

Proof. The proof of Proposition 9 appears in Appendix G of supplemental material section. $\qquad\square$

Proposition 10. *For the class-$\mathcal{I}_6$ index coding instances, the UMCD coding scheme is optimal.*

Proof. It can be easily seen that subgraph $M = \{2i : i \in [2l]\} \cup \{4l + 1\}$ is acyclic. Therefore, $\beta_{\text{MAIS}}(\mathcal{I}_6(l)) \geq 2l + 1$, which is met by the UMCD coding scheme. This proves the optimality of the UMCD coding for $\mathcal{I}_6(l)$. $\qquad\square$

C. The UMCD versus the ICC Coding Scheme

In this subsection, a class of index coding instances are characterized to show that the gap between the broadcast rates of the ICC and the proposed UMCD coding schemes can grow linearly with the number of messages.

Definition 18 ($\mathcal{I}_7(l)$**: Class-$\mathcal{I}_7$ Index Coding Instances).** Class-$\mathcal{I}_7$ index coding instances are defined as $\mathcal{I}_7(l) = \{(i|A_i), i \in [5l + 3]\}$, where the side information sets are as follows

$$A_i = \begin{cases} E_{L_1} \cup (L_2 \backslash \{i + (2l+1)\}) \cup L_3, & \forall i \in O_{L_1}, \\ O_{L_2} \cup (L_1 \backslash \{i\}) \cup L_3, & \forall i \in E_{L_1}, \\ O_{L_2} \cup (L_1 \backslash \{i - (2l+1)\}) \cup L_3, & \forall i \in E_{L_2}, \\ E_{L_1} \cup (L_2 \backslash \{i\}) \cup L_3, & \forall i \in O_{L_2}, \\ \{2(i-4l)+3, 2(i-3l-2)\}, & \forall i \in L_3, \end{cases}$$
$$(22)$$

where, $L_1 = [2l+1]$, $L_2 = [2l+2 : 4l+2]$, and $L_3 = [4l+3 : 5l+3]$. Here, O_{L_i}, E_{L_i}, respectively, denote a set whose elements are the odd and even numbers inside L_i, $i = 1, 2$.

Example 10. For the $\mathcal{I}_7(l = 1)$, we have $m = 8$, and

$$A_1 = \{2, 5, 6, 7, 8\}, \quad A_2 = \{1, 3, 5, 7, 8\}, \quad A_3 = \{2, 4, 5, 7, 8\},$$
$$A_4 = \{2, 3, 5, 7, 8\}, \quad A_5 = \{2, 3, 6, 7, 8\}, \quad A_6 = \{1, 2, 5, 7, 8\},$$
$$A_7 = \{1, 4\}, \qquad\quad A_8 = \{3, 6\}.$$

Theorem 5. *For the class-$\mathcal{I}_7$ index coding instances, we have*

$$\beta_{ICC}(\mathcal{I}_7(l)) - \beta_{UMCD}(\mathcal{I}_7(l)) \geq 2l - \frac{1}{2}, \quad (23)$$

which scales as $\mathcal{O}(m)$. This means the gap between the broadcast rates of the ICC coding scheme and our proposed UMCD coding scheme grows linearly with the number of messages.

Proof. The proof of Theorem 5 appears in Appendix H of supplemental material section. $\qquad\square$

Proposition 11. *For the class-$\mathcal{I}_7$ index coding instances, the UMCD coding scheme is optimal.*

Proof. It can be easily seen that subgraph $M = L_3 \cup \{2\}$ is acyclic. Therefore, $\beta_{\text{MAIS}}(\mathcal{I}_7(l)) \geq l + 2$, which is met by the broadcast rate of the UMCD coding scheme. This proves the optimality of the UMCD coding for $\mathcal{I}_7(l)$. $\qquad\square$

VII. EXTENSIONS OF THE UMCD CODING SCHEME

Inspired by the PPC and FPCC coding schemes, in this section, we extend the UMCD coding scheme to its partial and fractional versions, where each is, respectively, motivated through Examples 11 and 12.

A. Partial UMCD (P-UMCD) Coding Scheme

Example 11. Consider the index coding instance $\mathcal{I}_8 = \{(1|2, 5), (2|3, 4), (3|2, 4), (4|2, 3), (5|1, 4)\}$. For this instance $\beta_{\text{UMCD}}(\mathcal{I}_8) = 3$, while $\beta(\mathcal{I}_8) = 2$, indicating that the UMCD is suboptimal. However, we first partition $\mathcal{I}_8$ into subinstances $M_1 = \{1, 5\}$ and $M_2 = \{2, 3, 4\}$. Then we apply the UMCD coding scheme for each subinstance separately, which results in $\beta_{\text{UMCD}}(M_1) = \beta_{\text{UMCD}}(M_2) = 1$. This achieves the broadcast rate.

Definition 19 (P-UMCD Coding Scheme). Given an index coding instance $\mathcal{I} = \{(i|A_i), i \in [m]\}$, the P-UMCD coding scheme first partitions $\mathcal{I}$ into subinstances $M_1, \ldots, M_n, n \leq m$, satisfying the condition in (2). Then, the UMCD coding scheme is used for solving each subinstance individually. So, the broadcast rate of the P-UMCD will be achieved as below

$$\beta_{\text{P-UMCD}}(M_j, j \in [n]) = \sum_{j \in [n]} \beta_{\text{UMCD}}(M_j). \quad (24)$$

Theorem 6. *Given an index coding instance $\mathcal{I}$, the broadcast rate $\beta(\mathcal{I})$ is upper bounded by $\beta_{P\text{-}UMCD}(\mathcal{I})$, which is the solution to the following optimization problem*

$$\min_{M_j, j \in [n]} \beta_{P\text{-}UMCD}(M_j, j \in [n]), \quad (25)$$

subject to the constraint in (2).

Proposition 12. *For the index coding instance $\mathcal{I}$, we have $\beta_{P\text{-}UMCD}(\mathcal{I}) \leq \beta_{PCC}(\mathcal{I})$.*

Proof. This can directly be concluded from Theorem 3, which states that for each subinstance $M_j, j \in [n]$, we always have $\beta_{\text{UMCD}}(M_j) \leq \beta_{\text{MDS}}(M_j)$. $\square$

B. Fractional Partial UMCD (FP-UMCD) Coding Scheme

Example 12. Consider the index coding instance $\mathcal{I}_9 = \{(1|2,5), (2|1,3), (3|2,4), (4|3,5), (5|1,4)\}$. For this instance, we have $\beta_{\text{UMCD}}(\mathcal{I}_9) = \beta_{\text{P-UMCD}}(\mathcal{I}_9) = 3$, while $\beta(\mathcal{I}_9) = 2.5$, which requires a vector index code to be optimal. This can be achieved by applying time sharing over the subinstances $M_1 = \{1,2\}$, $M_2 = \{2,3\}$, $M_3 = \{3,4\}$, $M_4 = \{4,5\}$, and $M_5 = \{1,5\}$, each solved by the UMCD coding scheme.

Definition 20 (FP-UMCD Coding Scheme). Given an index coding instance $\mathcal{I} = \{(i|A_i), i \in [m]\}$, the FP-UMCD coding scheme applies time sharing over the UMSD coding solution of the subinstances $M_j, j \in [n]$ for some $n \in 2^{[m]}$, as follows

$$\beta_{\text{FP-UMCD}}(M_j, \gamma_j, j \in [n]) = \sum_{j \in [n]} \gamma_j \beta_{\text{UMCD}}(M_j), \quad (26)$$

such that $\gamma_j \in [0,1], \quad \sum_{j \in P_i} \gamma_j \geq 1, \quad \forall i \in [m], \quad (27)$

where $P_i = \{j \in [n], i \in M_j\}$. Note that subsets $M_j, j \in [n]$ may have overlap each other.

Theorem 7. *Given an index coding instance $\mathcal{I}$, the broadcast rate $\beta(\mathcal{I})$ is upper bounded by $\beta_{FP\text{-}UMCD}(\mathcal{I})$, which is the solution of the following optimization problem*

$$\min_{M_j, \gamma_j, j \in [n]} \beta_{FP\text{-}UMCD}(M_j, \gamma_j, j \in [n]), \quad (28)$$

subject to the constraint in (27).

Proposition 13. *For an index coding instance $\mathcal{I}$, we have $\beta_{FP\text{-}UMCD}(\mathcal{I}) \leq \beta_{FPCC}(\mathcal{I})$.*

Proof. This can directly be concluded from Theorem 3, which states that for each subinstance $M_j, j \in [n]$, we always have $\beta_{\text{UMCD}}(M_j) \leq \beta_{\text{MDS}}(M_j)$. $\square$

Remark 6. Note that the broadcast rate of the UMCD algorithm for the subinstance M_i, i.e., $\beta_{UMCD}(M_i), i \in [n]$, is achieved independently of using the MCD algorithm. Thus, finding the optimal solution (optimal subinstances) of the P-UMCD and FP-UMCD schemes, respectively in (25) and (28)

can be achieved only using the MCM algorithm. Since the MCM is a polynomial-time algorithm, it can be shown that the computational complexity of finding the optimal solution of (25) and (28) can be achieved, respectively, from the optimal solution of the PCC and FPCC schemes, respectively in (33) and (35), by a polynomial reduction. Once the optimal solutions (25) and (28) are found using the MCM algorithm, then the encoding matrix for each optimal subinstance can be generated using the MCD algorithm. This highlights the importance of separating the MCM and the MCD algorithms in the UMCD coding scheme to reduce the computational complexity of the P-UMCD and FP-UMCD coding schemes.

Remark 7. Among all the existing coding schemes, only min-rank and composite coding schemes are optimal for index coding instances of up to and including five receivers. Now, we find that the proposed FP-UMCD coding scheme can also achieve the broadcast rate of the instances with up to and including five receivers (9846 non-isomorphic instances). However, for six receivers, there are some index coding instances for which the FP-UMCD coding scheme is suboptimal. One of these instances is illustrated in the following example.

Example 13. For the index coding instance $\mathcal{I}_{10} = \{(1|4,5), (2|1,6), (3|1,2,4,5,6), (4|1,2,3), (5|2,3), (6|3,4)\}$, we have $\beta(\mathcal{I}_{10}) = 3$, which can be achieved by index code $\{y_1 = x_1 + x_4 + x_5, y_2 = x_1 + x_2 + x_6, y_3 = x_2 + x_3 + x_5\}$. However, it can be verified that $\beta_{\text{FP-UMCD}}(\mathcal{I}_{10}) = \frac{10}{3} < \beta_{\text{ICC}}(\mathcal{I}_{10}) = \beta_{\text{R}}(\mathcal{I}_{10}) = 3.5$, implying that while the proposed FP-UMCD coding scheme outperforms both the recursive and ICC coding schemes, it is still suboptimal for $\mathcal{I}_{10}$.

VIII. CONCLUSION

In this paper, a new index coding scheme, referred to as update-based maximum column distance (UMCD) was proposed in which for each step of transmission, a linear coded message is designed with the aim of satisfying at least one receiver with the minimum size of side information. The problem is updated after each transmission using the polynomial-time Hopcroft-Karp algorithm, which is able to identify the satisfied receivers in each transmission. The maximum column distance (MCD) algorithm was proposed to generate the encoding matrix of the UMCD coding scheme such that each subset of its columns achieves its maximum possible rank. Several index coding instances were provided to show that the UMCD can outperform the ICC [14] and recursive [16] schemes. Moreover, we proved that the broadcast rate of the proposed UMCD scheme is never larger than the MDS coding scheme. Then, we characterized two classes of index coding instances for which the gap between the broadcast rates of the recursive and ICC coding schemes and the UMCD coding scheme grows linearly with the number of messages. The UMCD coding scheme was extended to its fractional version by dividing the instance into subinstances, and then applying the time sharing over their UMCD solutions. The fractional UMCD is optimal for all index coding instances with up to and including five receivers. Extending the UMCD coding scheme for the more general index coding scenarios, including the groupcast index coding [22]–[25], the secure index coding [26]–[30], and the distributed index coding [11], [31] would be directions for future studies.

APPENDIX A
PROOF OF PROPOSITION 1

First note that we must have rank $\boldsymbol{H}^{\{i\}} = t, \forall i \in [m]$. For the *if* condition, we suppose that (3) holds. Now, we show that there is a decoder function $\psi_\mathcal{I}^i(\boldsymbol{y}, S_i) = \boldsymbol{x}_i$, which can correctly decode $\boldsymbol{x}_i$ as follows

$$\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x} = \sum_{j \in [m]} \boldsymbol{H}^{\{j\}}\boldsymbol{x}_j - \sum_{j \in A_i} \boldsymbol{H}^{\{j\}}\boldsymbol{x}_j$$
$$= \boldsymbol{H}^{\{i\}}\boldsymbol{x}_i + \sum_{j \in B_i} \boldsymbol{H}^{\{j\}}\boldsymbol{x}_j. \quad (29)$$

Let rank $\boldsymbol{H}^{B_i} = c_i t$, then we partition $B_i = C_i \cup (B_i \backslash C_i)$ such that $|C_i| = c_i$ and $\{\boldsymbol{H}^{\{j\}} : j \in C_i\}$ represents the set of local encoding matrices whose columns are linearly independent, i.e., rank $\boldsymbol{H}^{C_i} = c_i t$. Then, the remaining local encoding matrices $\boldsymbol{H}^{\{l\}}, \forall l \in (B_i \backslash C_i)$ can be expressed as

$$\boldsymbol{H}^{\{l\}} = \sum_{j \in C_i} \boldsymbol{H}^{\{j\}}\boldsymbol{p}_{l,j},$$

where, $\boldsymbol{p}_{l,j} \in \mathbb{F}_q^{t \times 1}, \forall j \in C_i$. So, (29) will be equal to

$$\boldsymbol{H}^{\{i\}}\boldsymbol{x}_i + \sum_{j \in C_i} \boldsymbol{H}^{\{j\}}(\boldsymbol{x}_j + \sum_{l \in B_i \backslash C_i} \boldsymbol{p}_{l,j}\boldsymbol{x}_l). \quad (30)$$

Since (3) holds, then all the columns of $\boldsymbol{H}^{\{i\} \cup C_i}$ are linearly independent. Hence, the set of messages $\{\boldsymbol{x}_i\} \cup \{\boldsymbol{x}_j + \sum_{l \in B_i \backslash C_i} \boldsymbol{p}_{l,j}\boldsymbol{x}_l, j \in C_i\}$ can be decoded by receiver u_i, which completes the proof for the *if* condition.

Conversely, according to the polymatroidal bound [18], in order to decode all the messages correctly for the described system model, the following constraint must be met by any polymatroidal function $f : 2^{[m]} \to r$

$$f(\{i\} \cup B_i) \geq f(\{i\}) + f(B_i), \quad \forall i \in [m].$$

Since the rank function is polymatroidal, by setting $f(L) = \text{rank}(\boldsymbol{H}^L)$, we must have,

$$\text{rank } \boldsymbol{H}^{\{i\} \cup B_i} \geq \text{rank } \boldsymbol{H}^{\{i\}} + \text{rank } \boldsymbol{H}^{B_i}$$
$$= t + \text{rank } \boldsymbol{H}^{B_i}. \quad (31)$$

On the other hand, it can be easily observed that

$$\text{rank } \boldsymbol{H}^{\{i\} \cup B_i} \leq \text{rank } \boldsymbol{H}^{\{i\}} + \text{rank } \boldsymbol{H}^{B_i}$$
$$= t + \text{rank } \boldsymbol{H}^{B_i}. \quad (32)$$

Now, combining (31) and (32) will complete the proof.

APPENDIX B
A BRIEF OVERVIEW OF THE PCC, FPCC, RECURSIVE AND ICC CODING SCHEMES

Definition 21 (PCC Coding Scheme). In the PCC scheme, the index coding instance $\mathcal{I}$ is partitioned into $n \leq m$ subinstances $M_i, i \in [n]$, satisfying the condition in (2). Then, the PCC coding scheme solves each subinstace using the MDS code. So, the broadcast rate of the PCC scheme, $\beta_{\text{PCC}}(\mathcal{I})$, will be achieved by solving the following optimization problem

$$\min_{M_j, j \in [n]} \sum_{j \in [n]} \beta_{\text{MDS}}(M_j), \quad (33)$$

subject to the constraint in (2). Each subset $M_i, j \in [n]$ is called a partial clique set.

Remark 8. For the broadcast rate of the MDS code for each subset M_j, we have

$$\beta_{\text{MDS}}(M_j) = |M_j| - \min_{i \in M_j} |M_j \cap A_i|, \quad (34)$$

where $M_j \cap A_i$ is the local side information set of receiver u_i with regards to subset M_j.

Definition 22 (FPCC Coding Scheme). The FPCC scheme is an extension of the PPC scheme to the vector version through time-sharing over the MDS solutions of the subinstances $M_j, j \in [n]$ for some $n \in 2^m$. Thus, the broadcast rate of the FPCC scheme, $\beta_{\text{FPCC}}(\mathcal{I})$, will be equal to the solution of the following optimization problem

$$\min_{M_j, \gamma_j, j \in [n]} \sum_{j \in [n]} \gamma_j \beta_{\text{MDS}}(M_j), \quad (35)$$

subject to the constraints

$$\sum_{j \in P_i} \gamma_j \geq 1, \quad \forall i \in [m], \quad (36)$$

$$\gamma_j \in [0, 1], \quad (37)$$

where $P_i = \{j \in [n], i \in M_j\}$.

Definition 23 (Recursive Coding Scheme). The recursive scheme begins with fixing an initial broadcast rate $\beta_{\text{R}}(\{i\}) = 1, \forall i \in [m]$. Now, let $M_j \subsetneq M, j \in [n]$ for some $n \in 2^{|M|}$. Then, the broadcast rate for subinstance $\beta_{\text{R}}(M \subseteq [m])$ is recursively achieved as follows

$$\beta_{\text{R}}(M) = \min_{M_j, \gamma_j, j \in [n]} \max_{i \in M} \sum_{j \in [n], M_j \nsubseteq A_i} \gamma_j \beta_{\text{R}}(M_j), \quad (38)$$

subject to

$$\sum_{j \in P_i} \gamma_j \geq 1, \quad \forall i \in M, \quad (39)$$

$$\gamma_j \in [0, 1], \quad (40)$$

where $P_i = \{j \in [n], i \in M_j\}$.

Proposition 14. *Given an index coding instance $\mathcal{I}$, we have [18]*

$$\beta_R(\mathcal{I}) \leq \beta_{FPCC}(\mathcal{I}) \leq \beta_{PCC}(\mathcal{I}) \leq \beta_{MDS}(\mathcal{I}). \quad (41)$$

Definition 24 (ICC Coding Scheme). Given an index coding instance $\mathcal{I} = \{(i|A_i), i \in [m]\}$, the ICC coding scheme first identifies all the ICC-structured subgraphs such as $M \subseteq [m]$, as follows:

- Inner vertex set $J \subseteq M$, which consists of vertices, where there exists a path between each ordered pair of vertices $(i, j) \in J, i \neq j$ such that the path does not include any other vertex inside $J \backslash \{i, j\}$.
- J-path condition: There is only one J-path between any pair of vertices inside inner vertex set J, where the J-path is defined as follows: A path in which only the first and the last vertices (distinct vertices) belong to J.
- J-cycle condition: There is no J-cycle, where J-cycle is defined as follows: If in the J-path, the first and the last vertices are the same, it is considered as a J-cycle.

It has been shown that $\beta_{\text{ICC}}(M) = |M| - |J| + 1$, where all the savings in the transmissions are due to the inner vertex set.

Then, the broadcast rate of the ICC coding scheme for $\mathcal{I}$ is achieved as follows

$$\min_{M_j, \gamma_j, j \in [n]} \sum_{j \in [n]} \gamma_j \beta_{\text{ICC}}(M_j), \tag{42}$$

subject to the constraints in (36) and (37).

Conjecture 1. *Given an index coding instance $\mathcal{I}$, it is conjectured in [14] that*

$$\beta_{ICC}(\mathcal{I}) \leq \beta_{FPCC}(\mathcal{I}). \tag{43}$$

Remark 9. Neither the recursive nor the ICC is always outperformed by the other scheme. Consider the index coding instances $\mathcal{I}_4$ and $\mathcal{I}_5$, where $3 = \beta_{\text{ICC}}(\mathcal{I}_4) < \beta_{\text{R}}(\mathcal{I}_4) = 3.5$ and $2 = \beta_{\text{R}}(\mathcal{I}_5) < \beta_{\text{ICC}}(\mathcal{I}_5) = 2.5$.

Now, we provide Propositions 15 and 16, which will be used in the proof of Lemmas 4 and 6.

Proposition 15. *If for the constraint $\sum_{j \in P_i} \gamma_j \geq 1$ in the optimization problem of the FPCC, recursive, ICC, and the proposed FP-UMCD schemes, we consider only its equality case, i.e., $\sum_{j \in P_i} \gamma_j = 1, \forall i \in [m]$, the solution does not change.*

The proof can be achieved using standard techniques in the linear programming (LP).

Definition 25 (Minimal Partial Clique Set). The partial clique set $M \subseteq [m]$ is said to be minimal if its broadcast rate $\beta_{\text{MDS}}(M)$ cannot be reduced by being further partitioned into any subsets $M_j \subseteq M, j \in [n]$. This means

$$\beta_{\text{MDS}}(M) \leq \sum_{j \in [n]} \beta_{\text{MDS}}(M_j). \tag{44}$$

For example, all the cycles and cliques are minimal partial clique sets. Note that, any partial clique can be partitioned into minimal partial cliques.

Proposition 16. *The optimal solution of the FPCC scheme in (35) can always be expressed in minimal partial clique sets.*

Proof. First, we suppose that subsets $M_j, j \in [n]$ give the optimal solution in (35) such that subsets $M_j, j \in [r - 1]$ are minimal partial clique set, while $M_j, j \in [r : n]$ are not minimal. Thus,

$$\beta_{\text{FPCC}}(\mathcal{I}) = \sum_{j \in [n]} \gamma_j \beta_{\text{MDS}}(M_j) \tag{45}$$

$$= \sum_{j \in [r-1]} \gamma_j \beta_{\text{MDS}}(M_j) + \sum_{j \in [r:n]} \gamma_j \beta_{\text{MDS}}(M_j). \tag{46}$$

Now, assume that each non-minimal partial clique set $M_j, j \in [r : n]$ is further partitioned into $p_j + 1$ minimal partial clique sets $M_{j,z}, z \in [0 : p_j]$. Then, for the second term in (46), we have

$$\sum_{j \in [r:n]} \gamma_j \beta_{\text{MDS}}(M_j) \geq \sum_{j \in [r:n]} \gamma_j \sum_{z \in [0:p_j]} \beta_{\text{MDS}}(M_{j,z}) \tag{47}$$

Now, if for all $z \in [0 : p_j], \forall j \in [r : n]$, we set

$$M'_{j+z+\sum_{z \in [j-1]} p_z} = M_{j,z}, \tag{48}$$

$$\gamma'_{j+z+\sum_{z \in [j-1]} p_z} = \gamma_j, \tag{49}$$

then the right hand side of (47) will be equal to

$$\sum_{j \in [r:r+b]} \gamma'_j \beta_{\text{MDS}}(M'_j), \tag{50}$$

where $b = n + \sum_{j \in [r:n]} p_j$. Since subsets $M_{j,z}$ are minimal partial clique set, so are subsets M'_j. This means the second term of (46) can be expressed in minimal partial clique sets.

Now, from (45), ..., (50), we have

$$\sum_{j \in [n]} \gamma_j \beta_{\text{MDS}}(M_j) \geq \sum_{j \in [r+b]} \gamma'_j \beta_{\text{MDS}}(M'_j), \tag{51}$$

where for $j \in [r - 1]$, we set $M'_j = M_j$ and $\gamma'_j = \gamma_j$. This means that if subsets $M_j, j \in [n]$ with coefficients γ_j give the optimal solution in (35), so do subsets $M'_j, j \in [r + b]$ with coefficients γ'_j. Thus, the optimal solution of (35) can always be expressed in minimal partial cliques. □

Remark 10. In [14], the authors proved that for any minimal partial clique set, the ICC scheme performs at least as well as the FPCC scheme. Then, they provided Conjecture 1. Now, having proved Proposition 16, this conjecture is settled in the affirmative.

Definition 26 (Minimal Recursive Set). The recursive set $M \subseteq [m]$ is said to be minimal if its broadcast rate $\beta_{\text{R}}(M)$ cannot be reduced by being further partitioned into any subsets $M_j \subseteq M, j \in [n]$. This means

$$\beta_{\text{R}}(M) \leq \sum_{j \in [n]} \beta_{\text{R}}(M_j). \tag{52}$$

For instance, all the cliques and minimal cycles are minimal recursive sets.

Proposition 17. *The optimal solution of the recursive scheme in (38) can always be expressed in minimal recursive sets.*

Proof. The proof can be easily achieved by replacing β_{MDS} with β_{R} in the proof of Proposition 16. □

APPENDIX C
EXAMPLE 3

Since the proposed UMCD scheme is a scalar linear code, we make a comparison between the broadcast rates of the UMCD and the scalar binary minrank for the index coding instance $\mathcal{I}_3 = \{(1|3, 5, 6), (2|1, 4, 5, 6), (3|2, 4, 6), (4|1, 2, 3, 5), (5|1, 2, 3, 4), (6|1, 2, 4, 5)\}$, depicted in Figure 2c. One can verify that the broadcast rate of the scalar binary minrank is 3. Now, assume that the server first targets receivers with the minimum size of side information. It begins with receiver u_1 and transmits a linear combination of messages x_1 (requested message) and $x_i, i \in A_1$ as follows

$$y_1 = [h_{1,1} \ 0 \ h_{1,3} \ 0 \ h_{1,5} \ h_{1,6}] \ \boldsymbol{x}, \tag{53}$$

where $h_{1,j} \in \mathbb{F}_q$ and $\boldsymbol{x} = [x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6]$. This transmission satisfies only receiver u_1. Then, the server targets receiver u_3 by sending a linear combination of messages x_3 (requested) and $x_i, i \in A_3$ as follows

$$y_2 = [0 \ h_{2,2} \ h_{2,3} \ h_{2,4} \ 0 \ h_{2,6}] \ \boldsymbol{x}. \tag{54}$$

Now, we set $q = 3$, and fix $h_{1,1} = h_{1,3} = h_{1,5} = h_{1,6} = h_{2,2} = h_{2,3} = h_{2,4} = 1$ and $h_{2,6} = 2$. The reason of fixing $h_{2,6} = 2$ is to make the two column vectors $\boldsymbol{H}^{\{3\}} = [h_{1,3}\ h_{2,3}]^T$ and $\boldsymbol{H}^{\{6\}} = [h_{1,6}\ h_{2,6}]^T$ linearly independent so that receiver u_6 is able to to decode its requested message x_6. Now, it can be checked that the remaining receivers are able to decode their requested message upon receiving the coded messages y_1 and y_2. This index code achieves the MAIS bound and, so it is optimal for $\mathcal{I}_3$. This simple instance illustrates how the proposed UMCD coding scheme can outperform the scalar binary minrank.

APPENDIX D
PROOF OF PROPOSITIONS 3 AND 4

A. Proof of Proposition 3

First, since L is a circuit set of $\boldsymbol{H}_{[k-1]}$ and any circuit set is a minimal dependent set, we have

$$\boldsymbol{H}_{[k-1]}^{\{i\}} = \boldsymbol{H}_{[k-1]}^{L\setminus\{i\}} \boldsymbol{f},$$

where $\boldsymbol{f} = [f_1, \ldots, f_{|L|-1}]^T$ is a unique vector and f_i must be nonzero for all $i \in [|L|-1]$, since otherwise it contradicts (11). Note that vector $\boldsymbol{f}$ can be achieved using reduced row echelon form (rref) as follows

$$\text{rref}\ \left[\ \boldsymbol{H}_{[k-1]}^{L\setminus\{i\}}\ \Big|\ \boldsymbol{H}_{[k-1]}^{\{i\}}\ \right] = \left[\begin{array}{c|c} \boldsymbol{I}_{|L|-1} & \boldsymbol{f} \\ \hline \boldsymbol{0}_{|L|-1} & 0 \end{array} \right]$$
$$\overset{r}{=} \left[\ \boldsymbol{I}_{|L|-1}\ \Big|\ \boldsymbol{f}\ \right], \qquad (55)$$

where $\boldsymbol{H}_1 \overset{r}{=} \boldsymbol{H}_2$ means that matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ have an equal rank. Now, we have

$$\text{rref}\ \boldsymbol{H}_{[k]}^L = \text{rref}\ \left[\begin{array}{c|c} \boldsymbol{H}_{[k-1]}^{L\setminus\{i\}} & \boldsymbol{H}_{[k-1]}^{\{i\}} \\ \hline \boldsymbol{H}_{\{k\}}^{L\setminus\{i\}} & h_{k,i} \end{array} \right]$$
$$\overset{r}{=} \text{rref}\ \left[\begin{array}{c|c} \boldsymbol{I}_{|L|-1} & \boldsymbol{f} \\ \hline \boldsymbol{H}_{\{k\}}^{L\setminus\{i\}} & h_{k,i} \end{array} \right] \qquad (56)$$
$$= \text{rref}\ \left[\begin{array}{c|c} \boldsymbol{I}_{|L|-1} & \boldsymbol{f} \\ \hline \boldsymbol{0}_{|L|-1} & h_{k,i}^* \end{array} \right], \qquad (57)$$

where (56) is due to (55) and (57) is achieved by running rref over the last row such that

$$h_{k,i}^* = h_{k,i} - \boldsymbol{H}_{\{k\}}^{L\setminus\{i\}} \boldsymbol{f}.$$

Thus, only the value $h_{k,i} = \boldsymbol{H}_{\{k\}}^{L\setminus\{i\}} \boldsymbol{f}$, denoted by $h_{k,i}(L)$ will cause $h_{k,i}^* = 0$, which keeps the rank unchanged. In other words, choosing any value $h_{k,i} \in \mathbb{F}_q\setminus\{h_{k,i}(L)\}$ will lead to

$$\text{rank}\ \boldsymbol{H}_{[k]}^L = \text{rank}\ \boldsymbol{H}_{[k-1]}^L + 1 = |L|,$$

which means while L is a circuit set of $\boldsymbol{H}_{[k-1]}$, it will be an independent set of $\boldsymbol{H}_{[k]}$. Note that any field of size $q \geq 2$ guarantees that we can always find a value inside $\mathbb{F}_q\setminus\{h_k^L(\{i\})\}$.

B. Proof of Proposition 4

Since sets $L\setminus\{i\}$ and L, respectively, are a basis and a dependent set of $\boldsymbol{H}_{[k-1]}$, column $\boldsymbol{H}_{[k-1]}^{\{i\}}$ is linearly dependent on columns in $\boldsymbol{H}_{[k-1]}^{L\setminus\{i\}}$. So, set $L\setminus\{i\}$ can be partitioned into two subsets L' and L'' such that sets $L' \cup \{i\}$ and $L'' \cup \{i\}$, respectively, will be a circuit and an independent set of $\boldsymbol{H}_{[k-1]}$. For the circuit set $L' \cup \{i\}$, based on Proposition 3, if $h_{k,i} \in \mathbb{F}_q\setminus\{h_{k,i}(L' \cup \{i\})\}$, then $L' \cup \{i\}$ will be an independent set of $\boldsymbol{H}_{[k]}$.

It is obvious that any independent set of $\boldsymbol{H}_{[k-1]}$ will be an independent set of $\boldsymbol{H}_{[k]}$ as well. Thus, set $L'' \cup \{i\}$ is also an independent set of $\boldsymbol{H}_{[k]}$.

Thus, by setting $h_{k,i} \in \mathbb{F}_q\setminus Z_{k,i}^L$, where $Z_{k,i}^L = \{h_{k,i}(L' \cup \{i\})\}$, column $\boldsymbol{H}_{[k]}^{\{i\}}$ will be linearly independent of the columns in $\boldsymbol{H}_{[k]}^{L\setminus\{i\}}$. Now, since $L\setminus\{i\} \in \mathcal{B}_{k-1}$, by setting $h_{k,i} \in \mathbb{F}_q\setminus Z_{k,i}^L$, we have $L \in \mathcal{B}_k$, which completes the proof.

APPENDIX E
MORE DISCUSSION ON THE FIELD SIZE AND COMPLEXITY OF THE MCD ALGORITHM

Remark 11. According to the UMCD algorithm, the nonzero elements of matrix $\boldsymbol{G}$ are determined by the receivers with the minimum size of side information. This implies that many elements of $\boldsymbol{G}$ can be zero, which may reduce both the computational complexity and required filed size for the MCD algorithm.

Example 14. Consider the binary matrix $\boldsymbol{G}$ in

- Example 4, equation (8). First, it can be verified that for any $q \geq 2$, the size of the veto set for all the nonzero elements is always one. Thus, fixing the field size as $q \geq 2$ guarantees the existence of at least one element outside of the veto set for each nonzero element. Second, determining the veto value for only two elements $h_{3,2}$ and $h_{3,5}$ requires the rref operation over two submatrices of size 3×2, implying the low-complexity of the MCD algorithm for this case.

- Example 5, equation (9). First, it can be verified that for any $q \geq 3$, the maximum size of the veto set for the nonzero elements is always two. Thus, fixing the field size as $q \geq 3$ guarantees the existence of at least one element outside of the veto set for each nonzero element. Second, determining the veto value for only two elements $h_{2,1}$ and $h_{2,5}$ requires the rref operation over two submatrices of size 2×2, implying the low-complexity of the MCD algorithm for this case.

Remark 12. In the MCD algorithm, the aim is to design the elements of the encoding matrix $\boldsymbol{H}$ such that each of its submatrices achieves its maximum possible rank. However, in the following, we show that, for the UMCD scheme with the binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$, to satisfy all the receivers $u_i, i \in [m]$, we only need to make sure that m specific square submatrices of $\boldsymbol{H} \in \mathbb{F}_q^{r \times m}$ are full-rank. This can significantly reduce the complexity of the MCD algorithm as well as the required field size. Let $U_k = \{i \in [m], u_i \text{ is satisfied at transmission } k\}$ denote the set of receivers who are satisfied at transmission $k \leq r$ by the UMCD algorithm. This means that

$$\text{mcm}(\boldsymbol{G}_{[k]}^{\{i\}\cup B_i}) = 1 + \text{mcm}(\boldsymbol{G}_{[k]}^{B_i}), \quad i \in U_k. \qquad (58)$$

Now, assume that $\mathrm{mcm}(\boldsymbol{G}_{[k]}^{\{i\}\cup B_i}) = l_i$. This implies that there exists a square submatrix $\boldsymbol{G}_{L_i}^{\{i\}\cup B_i'}$, where $L_i \subseteq [k], |L_i| = l_i$ and $B_i' \subseteq B_i, |B_i'| = l_i - 1$, such that

$$\mathrm{mcm}(\boldsymbol{G}_{L_i}^{\{i\}\cup B_i'}) = 1 + \mathrm{mcm}(\boldsymbol{G}_{L_i}^{B_i'}), \quad i \in U_k. \quad (59)$$

Thus, if the elements of the encoding matrix $\boldsymbol{H}$ are designed such that all the submatrices $\boldsymbol{H}_{L_i}^{\{i\}\cup B_i'}$ are full-rank (which is possible due to (59)), then all the receivers $u_i, i \in U_k$ are able to decode their requested message in the k-th transmission.

Example 15. As seen in Example 5, in the first transmission, receiver u_5 and in the second transmission, other receivers are satisfied. Thus, $U_1 = \{5\}$ and $U_2 = \{1, 2, 3, 4\}$. It can be verified that for any field size $q \geq 2$, the decoding condition in (58) will be met by assigning value one to all nonzero elements of the encoding matrix. In fact, in the MCD algorithm, since columns 1 and 5 must be linearly independent, the size of the veto set for one of the elements $h_{2,1}$ and $h_{2,5}$ will be equal to two. Thus, $q \geq 3$ is required (and sufficient as said in Example 14). However, based on (58), the linear independence of columns 1 and 5 is not required (because $1 \in A_5$ and $5 \in A_1$). Thus, any field size $q \geq 2$ is sufficient for satisfying the decoding condition (58) for all receivers.

APPENDIX F
PROOF OF LEMMAS 2 AND 3

First, we begin with the following definitions and remarks.

Definition 27 (Potentially Full-rank Binary Matrix). Let $r \leq m$. We say that binary matrix $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ is potentially full-rank, if $\mathrm{mcm}(\boldsymbol{G}) = r$.

Definition 28 (Potential Pivot Column). Let $i \in [m]$. We say that the i-th column of $\boldsymbol{G} \in \mathbb{F}_2^{r \times m}$ is a potential pivot column if $\mathrm{mcm}(\boldsymbol{G}^{[m]}) = 1 + \mathrm{mcm}(\boldsymbol{G}^{[m]\setminus\{i\}})$.

Remark 13. It can be easily shown that the mcm function captures some properties of the rank function such as follows:

(i) Assume that $\boldsymbol{G}'$ is an $k \times r$ submatrix of the square matrix $\boldsymbol{G}_{r \times r}$, where $k \leq r$, and $\boldsymbol{G}$ is potentially full-rank, i.e., $\mathrm{mcm}(\boldsymbol{G}) = r$. Then $\mathrm{mcm}(\boldsymbol{G}') = k$. Moreover, every column of $\boldsymbol{G}$ is a potential pivot column.

(ii) Let $B \subseteq C$ and $i \notin C$. If $\mathrm{mcm}(\boldsymbol{G}^{\{i\}\cup C}) = 1 + \mathrm{mcm}(\boldsymbol{G}^C)$, then $\mathrm{mcm}(\boldsymbol{G}^{\{i\}\cup B}) = 1 + \mathrm{mcm}(\boldsymbol{G}^B)$.

Remark 14. Let set $C_K \subset F_K$ be obtained by removing $d-1$ elements from F_K in (16). Then, it can be easily verified that (i) C_K satisfies the linear code condition in (16) with $d = 1$, and (ii) for any set C_K satisfying (16) with $d = 1$, we have $\mathrm{mcm}(\boldsymbol{G}_{[k]}^{C_{[k]}}) = k, \forall k \in [r]$. This means, $\boldsymbol{G}_{[k]}^{C_{[k]}}$ is potentially full-rank.

A. Proof of Lemma 2

Let $B_i' = F_{[k]} \cap B_i$ denote the indices of nonzero columns related to B_i. So,

$$\mathrm{mcm}(\boldsymbol{G}_{[k]}^{\{i\}\cup B_i}) = \mathrm{mcm}(\boldsymbol{G}_{[k]}^{\{i\}\cup B_i'}). \quad (60)$$

Since $|F_{[k]}| = d - 1 + k$ and $|A_i| \geq d - 1$, then $|F_{[k]}\setminus A_i| \leq k$, and so, we have

$$|\{i\} \cup B_i'| \leq k. \quad (61)$$

This means that there exists a set $C_K \subseteq F_K$ such that $i \in C_{[k]}$, $B_i' \subseteq C_{[k]}$, and $|C_{[k]}| = k$. This set C_K can be achieved by removing $d-1$ elements from F_K. Based on Remark 14, we have

$$\mathrm{mcm}(\boldsymbol{G}_{[k]}^{C_{[k]}}) = k. \quad (62)$$

And because $|C_{[k]}| = k$, then $\boldsymbol{G}_{[k]}^{C_{[k]}}$ is a square matrix, and based on Remark 13-(i), each of its columns is a potential pivot column. Since $i \in C_{[k]}$, then its corresponding column will be a potential pivot column. Note, since $B_i' \subseteq C_{[k]}$, based on Remark 13-(ii), the corresponding column of the i-th element in matrix $\boldsymbol{G}^{\{i\}\cup B_i'}$ is a potential pivot column.

B. Proof of Lemma 3

The proof is achieved by induction.

- Consider the condition in (16) for $r = 1$. Since $|A_i| \geq |A|_{\min}, \forall i \in [m]$, we have $|F_{\{k\}}| = |G_k| = |\{i\} \cup A_i| \geq |A|_{\min} + 1$.

- As the induction hypothesis, we suppose that the condition in (16) holds for $r = k-1$. Thus, (16) holds for $K = [k-1]$, which means

$$|F_{[k-1]}| \geq |A|_{\min} + k - 1. \quad (63)$$

- Now, we need to prove that the condition in (16) will also hold for $r = k$. Now, assume that the linear code condition does not hold for $K = [k]$. So,

$$|F_{[k]}| \leq |A|_{\min} + k - 1. \quad (64)$$

Based on (63) and (64), we have

$$F_{[k]} = F_{[k-1]} = |A|_{\min} + k - 1. \quad (65)$$

Let w denote the index of receiver u_w, which is selected by the UMCD coding scheme for the k-th transmission. Now, from (65), we must have

$$G_k = \{w\} \cup A_w \subseteq F_{[k-1]}. \quad (66)$$

Now, since $|A_w| \geq |A|_{\min} = d - 1$, the three conditions of Lemma 2 are met for $r = k-1$ and $d = |A|_{\min} + 1$. So, we will have

$$\mathrm{mcm}(\boldsymbol{G}_{[k-1]}^{\{w\}\cup B_w}) = 1 + \mathrm{mcm}(\boldsymbol{G}_{[k-1]}^{B_w}). \quad (67)$$

Hence, receiver u_w is able to decode its requested messages from the first $k-1$ transmissions. However, this contradicts the UMCD coding scheme's logic, where at each step of transmission, it picks a receiver which has not been satisfied by the previous transmissions (as it can be seen in Algorithm 1 that in each transmission, we remove i from N as $N \leftarrow N\setminus\{i\}$ when u_i is satisfied). This completes the proof.

REFERENCES

[1] A. Sharififar, N. Aboutorab, and P. Sadeghi, "Update-based maximum column distance coding scheme for index coding problem," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 575–580.

[2] Y. Birk and T. Kol, "Informed-Source Coding-On-Demand (ISCOD) over Broadcast Channels," *in Proc. IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1257–1264, 1998.

[3] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2745–2759, 2005.

[4] A. Sharififar, P. Sadeghi, and N. Aboutorab, "Broadcast rate requires nonlinear coding in a unicast index coding instance of size 36," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 208–213.

[5] ——, "On the optimality of linear index coding over the fields with characteristic three," *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 3250–3255, 2022.

[6] ——, "On the optimality of linear index coding over the fields with characteristic three," *IEEE Open Journal of the Communications Society*, pp. 1–1, 2022.

[7] H. Maleki, V. R. Cadambe, and S. A. Jafar, "Index Coding — An Interference Alignment Perspective," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5402–5432, 2014.

[8] E. Lubetzky and U. Stav, "Nonlinear index coding outperforming the linear optimum," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3544–3551, 2009.

[9] F. Arbabjolfaei, B. Bandemer, Y. Kim, E. Şaşoğlu, and L. Wang, "On the capacity region for index coding," in *2013 IEEE International Symposium on Information Theory*, 2013, pp. 962–966.

[10] Y. Liu, P. Sadeghi, and Y.-H. Kim, "Three-layer composite coding for index coding," in *2018 IEEE Information Theory Workshop (ITW)*. IEEE, 2018, pp. 1–5.

[11] Y. Liu, P. Sadeghi, F. Arbabjolfaei, and Y.-H. Kim, "Capacity theorems for distributed index coding," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 4653–4680, 2020.

[12] M. A. R. Chaudhry, Z. Asad, A. Sprintson, and M. Langberg, "On the complementary index coding problem," in *2011 IEEE International Symposium on Information Theory Proceedings*, 2011, pp. 244–248.

[13] Y. Birk and T. Kol, "Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2825–2830, 2006.

[14] C. Thapa, L. Ong, and S. J. Johnson, "Interlinked cycles for index coding: generalizing cycles and cliques," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3692–3711, 2017.

[15] H. Yu and M. J. Neely, "Duality codes and the integrality gap bound for index coding," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7256–7268, 2014.

[16] F. Arbabjolfaei and Y.-h. Kim, "Local time sharing for index coding," *in Proc. IEEE International Symposium on Information Theory*, pp. 286–290, 2014.

[17] J. E. Hopcroft and R. M. Karp, "A n5/2 algorithm for maximum matchings in bipartite," *12th Annual Symposium on Switching and Automata Theory (SWAT)*, vol. 24, no. 6, pp. 122–125, 1971.

[18] F. Arbabjolfaei and Y.-H. Kim, "Fundamentals of index coding," *Foundations and Trends® in Communications and Information Theory*, vol. 14, no. 3-4, pp. 163–346, 2018. [Online]. Available: http://dx.doi.org/10.1561/0100000094

[19] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Lexicographic products and the power of non-linear network coding," in *IEEE Annual Symposium on Foundations of Computer Science*. IEEE, 2011, pp. 609–618.

[20] S. E. Rouayheb, A. Sprintson, and C. Georghiades, "On the Index Coding Problem and Its Relation to Network Coding and Matroid Theory," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3187–3195, 2010.

[21] S. H. Dau, W. Song, and C. Yuen, "On the existence of mds codes over small fields with constrained generator matrices," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 1787–1791.

[22] A. S. Tehrani, A. G. Dimakis, and M. J. Neely, "Bipartite index coding," *in Proc. IEEE International Symposium on Information Theory*, pp. 2246–2250, 2012.

[23] K. Shanmugam, A. G. Dimakis, and M. Langberg, "Graph theory versus minimum rank for index coding," *in Proc. IEEE International Symposium on Information Theory*, pp. 291–295, 2014.

[24] S. Unal and A. B. Wagner, "A Rate-Distortion Approach to Index Coding," *IEEE Transactions on Information Theory*, vol. 62, no. 99, pp. 6359–6378, 2016.

[25] A. Sharififar, N. Aboutorab, Y. Liu, and P. Sadeghi, "Independent user partition multicast scheme for the groupcast index coding problem," in *2020 International Symposium on Information Theory and Its Applications (ISITA)*, 2020, pp. 314–318.

[26] V. Narayanan, V. M. Prabhakaran, J. Ravi, V. K. Mishra, B. K. Dey, and N. Karamchandani, "Private index coding," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 596–600.

[27] S. H. Dau, V. Skachek, and Y. M. Chee, "On the security of index coding with side information," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3975–3988, 2012.

[28] L. Ong, B. N. Vellambi, P. L. Yeoh, J. Kliewer, and J. Yuan, "Secure index coding: Existence and construction," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 2834–2838.

[29] M. M. Mojahedian, M. R. Aref, and A. Gohari, "Perfectly secure index coding," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7382–7395, 2017.

[30] Y. Liu, P. Sadeghi, N. Aboutorab, and A. Sharififar, "Secure index coding with security constraints on receivers," in *2020 International Symposium on Information Theory and Its Applications (ISITA)*, 2020, pp. 319–323.

[31] Y. Liu, P. Sadeghi, F. Arbabjolfaei, and Y.-H. Kim, "On the capacity for distributed index coding," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 3055–3059.

Arman Sharififar received his B.Sc. degree in Electrical Engineering from Bahonar University, Iran. He completed his M.Sc. degree in the field of coding and communication systems at Shiraz University, Iran. Currently, he is pursuing his PhD degree at the School of Engineering and Information Technology, the University of New South Wales, Canberra, Australia. His research interests include index and network coding, private and secured index coding, coded caching, and space-time coding in the MIMO systems.

Neda Aboutorab (S'09-M'12-SM'17) is currently a Senior Lecturer at the School of Engineering and Information Technology at the University of New South Wales, Canberra, Australia. She received her PhD in Electrical Engineering from the University of Sydney, Australia, in 2012. From 2012-2015 and before joining the University of New South Wales, she was a Postdoctoral Research Fellow at the Research School of Engineering, the Australian National University. Her research interests include index and network coding, applied information theory, big data caching and storage systems, wireless communications and signal processing.

Parastoo Sadeghi (Senior Member, IEEE) received the bachelor's and master's degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1995 and 1997, respectively, and the Ph.D. degree in electrical engineering from the University of New South Wales, Sydney, NSW, Australia, in 2006. She is currently a Professor with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT, Australia. She has co-authored the book Hilbert Space Methods in Signal Processing (Cambridge University Press, 2013) and around 190 refereed journal articles and conference papers. Her research interests include information theory, data privacy, index coding, and network coding. She was a recipient of the 2019 Future Fellowship from the Australian Research Council. From 2016 to 2019, she served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. From 2019 to 2020, she served as a member on the Board of Governors of the IEEE Information Theory Society. She was the General Co-chair of the 2021 IEEE International Symposium on Information Theory.

APPENDIX G
PROOF OF PROPOSITION 9

The proof of Proposition 9 can be directly concluded from the following Lemmas 4 and 5. In Lemma 4, we prove that $\beta_{\text{FPCC}}(\mathcal{I}_6(l)) = 3l + \frac{1}{2}$. Then, in Lemma 5, it will be shown that the recursive scheme cannot outperform the FPCC scheme for the class-$\mathcal{I}_6$ index coding instances. The associated graph $\mathcal{G}_{\mathcal{I}_6}$ is depicted in Figure 5.

Lemma 4. *For the class-$\mathcal{I}_6$ index coding instances, we have* $\beta_{FPCC}(\mathcal{I}_6(l)) = 3l + \frac{1}{2}$.

Proof. Regarding Proposition 16, we only consider the minimal partial clique sets in class-$\mathcal{I}_6$ index coding instances. It can be verified that the minimal partial clique sets can be all characterized as follows

$$\begin{cases} M_{i,j} & = \{2i-1, 2i, 2j-1, 2j\}, & \forall i \neq j \in [2l], \\ M_{i,2l+1} & = \{2i-1, 4l+1\}, & \forall i \in [2l], \\ M_i & = \{i\}, & \forall i \in [4l+1]. \end{cases} \tag{68}$$

Moreover, it can also be easily seen that

$$\begin{cases} \beta_{\text{MDS}}(M_{i,j}) & = 3, & \forall i \neq j \in [2l], \\ \beta_{\text{MDS}}(M_{i,2l+1}) & = 1, & \forall i \in [2l], \\ \beta_{\text{MDS}}(M_i) & = 1, & \forall i \in [4l+1]. \end{cases}$$

Thus, the optimal solution for the FPCC scheme in (35) will be equal to (note, since $M_{i,j} = M_{j,i}, \forall i \neq j \in [2l]$, to avoid repetition, we only consider $i < j \in [2l]$)

$$\begin{aligned} \beta_{\text{FPCC}}(\mathcal{I}_6(l)) &= \min_{\substack{\gamma_{i,j},\ i<j\in[2l] \\ \gamma_{i,2l+1},\ i\in[2l] \\ \gamma_i,\ i\in[4l+1]}} \Big(\sum_{i<j\in[2l]} \gamma_{i,j}\ \beta_{\text{MDS}}(M_{i,j}) \\ &\quad + \sum_{i\in[2l]} \gamma_{i,2l+1}\ \beta_{\text{MDS}}(M_{i,2l+1}) + \sum_{i\in[4l+1]} \gamma_i\ \beta_{\text{MDS}}(M_i) \Big) \\ &= \min_{\substack{\gamma_{i,j},\ i<j\in[2l] \\ \gamma_{i,2l+1},\ i\in[2l] \\ \gamma_i,\ i\in[4l+1]}} \Big(\sum_{i<j\in[2l]} 3\gamma_{i,j} + \sum_{i\in[2l]} \gamma_{i,2l+1} + \sum_{i\in[4l+1]} \gamma_i \Big), \end{aligned} \tag{69}$$

subject to the constraints in (36), which will be expressed as follows (according to Proposition 15, we just consider the equality case and also we let $\gamma_{i,j} = \gamma_{j,i}, i < j \in [2l]$)

$$\Big(\sum_{\substack{j\in[2l] \\ j\neq i}} \gamma_{\frac{i+1}{2},j} \Big) + \gamma_{i,2l+1} + \gamma_i = 1, \ \forall i = 2i_1 - 1, i_1 \in [2l], \tag{70}$$

$$\Big(\sum_{\substack{j\in[2l] \\ j\neq i}} \gamma_{\frac{i}{2},j} \Big) + \gamma_i = 1, \qquad \forall i = 2i_1, i_1 \in [2l], \tag{71}$$

$$\Big(\sum_{i\in[2l]} \gamma_{i,2l+1} \Big) + \gamma_{4l+1} = 1, \qquad i = 4l+1. \tag{72}$$

Now, we add all the constraints in (70), (71) and (72) with each other, which will result in

$$\sum_{i<j\in[2l]} 4\gamma_{i,j} + \sum_{i\in[2l]} 2\gamma_{i,2l+1} + \sum_{i\in[4l+1]} \gamma_i = 4l+1. \tag{73}$$
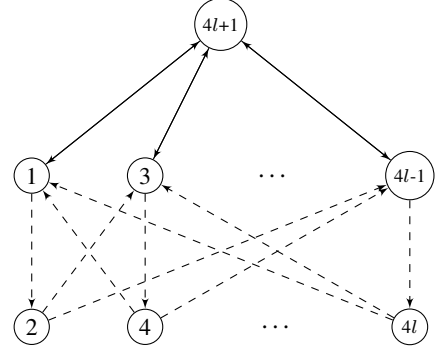

Figure 5: The class-$\mathcal{I}_6$ index coding instances for which $\beta_{\text{R}}(\mathcal{I}_6(l)) - \beta_{\text{UMCD}}(\mathcal{I}_6(l)) = l - \frac{1}{2}$.

Now, using (73), we can express (69) as follows

$$\begin{aligned} &\sum_{i<j\in[2l]} 3\gamma_{i,j} + \sum_{i\in[2l]} \gamma_{i,2l+1} + \sum_{i\in[4l+1]} \gamma_i = 4l+1 \\ &- \Big(\sum_{i<j\in[2l]} \gamma_{i,j} + \sum_{i\in[2l]} \gamma_{i,2l+1} \Big), \end{aligned} \tag{74}$$

Moreover, adding all the constraints in (36) and (72) will result in

$$\sum_{i<j\in[2l]} \gamma_{i,j} + \sum_{i\in[2l]} \gamma_{i,2l+1} + \sum_{i\in[2l+1]} \gamma_{2i-1} = \frac{2l+1}{2}, \tag{75}$$

From (74) and (75), we have

$$\begin{aligned} \beta_{\text{FPCC}}(\mathcal{I}_6(l)) &= 3l + \frac{1}{2} - \min_{\gamma_{2i-1}, i\in[2l+1]} \sum_{i\in[2l+1]} \gamma_{2i-1} \\ &= 3l + \frac{1}{2}, \end{aligned} \tag{76}$$

where (76) is achieved by setting $\gamma_{2i-1} = 0, \forall i \in [2l+1]$. This completes the proof. □

Lemma 5. *For the class-$\mathcal{I}_6$ index coding instances, we have* $\beta_R(\mathcal{I}_6(l)) = \beta_{FPCC}(\mathcal{I}_6(l))$.

Proof. First, it can be seen that the subsets in (68) are all minimal recursive sets for which the broadcast rate of the recursive and the MDS coding schemes are equal. Second, since none of the sets $M_{i,j}, i \neq j \in [2l]$ and $M_{i,2l+1}, i \in [2l]$ in (68) are subset of any side information set $A_i, i \in [m]$, then due to (38), the broadcast rate of the recursive coding scheme will be equal to the time sharing over the subsets in (68). Therefore, the recursive optimization problem will be reduced to the FPCC optimization problem, which completes the proof. □

APPENDIX H
PROOF OF THEOREM 5

The proof of Theorem 5 can directly be concluded from Propositions 18 and 19.

Proposition 18. *The broadcast rate of the proposed UMCD coding scheme for the class-$\mathcal{I}_7$ index coding instances is*

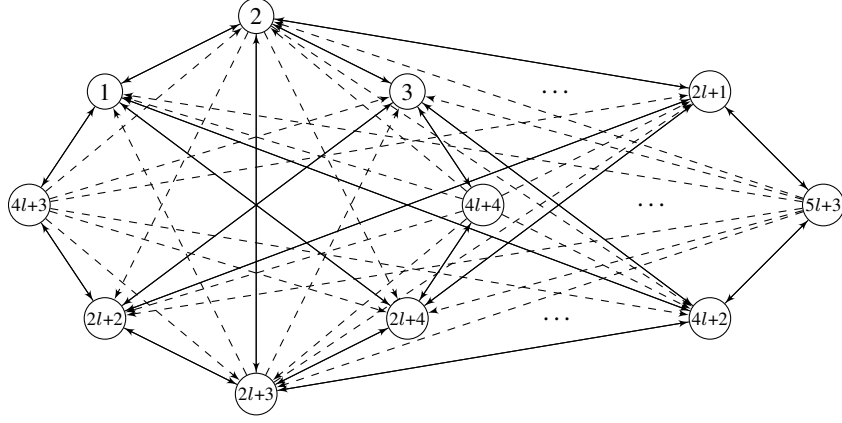$$\beta_{UMCD}(\mathcal{I}_7(l)) = l + 2.$$

Figure 6: The class-$\mathcal{I}_7$ index coding instances for which $\beta_{\text{ICC}}(\mathcal{I}_7(l)) - \beta_{\text{UMCD}}(\mathcal{I}_7(l)) \geq 2l - \frac{1}{2}$.

Proof. First, without loss of generality, we rearrange the elements of the message vector as follows

$$\boldsymbol{x} = [\boldsymbol{o}_{l_1} \mid \boldsymbol{e}_{l_2} \mid \boldsymbol{l}_3 \mid \boldsymbol{e}_{l_1} \mid \boldsymbol{o}_{l_2}], \tag{77}$$

where $\boldsymbol{o}_{l_1}, \boldsymbol{e}_{l_2}, \boldsymbol{l}_3, \boldsymbol{e}_{l_1}, \boldsymbol{o}_{l_2}$ represent the vectors whose elements are selected, respectively, from the sets O_{L_1}, E_{L_2}, L_3, E_{L_1} and O_{L_2} in an ascending order. Now, note that $|A_i| = 4l + 2$, $\forall i \in L_1 \cup L_2$, and $|A_i| = 2$, $\forall i \in L_3$. Therefore, the UMCD coding scheme begins with the receivers indexed by $W = L_3$, having the minimum size of side information. By choosing $w = i + 4l + 1$, in the i-th transmission, $i \in [l+1]$, it can be seen that the binary matrix for the first $l+1$ transmissions will be equal to

$$\boldsymbol{G}_{[l+1]} = \begin{bmatrix} \boldsymbol{I}_{l+1} \mid \boldsymbol{I}_{l+1} \mid \boldsymbol{I}_{l+1} \mid \boldsymbol{0}_{(l+1) \times l} \mid \boldsymbol{0}_{(l+1) \times l} \end{bmatrix},$$

which satisfies all receivers $u_i, \forall i \in L_3$. Since the size of the side information of the remaining receivers is equal, one of them is selected randomly. Assume that $w = j$ for some $j \in O_{L_1}$. Then, the binary matrix $\boldsymbol{G}_{[l+2]}$ will be

$$\begin{bmatrix} \boldsymbol{I}_{l+1} & \boldsymbol{I}_{l+1} & \boldsymbol{I}_{l+1} & \boldsymbol{0}_{(l+1) \times l} & \boldsymbol{0}_{(l+1) \times l} \\ \hline \boldsymbol{0}_{l+1}(\frac{j+1}{2}) & \boldsymbol{1}_{l+1}(\frac{j+1}{2}) & \boldsymbol{1}_{l+1} & \boldsymbol{1}_l & \boldsymbol{1}_l \end{bmatrix}, \tag{78}$$

where, $\boldsymbol{1}_l$ and $\boldsymbol{0}_l$, respectively, denote a row vector of size l whose elements are all set to one and zero. Moreover, $\boldsymbol{1}_l(j)$ and $\boldsymbol{0}_l(j)$, respectively, represent a row vector of size l whose elements are all set to one and zero, except the j-th element. Now, we show that all the remaining unsatisfied receivers will be able to decode their requested message using $\boldsymbol{G}_{[l+2]}$. It can be verified that binary matrix $\boldsymbol{G}_{[l+2]}^{\{i\} \cup B_i}$ for the remaining receivers is as follows

$$\begin{cases} \begin{bmatrix} \boldsymbol{I}_{l+1} & \boldsymbol{0}_{l+1}^T(\frac{i+1}{2}) \\ \hline \boldsymbol{0}_{l+1}(\frac{j+1}{2}) & 1 \end{bmatrix}, & \forall i \in O_{L_1} \backslash \{j\}, \\[2em] \begin{bmatrix} \boldsymbol{I}_{l+1} & \boldsymbol{0}_{l+1}^T \\ \hline \boldsymbol{1}_{l+1}(\frac{j+1}{2}) & 1 \end{bmatrix}, & \forall i \in E_{L_1}, \\[2em] \begin{bmatrix} \boldsymbol{0}_{l+1}^T(\frac{i-2k}{2}) & \boldsymbol{I}_{l+1} \\ \hline 0 & \boldsymbol{1}_{l+1}(\frac{j+1}{2}) \end{bmatrix}, & \forall i \in E_{L_2} \backslash \{j+2k+1\}, \\[2em] \begin{bmatrix} \boldsymbol{0}_{l+1}^T(\frac{i-2k}{2}) & \boldsymbol{I}_{l+1} \\ \hline 1 & \boldsymbol{1}_{l+1}(\frac{j+1}{2}) \end{bmatrix}, & i = j+2k+1 \in E_{L_2}, \\[2em] \begin{bmatrix} \boldsymbol{I}_{l+1} & \boldsymbol{0}_{l+1}^T \\ \hline \boldsymbol{0}_{l+1}(\frac{j+1}{2}) & 1 \end{bmatrix}, & \forall i \in O_{L_2}. \end{cases}$$

Now, it can be easily observed that for each $\boldsymbol{G}_{[l+2]}^{\{i\} \cup B_i}$, the elements of the main diagonal can be all fixed to one. Thus, the decoding condition (6) will hold for all the receivers, which completes the proof. One can check that for the other possible selection of w in transmission $l + 2$, the broadcast rate will be the same. $\square$

Proposition 19. *For the broadcast rate of the ICC coding scheme for the class-$\mathcal{I}_7$ index coding instances, we have* $\beta_{ICC}(\mathcal{I}_7(l)) \geq \frac{3(l+1)}{2}$.

The proof of Proposition 19 can be directly concluded from the following Lemmas 6 and 7. In Lemma 6, we prove that $\beta_{FPCC}(\mathcal{I}_7(l)) \geq \frac{3(l+1)}{2}$. Then, in Lemma 7, it will be shown that the ICC scheme cannot outperform the FPCC scheme for the class-$\mathcal{I}_7$ index coding instances.

Lemma 6. *For the class-$\mathcal{I}_7$ index coding instance, we have* $\beta_{FPCC}(\mathcal{I}_7(l)) \geq \frac{3(l+1)}{2}$.

Proof. First, we note that any subset $M_j \subseteq L_1 \cup L_2, j \in [n]$ can be expressed as follows

$$M_j = O_{L_1,j} \cup E_{L_1,j} \cup E_{L_2,j} \cup O_{L_2,j}, \tag{79}$$

where, $O_{L_{i},j} \subseteq O_{L_i}$ and $E_{L_{i},j} \subseteq E_{L_i}$ for $i = 1, 2$ and $\forall j \in \mathcal{P}([5l+3])$. Now, based on (79) and (22), it can be seen that the local side information $M_j \cap A_i$ will be equal to

$$\begin{cases} E_{L_1,j} \cup O_{L_2,j} \cup E_{L_2,j}\backslash\{i+(2l+1)\}, & \forall i \in O_{L_1,j}, \\ E_{L_1,j}\backslash\{i\} \cup O_{L_2,j} \cup O_{L_1,j}, & \forall i \in E_{L_1,j}, \\ E_{L_1,j} \cup O_{L_2,j} \cup O_{L_1,j}\backslash\{i-(2l+1)\}, & \forall i \in E_{L_2,j}, \\ E_{L_1,j} \cup O_{L_2,j}\backslash\{i\} \cup E_{L_2,j}, & \forall i \in O_{L_2,j}. \end{cases}$$

So, $|M_j \cap A_i|$ will be equal to

$$\begin{array}{ll} e_{L_1,j} + o_{L_2,j} + e_{L_2,j} - |E_{L_2,j} \cap \{i+(2l+1)\}|, & \forall i \in O_{L_1,j}, \\ e_{L_1,j} + o_{L_2,j} + o_{L_1,j} - |E_{L_1,j} \cap \{i\}|, & \forall i \in E_{L_1,j}, \\ e_{L_1,j} + o_{L_2,j} + o_{L_1,j} - |O_{L_1,j} \cap \{i-(2l+1)\}|, & \forall i \in E_{L_2,j}, \\ e_{L_1,j} + o_{L_2,j} + e_{L_2,j} - |O_{L_2,j}\backslash\{i\}|, & \forall i \in O_{L_2,j}, \end{array}$$

where $e_{L_i,j} = |e_{L_i,j}|$, $o_{L_i,j} = |O_{L_i,j}|$ for $i = 1, 2$. Now, for the minimum size of the local side information, we have

$$\begin{aligned} \min_{i \in M_j} |M_j \cap A_i| &= e_{L_1,j} + o_{L_2,j} + \min\Big\{ e_{L_2,j} \\ &\quad - |E_{L_2,j} \cap \{i+(2l+1)\}|, \ o_{L_1,j} - 1, \\ &\quad o_{L_1,j} - |O_{L_1,j} \cap \{i-(2l+1)\}|, \ e_{L_2,j} - 1 \Big\} \\ &= e_{L_1,j} + o_{L_2,j} - 1 + \min\big\{ o_{L_1,j}, e_{L_2,j} \big\}. \quad (80) \end{aligned}$$

Now, for the broadcast rate of the MDS code (34) for each subset M_j in (79), we have

$$\begin{aligned} \beta_{\text{MDS}}(M_j) &= |M_j| - \min_{i \in M_j} |M_j \cap A_i| \\ &= e_{L_2,j} + o_{L_1,j} + 1 - \min\big\{ o_{L_1,j}, e_{L_2,j} \big\} \\ &\geq e_{L_2,j} + o_{L_1,j} + 1 - \frac{e_{L_2,j} + o_{L_1,j}}{2} \\ &= \frac{e_{L_2,j} + o_{L_1,j}}{2} + 1. \quad (81) \end{aligned}$$

It can also be verified that all the minimal partial clique sets which include receivers indexed by L_3, are characterized as follows

$$\begin{cases} M'_j = \{2j-1, 4l+2+j\}, & \forall j \in [l+1], \\ M''_j = \{2l+2j, 4l+2+j\}, & \forall j \in [l+1], \\ M'''_j = \{4l+2+j\}, & \forall j \in [l+1], \end{cases}$$

where their MDS broadcast rate is as below

$$\beta_{\text{MDS}}(M'_j) = \beta_{\text{MDS}}(M''_j) = \beta_{\text{MDS}}(M'''_j) = 1, \forall j \in [l+1]. \quad (82)$$

Now, having (80) and (82), the broadcast rate of the FPCC scheme is equal to

$$\begin{aligned} \beta_{\text{FPCC}}(\mathcal{I}_7(l)) &= \min_{\substack{\gamma_j, M_j, j\in[n] \\ M'_j, M''_j, M'''_j, \\ \gamma'_j, \gamma''_j, \gamma'''_j, j\in[l+1]}} \Bigg[\sum_{j\in[n]} \gamma_j \beta_{\text{MDS}}(M_j) \\ &\quad + \sum_{j\in[l+1]} \Big(\gamma'_j \beta_{\text{MDS}}(M'_j) + \gamma''_j \beta_{\text{MDS}}(M''_j) + \gamma'''_j \beta_{\text{MDS}}(M'''_j) \Big) \Bigg] \\ &\geq \min_{\substack{\gamma_j, e_{L_2,j}, o_{L_1,j} \\ j\in[n], \\ \gamma'_j, \gamma''_j, \gamma'''_j \\ j\in[l+1]}} \Bigg[\frac{1}{2} \sum_{j\in[n]} \gamma_j \Big(e_{L_2,j} + o_{L_1,j} + 1 \Big) \\ &\quad + \sum_{j\in[l+1]} \Big(\gamma'_j + \gamma''_j + \gamma'''_j \Big) \Bigg], \quad (83) \end{aligned}$$

subject to the following constraints

$$\Big(\sum_{j\in P_i} \gamma_j \Big) + \gamma'_{\frac{i+1}{2}} = 1, \qquad \forall i \in O_{L_1}, \quad (84)$$

$$\Big(\sum_{j\in P_i} \gamma_j \Big) + \gamma''_{\frac{i-2l}{2}} = 1, \qquad \forall i \in E_{L_2}, \quad (85)$$

$$\Big(\sum_{j\in P_i} \gamma_j \Big) = 1, \qquad \forall i \in E_{L_1} \cup O_{L_2}, \quad (86)$$

$$\gamma'_i + \gamma''_i + \gamma'''_i = 1, \qquad \forall i \in L_3. \quad (87)$$

Now, we add all the constraints in (84) and (85) with each other, which results in

$$\begin{aligned} 2(l+1) &= \sum_{i\in O_{L_1}} \sum_{j\in P_i} \gamma_j + \sum_{i\in O_{L_1}} \gamma'_{\frac{i+1}{2}} \\ &\quad + \sum_{i\in E_{L_2}} \sum_{j\in P_i} \gamma_j + \sum_{i\in E_{L_2}} \gamma''_{\frac{i-2l}{2}} \\ &= \sum_{j\in[n]} \gamma_j o_{L_1,j} + \sum_{i\in[l+1]} \gamma'_j + \sum_{j\in[n]} \gamma_j e_{L_2,j} + \sum_{i\in[l+1]} \gamma''_j \\ &= \sum_{j\in[n]} \gamma_j \Big(e_{L_2,j} + o_{L_1,j} \Big) + \sum_{i\in[l+1]} \Big(\gamma'_j + \gamma''_j \Big). \end{aligned}$$

So, we have

$$\frac{1}{2} \sum_{j\in[n]} \gamma_j \Big(e_{L_2,j} + o_{L_1,j} \Big) = (l+1) - \frac{1}{2} \sum_{i\in[l+1]} \Big(\gamma'_j + \gamma''_j \Big). \quad (88)$$

Now, using (88), for the broadcast rate of the FPCC scheme in (83), we have

$$\begin{aligned} \beta_{\text{FPCC}}(\mathcal{I}_7(l)) &\geq (l+1) + \min_{\substack{\gamma'_j, \gamma''_j, \gamma'''_j \\ j\in[n]}} \Bigg[\sum_{j\in[n]} \gamma_j \\ &\quad + \frac{1}{2} \sum_{i\in[l+1]} \Big(\gamma'_j + \gamma''_j \Big) + \sum_{i\in[l+1]} \gamma'''_j \Bigg] \\ &\geq (l+1) + \frac{1}{2} \min_{\substack{\gamma'_j, \gamma''_j, \gamma'''_j \\ j\in[n]}} \sum_{i\in[l+1]} \Big(\gamma'_j + \gamma''_j + \gamma'''_j \Big) \\ &= (l+1) + \frac{1}{2}(l+1) \quad (89) \\ &= \frac{3}{2}(l+1), \end{aligned}$$

where (89) is due to (87). This completes the proof. □

Now, we provide the following definitions, which will be used in the proof of Lemma 7.

Definition 29 ($V_i^{\text{out}}, V_i^{\text{in}}$: Outgoing and Incoming Neighbor Set of i). In an arbitrary graph $\mathcal{G}_{\mathcal{I}} = ([m], E)$, for any vertex $i \in [m]$, we define an outgoing neighbor set as $V_i^{\text{out}} = \{j \in [m]\backslash\{i\}, (i,j) \in E\}$ and an incoming neighbor set as $V_i^{\text{in}} = \{j \in [m]\backslash\{i\}, (j,i) \in E\}$.

For any vertex i, its neighbor set can be partitioned into three subsets $V_{i,1} = V_i^{\text{out}} \cap V_i^{\text{in}}$, $V_{i,2} = V_i^{\text{out}}\backslash V_{i,1}$, and $V_{i,3} = V_i^{\text{in}}\backslash V_{i,1}$. It is obvious that vertex i forms a pairwise clique with the vertices in $V_{i,1}$.

Definition 30 (Clique-outgoing Vertex). Vertex $i \in [m]$ is said to be a clique-outgoing vertex, if $V_{i,2} = \emptyset$.

Definition 31 (Clique-incoming Vertex). Vertex $i \in [m]$ is said to be a clique-incoming vertex, if $V_{i,3} = \emptyset$.

Lemma 7. *For the class-$\mathcal{I}_7$ index coding instances, we have* $\beta_{ICC}(\mathcal{I}_7(l)) = \beta_{FPCC}(\mathcal{I}_7(l))$.

Proof. We show that for the class-$\mathcal{I}_7$ index coding instances, each ICC-structured subgraph is reduced to a clique, and since the broadcast rate of the ICC and FPCC schemes for a clique is equal, this will complete the proof.

First, it can be verified that (i) the vertices $i \in O_{L_1} \cup E_{L_2}$ are all clique-outgoing vertices and (ii) the vertices $i \in K_3$ are all clique-incoming vertices. Now, suppose that subgraph $M \subseteq [m = 5l + 3]$ is an ICC-structured subgraph, subset $J \subseteq M$ is its inner vertex set and $i \in J$. Now, we note the followings.

- Let vertex i be a clique-outgoing vertex. (i) Since vertex i forms a clique with all of its outgoing neighbors $j_1 \in V_i^{\text{out}}$, to meet the J-cycle condition, we must have $j_1 \notin M\backslash J$. (ii) Since every path from vertex i will include at least one of its outgoing neighbors, then, to meet the inner vertex set definition, none of the vertices in $j_2 \in M\backslash V_i^{\text{out}}$ can be inside the inner vertex set, i.e., $j_2 \notin J$. This means that if a clique-outgoing vertex is inside the vertex set $i \in J$, then only its outgoing neighbors, which forms a pairwise clique with them, are allowed to be inside the inner vertex set J.

- Let vertex i be a clique-incoming vertex. With the same arguments, (i) for all of its incoming neighbors $j_1 \in V_i^{\text{in}}$, we must have $j_1 \notin M\backslash J$ and (ii) only its incoming neighbors, which forms a pairwise clique with them, are allowed to be inside the inner vertex set.

- The vertices inside $E_{L_1} \cup O_{L_2}$ form a clique with each other. Let $i_1, i_2 \in E_{L_1} \cup O_{L_2}$. Now, if $i_1 \in J$, then to meet the J-cycle condition, we must have $i_2 \notin M\backslash J$. Note, if two vertices $i_1 \in E_{L_1} \cup O_{L_2}$ and $i_3 \in O_{L_1} \cup E_{L_2} \cup K_3$ are inside the inner vertex set, then in the previous cases, we already proved that i_1 and i_3 must form a clique with each other.

Thus, this shows that any two vertices $j_1 \in J$ and $j_2 \in J$ must form a clique with each other to meet the definition of the ICC-structure. In other words, we showed that all the vertices inside the inner vertex set must form a pairwise clique with each other. This means that any ICC-structured subgraph will be reduced to a clique. This completes the proof. □