Index Coding and Error Correction

Son Hoang Dau, Vitaly Skachek, and Yeow Meng Chee
Division of Mathematical Sciences, School of Physical and Mathematical Sciences
Nanyang Technological University, 21 Nanyang Link, Singapore 637371
Emails: { DauS0002, Vitaly.Skachek, YMChee } @ntu.edu.sg

***Abstract*—A problem of index coding with side information was first considered by Y. Birk and T. Kol *(IEEE INFOCOM, 1998)*. In the present work, a generalization of index coding scheme, where transmitted symbols are subject to errors, is studied. Error-correcting methods for such a scheme, and their parameters, are investigated. In particular, the following question is discussed: given the side information hypergraph of index coding scheme and the maximal number of erroneous symbols δ, what is the shortest length of a linear index code, such that every receiver is able to recover the required information? This question turns out to be a generalization of the problem of finding a shortest-length error-correcting code with a prescribed error-correcting capability in the classical coding theory.**

The Singleton bound and two other bounds, referred to as the α-bound and the κ-bound, for the optimal length of a linear error-correcting index code (ECIC) are established. For large alphabets, a construction based on concatenation of an optimal index code with an MDS classical code, is shown to attain the Singleton bound. For smaller alphabets, however, this construction may not be optimal. A random construction is also analyzed. It yields another inexplicit bound on the length of an optimal linear ECIC. Finally, the decoding of linear ECIC's is discussed. The syndrome decoding is shown to output the exact message if the weight of the error vector is less or equal to the error-correcting capability of the corresponding ECIC.

I. Introduction

A. Background

The problem of Index Coding with Side Information (ICSI) was introduced by Birk and Kol [1]. During the transmission, each client might miss a certain part of the data, due to intermittent reception, limited storage capacity or any other reasons. Via a slow backward channel, the clients let the server know which messages they already have in their possession, and which messages they are interested to receive. The server has to find a way to deliver to each client all the messages he requested, yet spending a minimum number of transmissions. As it was shown in [1], the server can significantly reduce the number of transmissions by coding the messages.

Possible applications of index coding include communications scenarios, in which a satellite or a server broadcasts a set of messages to a set clients, such as daily newspaper delivery or video-on-demand. Index coding with side information can also be used in opportunistic wireless networks [2].

The ICSI problem has been a subject of several recent studies [3]–[8]. This problem can be viewed as a special case of the Network Coding (NC) problem [9], [10]. In particular, as it was shown in [7], every instance of the NC problem can be reduced to an instance of the ICSI problem.

B. Our contribution

In this work, we generalize the ICSI problem towards a setup with error correction. We extend some known results on index coding to a case where any receiver can correct up to a certain number of errors. The problem of designing such error-correcting index codes (ECIC's) naturally generalizes the problem of constructing classical error-correcting codes. We establish an upper bound (the κ-bound) and a lower bound (the α-bound) on the shortest length of a linear ECIC, which is able to correct any error pattern of size up to δ. We also derive an analog of the Singleton bound, and show that this bound is tight for codes over large alphabets. We also consider random ECIC's. By analyzing their parameters, we obtain an upper bound on their length. Finally, we discuss the decoding of linear ECIC's. We show that the syndrome decoding results in a correct result, provided that the number of errors does not exceed the error-correcting capability of the code.

The problem of error correction for NC was studied in several previous works. However, these results are not directly applicable for the ICSI problem. First, the existing works only consider the multicast scenario, while the ICSI problem, however, is a special case of the non-multicast NC problem. Second, the ICSI problem can be modeled by the NC scenario [8], yet, this requires that there are directed edges from particular sources to each sink, which provide the side information. The symbols transmitted on these special edges, unlike for error-correcting NC, are not allowed to be corrupted.

II. Preliminaries

Let $\mathbb{F}_q$ be the finite field of q elements, where q is a power of prime, and $\mathbb{F}_q^* = \mathbb{F}_q \backslash \{0\}$. Let $[n] = \{1, 2, \ldots, n\}$. For the vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}_q^n$, we use $\mathsf{d}(\boldsymbol{u}, \boldsymbol{v})$ to denote the the Hamming distance between $\boldsymbol{u}$ and $\boldsymbol{v}$. If $\boldsymbol{u} \in \mathbb{F}_q^n$ and $\boldsymbol{M} \subseteq \mathbb{F}_q^n$ is a set of vectors (or a vector subspace), then this notation can be extended to

$$\mathsf{d}(\boldsymbol{u}, \boldsymbol{M}) = \min_{\boldsymbol{v} \in \boldsymbol{M}} \mathsf{d}(\boldsymbol{u}, \boldsymbol{v}) \ .$$

Given q, k, and d, let $N_q[k, d]$ denote the length of the shortest linear code over $\mathbb{F}_q$ which has dimension k and minimum distance d. The *support* of a vector $\boldsymbol{u} \in \mathbb{F}_q^n$ is defined by $\mathsf{supp}(\boldsymbol{u}) \triangleq \{i \in [n] : u_i \neq 0\}$. The Hamming weight of $\boldsymbol{u}$ is defined by $\mathsf{wt}(\boldsymbol{u}) \triangleq |\mathsf{supp}(\boldsymbol{u})|$. Suppose $E \subseteq [n]$. We write $\boldsymbol{u} \lhd E$ whenever $\mathsf{supp}(\boldsymbol{u}) \subseteq E$.

We use $\boldsymbol{e}_i = (\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{n-i}) \in \mathbb{F}_q^n$ to denote the unit vector, which has a one at the ith position, and zeros

elsewhere. For a vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ and a subset $B = \{i_1, i_2, \ldots, i_b\}$ of $[n]$, where $i_1 < i_2 < \cdots < i_b$, let $\boldsymbol{y}_B$ denote the vector $(y_{i_1}, y_{i_2}, \ldots, y_{i_b})$.

For an $n \times N$ matrix $\boldsymbol{L}$, let $\boldsymbol{L}_i$ denote its ith row. For a set $E \subseteq [n]$, let $\boldsymbol{L}_E$ denote the $|E| \times N$ matrix obtained from $\boldsymbol{L}$ by deleting all the rows of $\boldsymbol{L}$ which are not indexed by the elements of E. For a set of vectors $\boldsymbol{M}$, we use notation $\mathsf{span}(\boldsymbol{M})$ to denote the linear space spanned by the vectors in $\boldsymbol{M}$. We also use notation $\mathsf{colspan}(\boldsymbol{L})$ for the linear space spanned by the columns of the matrix $\boldsymbol{L}$.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with a vertex set $\mathcal{V}$ and an edge set $\mathcal{E}$. A directed graph $\mathcal{G}$ is called *symmetric* if

$$(u, v) \in \mathcal{E} \quad \Leftrightarrow \quad (v, u) \in \mathcal{E} \ .$$

The *independence number* of an undirected graph $\mathcal{G}$ is denoted by $\alpha(\mathcal{G})$. There is a natural correspondence between undirected graphs and directed symmetric graphs. By using this correspondence, the definition of independence number is naturally extended to directed symmetric graphs.

III. Error-Correcting Index Coding with Side Information

Index Coding with Side Information problem considers the following communications scenario. There is a unique sender (or source) S, who has a vector of messages $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ in his possession. There are also m receivers $R_1, R_2, \ldots, R_m$, receiving information from S via a broadcast channel. For each $i \in [m]$, R_i has side information, i.e. R_i owns a subset of messages $\{x_j\}_{j \in \mathcal{X}_i}$, where $\mathcal{X}_i \subseteq [n]$. Each R_i, $i \in [m]$, is interested in receiving the message $x_{f(i)}$ (we say that R_i requires $x_{f(i)}$), where the mapping $f : [m] \to [n]$ satisfies $f(i) \notin \mathcal{X}_i$ for all $i \in [m]$. Hereafter, we use the notation $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m)$. An instance of the ICSI problem is given by a quadruple $(m, n, \mathcal{X}, f)$. An instance of the ICSI problem can also be conveniently described by the following directed hypergraph [8].

Definition 3.1: Let $(m, n, \mathcal{X}, f)$ be an instance of the ICSI problem. The corresponding *side information (directed) hypergraph* $\mathcal{H} = \mathcal{H}(m, n, \mathcal{X}, f)$ is defined by the vertex set $\mathcal{V} = [n]$ and the edge set $\mathcal{E}_{\mathcal{H}}$, where

$$\mathcal{E}_{\mathcal{H}} = \{(f(i), \mathcal{X}_i) \ : \ i \in [n]\} \ .$$

We often refer to $(m, n, \mathcal{X}, f)$ as an instance of the ICSI problem described by the hypergraph $\mathcal{H}$.

Each side information hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E}_{\mathcal{H}})$ can be associated with the directed graph $\mathcal{G}_{\mathcal{H}} = (\mathcal{V}, \mathcal{E})$ in the following way. For each directed edge $(f(i), \mathcal{X}_i) \in \mathcal{E}_{\mathcal{H}}$ there will be $|\mathcal{X}_i|$ directed edges $(f(i), v) \in \mathcal{E}$, for $v \in \mathcal{X}_i$. When $m = n$ and $f(i) = i$ for all $i \in [m]$, the graph $\mathcal{G}_{\mathcal{H}}$ is, in fact, the *side information graph*, defined in [3].

Due to noise, the symbols received by R_i, $i \in [m]$, may be subject to errors. Assume that S broadcasts a vector $\boldsymbol{y} \in \mathbb{F}_q^N$. Let $\boldsymbol{\epsilon}_i \in \mathbb{F}_q^N$ be the error affecting the information received by R_i, $i \in [m]$. Then R_i actually receives the vector $\boldsymbol{y}_i = \boldsymbol{y} + \boldsymbol{\epsilon}_i \in \mathbb{F}_q^N$, instead of $\boldsymbol{y}$.

Definition 3.2: Consider an instance of the ICSI problem described by $\mathcal{H} = \mathcal{H}(m, n, \mathcal{X}, f)$. A *$\delta$-error-correcting index code* ($(\delta, \mathcal{H})$-ECIC) over $\mathbb{F}_q$ for this instance is an encoding function

$$\mathfrak{E} \ : \ \mathbb{F}_q^n \to \mathbb{F}_q^N \ ,$$

such that for each receiver R_i, $i \in [m]$, there exists a decoding function

$$\mathfrak{D}_i \ : \ \mathbb{F}_q^N \times \mathbb{F}_q^{|\mathcal{X}_i|} \to \mathbb{F}_q \ ,$$

satisfying

$$\forall \boldsymbol{x}, \boldsymbol{\epsilon}_i \in \mathbb{F}_q^n, \ \mathsf{wt}(\boldsymbol{\epsilon}_i) \leqslant \delta \ : \ \mathfrak{D}_i(\mathfrak{E}(\boldsymbol{x}) + \boldsymbol{\epsilon}_i, \boldsymbol{x}_{\mathcal{X}_i}) = x_{f(i)} \ .$$

If $\delta = 0$, we refer to such $\mathfrak{E}$ as a *non-error-correcting* index code, or just $\mathcal{H}$-IC. The parameter N is called the *length* of the index code. In the scheme corresponding to the code $\mathfrak{E}$, S broadcasts a vector $\mathfrak{E}(\boldsymbol{x})$ of length N over $\mathbb{F}_q$.

Definition 3.3: A *linear index code* is an index code, for which the encoding function $\mathfrak{E}$ is a linear transformation over $\mathbb{F}_q$. Such a code can be described as

$$\forall \boldsymbol{x} \in \mathbb{F}_q^n \ : \ \mathfrak{E}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{L} \ ,$$

where $\boldsymbol{L}$ is an $n \times N$ matrix over $\mathbb{F}_q$. The matrix $\boldsymbol{L}$ is called the *matrix corresponding to the index code* $\mathfrak{E}$, while $\mathfrak{E}$ is referred to as the *linear index code based on* $\boldsymbol{L}$.

Definition 3.4: An *optimal* linear $(\delta, \mathcal{H})$-ECIC over $\mathbb{F}_q$ is a linear $(\delta, \mathcal{H})$-ECIC over $\mathbb{F}_q$ of the smallest possible length $\mathcal{N}_q(\mathcal{H}, \delta)$.

Hereafter, we assume that $\mathcal{X} = (\mathcal{X}_i)_{i \in [m]}$ is known to S. We also assume that the code $\mathfrak{E}$ is known to each receiver R_i, $i \in [m]$.

Definition 3.5: Suppose $\mathcal{H} = \mathcal{H}(m, n, \mathcal{X}, f)$ corresponds to an instance of the ICSI problem. Then the *min-rank* of $\mathcal{H}$ over $\mathbb{F}_q$ is defined as

$$\kappa_q(\mathcal{H}) \triangleq \min\{\mathsf{rank}_{\mathbb{F}_q}(\{\boldsymbol{v}_i + \boldsymbol{e}_{f(i)}\}_{i \in [m]}) \ : \ \boldsymbol{v}_i \in \mathbb{F}_q^n \ , \ \boldsymbol{v}_i \lhd \mathcal{X}_i\} \ .$$

Observe that $\kappa_q(\mathcal{H})$ generalizes the min-rank over $\mathbb{F}_q$ of the side information graph, which was defined in [3]. More specifically, when $m = n$ and $f(i) = i$ for all $i \in [m]$, $\mathcal{G}_{\mathcal{H}}$ becomes the side information graph, and $\kappa_q(\mathcal{H}) = \text{min-rank}_q(\mathcal{G}_{\mathcal{H}})$. The min-rank was shown in [3], [4] to be the smallest number of transmissions in a linear index code.

Lemma 3.1: ([3], [11]) Consider an instance of the ICSI problem described by $\mathcal{H} = \mathcal{H}(m, n, \mathcal{X}, f)$.

1) The matrix $\boldsymbol{L}$ corresponds to a linear $\mathcal{H}$-IC over $\mathbb{F}_q$ if and only if for each $i \in [m]$ there exists $\boldsymbol{v}_i \in \mathbb{F}_q^n$ such that $\boldsymbol{v}_i \lhd \mathcal{X}_i$ and $\boldsymbol{v}_i + \boldsymbol{e}_{f(i)} \in \mathsf{colspan}(\boldsymbol{L})$.
2) The smallest possible length of a linear $\mathcal{H}$-IC over $\mathbb{F}_q$ is $\kappa_q(\mathcal{H})$.

IV. Basic Properties

We define the set of vectors

$$\mathcal{I}(q,\mathcal{H}) \triangleq \left\{ \boldsymbol{z} \in \mathbb{F}_q^n \ : \ \exists i \in [m] \text{ s.t. } \boldsymbol{z}_{\mathcal{X}_i} = \boldsymbol{0},\ z_{f(i)} \neq 0 \right\}.$$

For all $i \in [m]$, we also define $\mathcal{Y}_i \triangleq [n] \backslash \Big(\{f(i)\} \cup \mathcal{X}_i \Big)$. Then the collection of supports of all vectors in $\mathcal{I}(q,\mathcal{H})$ is given by

$$\mathcal{J}(\mathcal{H}) \triangleq \bigcup_{i \in [m]} \Big\{ \{f(i)\} \cup Y_i \ : \ Y_i \subseteq \mathcal{Y}_i \Big\}. \tag{1}$$

Lemma 4.1: The matrix $\boldsymbol{L}$ corresponds to a $(\delta,\mathcal{H})$-ECIC over $\mathbb{F}_q$ if and only if

$$\mathsf{wt}\left(\boldsymbol{z}\boldsymbol{L}\right) \geq 2\delta + 1 \text{ for all } \boldsymbol{z} \in \mathcal{I}(q,\mathcal{H}) \ . \tag{2}$$

Equivalently, $\boldsymbol{L}$ corresponds to a $(\delta,\mathcal{H})$-ECIC over $\mathbb{F}_q$ if and only if

$$\mathsf{wt}\left(\sum_{i \in K} z_i \boldsymbol{L}_i \right) \geq 2\delta + 1, \tag{3}$$

for all $K \in \mathcal{J}(\mathcal{H})$ and for all choices of $z_i \in \mathbb{F}_q^*$, $i \in K$.

Proof: For each $\boldsymbol{x} \in \mathbb{F}_q^n$, we define

$$B(\boldsymbol{x},\delta) = \{ \boldsymbol{y} \in \mathbb{F}_q^N \ : \ \boldsymbol{y} = \boldsymbol{x}\boldsymbol{L} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \in \mathbb{F}_q^N,\ \mathsf{wt}(\boldsymbol{\epsilon}) \leq \delta \} \ ,$$

the set of all vectors resulting from at most δ errors in the transmitted vector associated with the information vector $\boldsymbol{x}$. Then the receiver R_i can recover $x_{f(i)}$ correctly if and only if

$$B(\boldsymbol{x},\delta) \cap B(\boldsymbol{x}',\delta) = \varnothing,$$

for every pair $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{F}_q^n$ satisfying:

$$\boldsymbol{x}_{\mathcal{X}_i} = \boldsymbol{x}'_{\mathcal{X}_i} \text{ and } x_{f(i)} \neq x'_{f(i)} \ .$$

(Observe that R_i is interested only in the bit $x_{f(i)}$, not in the whole vector $\boldsymbol{x}$.)

Therefore, $\boldsymbol{L}$ corresponds to a $(\delta,\mathcal{H})$-ECIC if and only if the following condition is satisfied: for all $i \in [m]$ and for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{F}_q^n$ such that $\boldsymbol{x}_{\mathcal{X}_i} = \boldsymbol{x}'_{\mathcal{X}_i}$ and $x_{f(i)} \neq x'_{f(i)}$, it holds

$$\forall \boldsymbol{\epsilon}, \boldsymbol{\epsilon}' \in \mathbb{F}_q^N,\ \mathsf{wt}(\boldsymbol{\epsilon}) \leqslant \delta,\ \mathsf{wt}(\boldsymbol{\epsilon}') \leqslant \delta \ : \\ \boldsymbol{x}\boldsymbol{L} + \boldsymbol{\epsilon} \neq \boldsymbol{x}'\boldsymbol{L} + \boldsymbol{\epsilon}' \ . \tag{4}$$

Denote $\boldsymbol{z} = \boldsymbol{x}' - \boldsymbol{x}$. Then, the condition in (4) can be reformulated as follows: for all $i \in [n]$ and for all $\boldsymbol{z} \in \mathbb{F}_q^n$ such that $\boldsymbol{z}_{\mathcal{X}_i} = \boldsymbol{0}$ and $z_{f(i)} \neq 0$, it holds

$$\forall \boldsymbol{\epsilon}, \boldsymbol{\epsilon}' \in \mathbb{F}_q^N,\ \mathsf{wt}(\boldsymbol{\epsilon}) \leqslant \delta,\ \mathsf{wt}(\boldsymbol{\epsilon}') \leqslant \delta \ : \ \boldsymbol{z}\boldsymbol{L} \neq \boldsymbol{\epsilon} - \boldsymbol{\epsilon}' \ . \tag{5}$$

The equivalent condition is that for all $\boldsymbol{z} \in \mathcal{I}(q,\mathcal{H})$,

$$\mathsf{wt}(\boldsymbol{z}\boldsymbol{L}) \geqslant 2\delta + 1 \ .$$

Inequality (3) follows from this condition in a straight-forward manner. ∎

Corollary 4.1: For all $i \in [m]$, let

$$\boldsymbol{M}_i \triangleq \text{span}\left(\{ \boldsymbol{L}_j \ : \ j \in \mathcal{Y}_i \} \right) \ .$$

Then, the matrix $\boldsymbol{L}$ corresponds to a $(\delta,\mathcal{H})$-ECIC over $\mathbb{F}_q$ if and only if

$$\forall i \in [m] \ : \ \mathsf{d}(\boldsymbol{L}_{f(i)}, \boldsymbol{M}_i) \geqslant 2\delta + 1 \ . \tag{6}$$

Example 4.1: Let $q = 2$, $m = n = 3$, and $f(i) = i$ for $i \in [3]$. Suppose $\mathcal{X}_1 = \{2,3\}$, $\mathcal{X}_2 = \{1,3\}$, and $\mathcal{X}_3 = \{1,2\}$. Let

$$\boldsymbol{L} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

Note that $\boldsymbol{L}$ generates a $[4,3,1]_2$ code, which has minimum distance one. However, the index code based on $\boldsymbol{L}$ can still correct one error. Indeed, let $\mathcal{H} = \mathcal{H}(3,3,\mathcal{X},f)$, we have

$$\mathcal{I}(2,\mathcal{H}) = \{100, 010, 001\}.$$

Since each row of $\boldsymbol{L}$ has weight at least three, it follows that $\mathsf{wt}(\boldsymbol{z}\boldsymbol{L}) \geq 3$ for all $\boldsymbol{z} \in \mathcal{I}(2,\mathcal{H})$. By Lemma 4.1, $\boldsymbol{L}$ corresponds to a $(1,\mathcal{H})$-ECIC over $\mathbb{F}_2$.

Example 4.2: Assume that $m = n$ and $f(i) = i$ for all $i \in [m]$. Furthermore, suppose that $\mathcal{X}_i = \varnothing$ for all $i \in [m]$ (i.e. there is no side information available to the receivers). Let $\mathcal{H} = \mathcal{H}(m,n,\mathcal{X},f)$. Then, $\mathcal{I}(q,\mathcal{H}) = \mathbb{F}_q^n \backslash \{\boldsymbol{0}\}$. Hence, by Lemma 4.1, the $n \times N$ matrix $\boldsymbol{L}$ corresponding to a $(\delta,\mathcal{H})$-ECIC over $\mathbb{F}_q$ (for some integer $\delta \geqslant 0$) is a generating matrix of an $[N, n, \geqslant 2\delta+1]_q$ linear code. Thus, the problem of designing an ECIC is reduced to the problem of constructing a classical linear error-correcting code.

V. The α-Bound and the κ-Bound

Let $(m,n,\mathcal{X},f)$ be an instance of the ICSI problem, and let $\mathcal{H}$ be the corresponding side information hypergraph. Next, we introduce the following definitions for the hypergraph $\mathcal{H}$.

Definition 5.1: A subset H of $[n]$ is called a *generalized independent set* in $\mathcal{H}$ if every nonempty subset K of H belongs to $\mathcal{J}(\mathcal{H})$.

Definition 5.2: A generalized independent set of the largest size in $\mathcal{H}$ is called a *maximum generalized independent set*. The size of a maximum generalized independent set in $\mathcal{H}$ is called the *generalized independence number*, and denoted by $\alpha(\mathcal{H})$.

When $m = n$ and $f(i) = i$ for all $i \in [n]$, the generalized independence number of $\mathcal{H}$ is equal to the maximum size of an acyclic induced subgraph of $\mathcal{G}_{\mathcal{H}}$, which was introduced in [3]. In particular, when $\mathcal{G}_{\mathcal{H}}$ is symmetric, $\alpha(\mathcal{H})$ is the independence number of $\mathcal{G}_{\mathcal{H}}$. We omit the proof.

Theorem 5.1 (α-bound): The length of an optimal linear $(\delta,\mathcal{H})$-ECIC over $\mathbb{F}_q$ satisfies

$$\mathcal{N}_q(\mathcal{H},\delta) \geqslant N_q[\alpha(\mathcal{H}), 2\delta+1] \ .$$

Proof: Consider an $n \times N$ matrix $\boldsymbol{L}$, which corresponds to a $(\delta, \mathcal{H})$-ECIC. Let $H = \{i_1, i_2, \ldots, i_{\alpha(\mathcal{H})}\}$ be a maximum generalized independent set in $\mathcal{H}$. Then, every subset $K \subseteq H$ satisfies $K \in \mathcal{J}(\mathcal{H})$. Therefore,

$$\mathsf{wt}\left(\sum_{i \in K} z_i \boldsymbol{L}_i\right) \geq 2\delta + 1$$

for all $K \subseteq H$, $K \neq \varnothing$, and for all choices of $z_i \in \mathbb{F}_q^*$, $i \in K$. Hence, the $\alpha(\mathcal{H})$ rows of $\boldsymbol{L}$, namely $\boldsymbol{L}_{i_1}, \boldsymbol{L}_{i_2}, \ldots, \boldsymbol{L}_{i_{\alpha(\mathcal{H})}}$, form a generator matrix of an $[N, \alpha(\mathcal{H}), 2\delta + 1]_q$ code. Therefore,

$$N \geq N_q[\alpha(\mathcal{H}), 2\delta + 1] \; .$$

■

The following proposition is based on the fact that concatenation of a δ-error-correcting code with an optimal (non-error-correcting) $\mathcal{H}$-IC yields a $(\delta, \mathcal{H})$-ECIC.

Proposition 5.2 (κ-bound): The length of an optimal $(\delta, \mathcal{H})$-ECIC over $\mathbb{F}_q$ satisfies

$$\mathcal{N}_q(\mathcal{H}, \delta) \leq N_q[\kappa_q(\mathcal{H}), 2\delta + 1] \; .$$

The proof of this proposition is omitted due to lack of space.

Corollary 5.1: The length of an optimal linear $(\delta, \mathcal{H})$-ECIC over $\mathbb{F}_q$ satisfies

$$N_q[\alpha(\mathcal{H}), 2\delta + 1] \leq \mathcal{N}_q(\mathcal{H}, \delta) \leq N_q[\kappa_q(\mathcal{H}), 2\delta + 1] \; .$$

Example 5.1: Let $q = 2$, $m = n = 5$, $\delta = 2$, and $f(i) = i$ for all $i \in [m]$. Assume

$$\mathcal{X}_1 = \{2, 5\} \; , \quad \mathcal{X}_2 = \{1, 3\} \; , \quad \mathcal{X}_3 = \{2, 4\} \; ,$$
$$\mathcal{X}_4 = \{3, 5\} \; , \quad \mathcal{X}_5 = \{1, 4\} \; .$$

Let $\mathcal{H} = \mathcal{H}(5, 5, \mathcal{X}, f)$. The side information graph $\mathcal{G}_{\mathcal{H}}$ of this instance is a pentagon. It is easy to verify that $\alpha(\mathcal{H}) = \alpha(\mathcal{G}) = 2$. It follows from Theorem 9 in [4] that $\kappa_2(\mathcal{H}) =$ min-rank$_2(\mathcal{G}_{\mathcal{H}}) = 3$. Thus, from [12] we have

$$N_2[2, 5] = 8 \quad \text{and} \quad N_2[3, 5] = 10 \; .$$

Due to Corollary 5.1, we have

$$8 \leq \mathcal{N}_2(\mathcal{H}, 2) \leq 10.$$

Using a computer search, we obtain that $\mathcal{N}_2(\mathcal{H}, 2) = 9$, and the corresponding optimal scheme is based on

$$\boldsymbol{L} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \; .$$

It is technical to verify that by Lemma 4.1, $\boldsymbol{L}$ corresponds to $(2, \mathcal{H})$-ECIC. The length of this ECIC lies strictly between the α-bound and the κ-bound.

Remark 5.1: Example 5.1 illustrates that over small alphabets, the concatenation of an optimal linear (non-error-correcting) index code and an optimal linear error-correcting code may fail to produce an optimal linear ECIC.

VI. The Singleton Bound

Theorem 6.1 (Singleton bound): The length of an optimal linear $(\delta, \mathcal{H})$-ECIC over $\mathbb{F}_q$ satisfies

$$\mathcal{N}_q(\mathcal{H}, \delta) \geq \kappa_q(\mathcal{H}) + 2\delta \; .$$

Proof: Let $\boldsymbol{L}$ be the $n \times \mathcal{N}_q(\mathcal{H}, \delta)$ matrix corresponding to some optimal $(\delta, \mathcal{H})$-ECIC. Let $\boldsymbol{L}'$ be the matrix obtained by deleting any 2δ columns from $\boldsymbol{L}$.

By Lemma 4.1, $\boldsymbol{L}$ satisfies for all $\boldsymbol{z} \in \mathcal{I}(q, \mathcal{H})$,

$$\mathsf{wt}(\boldsymbol{z}\boldsymbol{L}) \geqslant 2\delta + 1 \; .$$

We deduce that the rows of $\boldsymbol{L}'$ also satisfy that for all $\boldsymbol{z} \in \mathcal{I}(q, \mathcal{H})$,

$$\mathsf{wt}(\boldsymbol{z}\boldsymbol{L}') \geqslant 1 \; .$$

By Lemma 4.1, $\boldsymbol{L}'$ corresponds to a linear $\mathcal{H}$-IC. Therefore, by Lemma 3.1, part 2, $\boldsymbol{L}'$ has at least $\kappa_q(\mathcal{H})$ columns. We deduce that

$$\mathcal{N}_q(\mathcal{H}, \delta) - 2\delta \geq \kappa_q(\mathcal{H}) \; ,$$

which concludes the proof. ■

The corollary below shows that for sufficiently large alphabets, a concatenation of a classical MDS error-correcting code with an optimal (non-error-correcting) index code yields an optimal linear ECIC.

Corollary 6.1 (MDS error-correcting index code): For $q \geq \kappa_q(\mathcal{H}) + 2\delta - 1$,

$$\mathcal{N}_q(\mathcal{H}, \delta) = \kappa_q(\mathcal{H}) + 2\delta \; . \tag{7}$$

Proof: Follows from Theorem 6.1 and Proposition 5.2. ■

Remark 6.1: There exist hypergraph $\mathcal{H}$, such that $\mathcal{G}_{\mathcal{H}}$ is the (symmetric) odd cycle of length n, for which the α-bound is at least as good as the Singleton bound.

VII. Random codes

Theorem 7.1: Let $\mathcal{H} = \mathcal{H}(m, n, \mathcal{X}, f)$ describe an instance of the ICSI problem. Then there exists a $(\delta, \mathcal{H})$-ECIC over $\mathbb{F}_q$ of length N if

$$\sum_{i \in [m]} q^{n - |\mathcal{X}_i| - 1} < \frac{q^N}{V_q(N, 2\delta)} \; , \tag{8}$$

where

$$V_q(N, 2\delta) = \sum_{\ell=0}^{2\delta} \binom{N}{\ell} (q-1)^\ell$$

is the volume of the q-ary sphere in $\mathbb{F}_q^N$.

Idea of proof: We construct a random $n \times N$ matrix $\boldsymbol{L}$ over $\mathbb{F}_q$, row by row. Each row is selected independently of other rows, uniformly over $\mathbb{F}_q^N$. The result is obtained by bounding from above the probability of the event

$$\bigcup_{i \in [m]} E_i \; , \quad \text{where } E_i \triangleq \left\{\mathsf{d}(\boldsymbol{L}_{f(i)}, \boldsymbol{M}_i) < 2\delta + 1\right\} \; ,$$

and by making this probability less than 1.

Remark 7.1: The bound in Theorem 7.1 implies a bound on $\kappa_q(\mathcal{H})$, which is tight for some $\mathcal{H}$. Indeed, fix $\delta = 0$. Take $m = n = 2\ell + 1$ ($\ell \geq 2$), and $f(i) = i$ for all $i \in [n]$. Let $\mathcal{X}_1 = [n]\backslash\{1, 2, n\}$ and $\mathcal{X}_n = [n]\backslash\{1, n-1, n\}$. For $2 \leq i \leq n-1$, let $\mathcal{X}_i = [n]\backslash\{i-1, i, i+1\}$. Take $\mathcal{H} = \mathcal{H}(n, n, \mathcal{X}, f)$. Then $\mathcal{G}_{\mathcal{H}}$ is the complement of the (symmetric directed) odd cycle of length n. We have $|\mathcal{X}_i| = 2\ell - 2$ for all $i \in [n]$. Then (8) becomes

$$N > 2 + \log_q(2\ell + 1) \ .$$

If $q > 2\ell + 1$ then we obtain $N \geqslant 3$. Observe that in this case $\kappa_q(\mathcal{H}) = \text{min-rank}_q(\mathcal{G}_{\mathcal{H}}) = 3$ (see [8, Claim A.1]), and thus the bound is tight.

VIII. Syndrome decoding

Consider the $(\delta, \mathcal{H})$-ECIC based on a matrix $\boldsymbol{L}$. Suppose that the receiver R_i, $i \in [m]$, receives the vector

$$\boldsymbol{y}_i = \boldsymbol{x}\boldsymbol{L} + \boldsymbol{\epsilon}_i \ , \tag{9}$$

where $\boldsymbol{x}\boldsymbol{L}$ is the codeword transmitted by S, and $\boldsymbol{\epsilon}_i$ is the error pattern affecting this codeword.

In the classical coding theory, the transmitted vector $\boldsymbol{c}$, the received vector $\boldsymbol{y}$, and the error pattern $\boldsymbol{e}$ are related by $\boldsymbol{y} = \boldsymbol{c} + \boldsymbol{e}$. For index coding, however, this is no longer the case. The following theorem shows that, in order to recover the message $x_{f(i)}$ from $\boldsymbol{y}_i$ using (9), it is sufficient to find just one vector from a set of possible error patterns. This set is defined as follows:

$$\mathcal{L}_i(\boldsymbol{\epsilon}_i) = \{\boldsymbol{\epsilon}_i + \boldsymbol{z} \ : \ \boldsymbol{z} \in \mathsf{span}(\{\boldsymbol{L}_j\}_{j \in \mathcal{Y}_i})\} \ .$$

We henceforth refer to the set $\mathcal{L}_i(\boldsymbol{\epsilon}_i)$ as the *set of relevant error patterns*.

Lemma 8.1: Assume that the receiver R_i receives $\boldsymbol{y}_i$.

1) If R_i knows the message $x_{f(i)}$ then it is able to determine the set $\mathcal{L}_i(\boldsymbol{\epsilon}_i)$.
2) If R_i knows some vector $\hat{\boldsymbol{\epsilon}} \in \mathcal{L}_i(\boldsymbol{\epsilon}_i)$ then it is able to determine $x_{f(i)}$.

We now describe a syndrome decoding algorithm for linear error-correcting index codes. We have

$$\boldsymbol{y}_i - \boldsymbol{x}_{\mathcal{X}_i}\boldsymbol{L}_{\mathcal{X}_i} - \boldsymbol{\epsilon}_i \in \mathsf{span}\big(\{\boldsymbol{L}_{f(i)}\} \cup \{\boldsymbol{L}_j\}_{j \in \mathcal{Y}_i}\big) \ .$$

Let $\mathscr{C}_i = \mathsf{span}(\{\boldsymbol{L}_{f(i)}\} \cup \{\boldsymbol{L}_j\}_{j \in \mathcal{Y}_i})$, and let $\boldsymbol{H}^{(i)}$ be a parity check matrix of $\mathscr{C}_i$. We obtain that

$$\boldsymbol{H}^{(i)}\boldsymbol{\epsilon}_i^T = \boldsymbol{H}^{(i)}(\boldsymbol{y}_i - \boldsymbol{x}_{\mathcal{X}_i}\boldsymbol{L}_{\mathcal{X}_i})^T \ .$$

Let $\boldsymbol{\beta}_i$ be a column vector defined by

$$\boldsymbol{\beta}_i = \boldsymbol{H}^{(i)}(\boldsymbol{y}_i - \boldsymbol{x}_{\mathcal{X}_i}\boldsymbol{L}_{\mathcal{X}_i})^T \ .$$

Observe that each R_i is capable of determining $\boldsymbol{\beta}_i$. This leads us to the formulation of the decoding procedure for R_i in Figure 1.

Theorem 8.2: Let $\boldsymbol{y}_i = \boldsymbol{x}\boldsymbol{L} + \boldsymbol{\epsilon}_i$ be the vector received by R_i, and let $\mathsf{wt}(\boldsymbol{\epsilon}_i) \leqslant \delta$. Assume that the procedure in Figure 1 is applied to $(\boldsymbol{y}_i, \boldsymbol{x}_{\mathcal{X}_i}, \boldsymbol{L})$. Then, its output satisfies $\hat{x}_{f(i)} = x_{f(i)}$.

- *Input:* $\boldsymbol{y}_i$, $\boldsymbol{x}_{\mathcal{X}_i}$, $\boldsymbol{L}$.
- *Step 1*: Compute the syndrome
$$\boldsymbol{\beta}_i = \boldsymbol{H}^{(i)}(\boldsymbol{y}_i - \boldsymbol{x}_{\mathcal{X}_i}\boldsymbol{L}_{\mathcal{X}_i})^T \ .$$
- *Step 2*: Find the lowest Hamming weight solution $\hat{\boldsymbol{\epsilon}}$ of the system
$$\boldsymbol{H}^{(i)}\hat{\boldsymbol{\epsilon}}^T = \boldsymbol{\beta}_i \ .$$
- *Step 3*: Given that $\hat{\boldsymbol{x}}_{\mathcal{X}_i} = \boldsymbol{x}_{\mathcal{X}_i}$, solve the system for $\hat{x}_{f(i)}$:
$$\boldsymbol{y}_i = \hat{\boldsymbol{x}}\boldsymbol{L} + \hat{\boldsymbol{\epsilon}}.$$
- *Output:* $\hat{x}_{f(i)}$.

Fig. 1: Syndrome decoding procedure.

Remark 8.1: It is not impossible that $\hat{\boldsymbol{\epsilon}} \neq \boldsymbol{\epsilon}_i$. However, if $\mathsf{wt}(\boldsymbol{\epsilon}_i) \leq \delta$, it can be shown that $\hat{\boldsymbol{\epsilon}} \in \mathcal{L}_i(\boldsymbol{\epsilon}_i)$. Hence, by Lemma 8.1, we have $\hat{x}_{f(i)} = x_{f(i)}$.

IX. Acknowledgements

The authors would like to thank the authors of [4] for providing a preprint of their paper. This work is supported by the National Research Foundation of Singapore (Research Grant NRF-CRP2-2007-03).

References

[1] Y. Birk and T. Kol, "Informed-source coding-on-demand (ISCOD) over broadcast channels," in *Proc. IEEE Conf. on Comput. Commun. (INFOCOM)*, San Francisco, CA, 1998, pp. 1257–1264.
[2] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," in *Proc. ACM SIGCOMM*, 2006, pp. 243–254.
[3] Z. Bar-Yossef, Z. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," in *Proc. 47th Annu. IEEE Symp. on Found. of Comput. Sci. (FOCS)*, 2006, pp. 197–206.
[4] ——, "Index coding with side information," *IEEE Trans. Inform. Theory*, to appear.
[5] E. Lubetzky and U. Stav, "Non-linear index coding outperforming the linear optimum," *Proc. 48th Annu. IEEE Symp. on Found. of Comput. Sci. (FOCS)*, pp. 161–168, 2007.
[6] S. El Rouayheb, M. A. R. Chaudhry, and A. Sprintson, "On the minimum number of transmissions in single-hop wireless coding networks," in *Proc. IEEE Inform. Theory Workshop (ITW)*, 2007, pp. 120–125.
[7] S. El Rouayheb, A. Sprintson, and C. Georghiades, "On the relation between the index coding and the network coding problems," in *Proc. IEEE Symp. on Inform. Theory (ISIT)*, Toronto, Canada, 2008, pp. 1823–1827.
[8] N. Alon, A. Hassidim, E. Lubetzky, U. Stav, and A. Weinstein, "Broadcasting with side information," in *Proc. 49th Annu. IEEE Symp. on Found. of Comput. Sci. (FOCS)*, 2008, pp. 823–832.
[9] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1204–1216, 2000.
[10] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, pp. 782–795, 2003.
[11] S. H. Dau, V. Skachek, and Y. M. Chee, "Secure index coding with side information," available online at `http://arxiv.org/abs/1011.5566`.
[12] M. Grassl, "Bounds on the minimum distance of linear codes and quantum codes," available online at `http://www.codetables.de`.