Linear Codes for Broadcasting with Noisy Side Information: Bounds and Code Constructions

Suman Ghosh and Lakshmi Natarajan

Abstract—We consider the problem of communicating over noise free broadcast channels where each receiver possesses an erroneous version of the message symbols that it demands from the transmitter as side information, and the number of errors in this side information is upper bounded by a constant. This communication problem, which we refer to as broadcasting with noisy side information (BNSI), has applications in the retransmission phase of downlink networks, and to the best of our knowledge, has no known coding schemes available in the literature. In a BNSI network the transmitter can exploit the noisy side information at the receivers to reduce the number of uses of the broadcast channel. In this paper, using a known code design criterion, we analyze and construct linear coding schemes for BNSI networks. Using a representation of BNSI problems in terms of undirected bipartite graphs, we first derive lower bounds on the optimal codeword length of linear codes for these problems. We then utilize the parity-check matrices of appropriately chosen linear error correcting codes to construct valid encoder matrices for BNSI problems. We further optimize this technique by partitioning a BNSI problem into multiple subproblems and applying independent linear encoders for each of these subproblems. Finally, we show that BNSI problems form a strict subset of index coding problems by proving that any given linear BNSI problem is equivalent to a scalar linear index coding problem, albeit with a considerably larger number of receivers than the given BNSI problem.

I. INTRODUCTION

Broadcast channels, where a single transmitter simultaneously serves multiple receivers using a common communication medium, are encountered in several applications, such as content delivery systems, sensor networks and satellite communications. We study the family of broadcast channels where the transmitter is equipped with finitely many messages, each receiver demands a subset of these messages and has prior knowledge of a noisy version of the message symbols it demands. We assume that the number of symbol errors in the noisy side information at each receiver is at the most a known constant, and that the broadcast channel itself is error free. The objective of this network coding problem, which we refer to as broadcasting with noisy side information (BNSI), is to design a coding scheme with as small a codeword length as possible such that each receiver can decode its demanded messages from the broadcast codeword and its own noisy side information.

An application of the BNSI problem is in the retransmission phase of multiuser downlink channels. If the channel is in outage during the initial transmission, the decoded packets at the receivers might contain symbol errors. When a retransmission is requested, the transmitter can encode the information symbols to achieve bandwidth savings by exploiting the side information at each receiver, which is an erroneous version of each receiver’s own demanded messages available from the earlier failed transmission.

To the best of our knowledge, no prior work on designing coding schemes for the broadcast problem considered in this paper seems to be available in the literature. However, the code design problem for BNSI networks, as we show in Section V of this paper, is related to index coding [1], [2]. In the index coding problem, each receiver of a broadcast channel demands a subset of the finitely many messages available at the transmitter and is aided with the prior knowledge of some other subset of these messages. The index coding problem is well known and several code constructions and bounds on optimal codewidth are available, such as [1]–[9]. Index coding when the side information is noisy [10], [11] and when the broadcast channel is noisy [12], [13] have also been studied. Note that unlike index coding, in the BNSI problems considered in this paper the side information at each receiver corresponds to only those messages that it demands from the transmitter.

In our recent work [14], we had formulated a necessary and sufficient condition for a linear coding scheme to be a valid solution to a given BNSI problem, and identified the family of BNSI problems where linear coding, if carefully designed, can provide strict gains over uncoded transmission.

In this paper we design linear coding schemes and derive bounds on the optimal length of linear codes for BNSI problems. We utilize a representation of BNSI problems as undirected bipartite graphs for this purpose. We first derive lower bounds on the optimal linear codeword length based on the graph representation (Section III). Using parity-check matrices of appropriately chosen error correcting codes we provide constructions of linear coding schemes (Section IV-A), and then optimize this technique by partitioning a given BNSI problem into multiple subproblems and applying a separate code for each of these subproblems (Section IV-B). These code constructions also provide upper bounds on the optimal codewidth. We then show that for any given BNSI problem there exists an index coding problem, albeit with a larger number of receivers than the BNSI problem, such that a linear...
code is a solution for the BNSI problem if and only if it is a scalar linear solution for the corresponding index coding problem (Section V). The system model and the relevant background are described in Section II.

A combined and extended version of this paper and [14], including the proofs of all the theorems and lemmas is available in [15].

Notation: Matrices and row vectors are denoted by bold uppercase and lowercase letters, respectively. For any positive integer \( n \), the symbol \([n]\) is the set \( \{1, \ldots, n\} \). The symbol \( wt(x) \) is the Hamming weight of the vector \( x \), and \( F_q \) is the finite field of size \( q \). We represent the set \( \{ \text{number of errors in each side information vector is less than } s \} \) as \( E(s) \). For any matrix \( L \in F_q^{n \times N} \), \( \text{rowspan}\{L\} \) denotes the rowspace of \( L \), and \( L^T \) is the transpose of \( L \). For any \( A \subset [n] \), \( L(A) \) is the submatrix of \( L \) composed of the rows indexed by \( A \). For \( x \in F_q^n \), \( x_A \) denotes the vector \( (x_i, i \in A) \).

II. SYSTEM MODEL AND PRELIMINARIES

We consider a noiseless broadcast channel with a single transmitter and \( m \) receivers \( u_1, \ldots, u_m \). There are \( n \) messages available at the transmitter over the finite field \( F_q \), denoted as the vector \( x = (x_1, \ldots, x_n) \in F_q^n \). Consider \( m \) subsets \( X_1, \ldots, X_m \) \( \subseteq [n] \), such that the messages demanded by user \( u_i \) is \( x_{X_i} \), for \( i \in [m] \). We represent the demands of all the users by the tuple \( X = (X_1, \ldots, X_m) \). For each \( i \in [m] \), the \( i \)-th receiver possesses a noisy version of its demanded messages, i.e., the side information at \( u_i \) is a vector \( x_{X_i} + \epsilon_i \), where \( \epsilon_i \in F_q^{n \times 1} \) is the error vector. Neither the transmitter nor the receivers know the exact realization of the side information error vectors \( \epsilon_1, \ldots, \epsilon_m \). However, we will assume that the number of errors in each side information vector is less than or equal to a constant, say \( \delta \), i.e., \( wt(\epsilon_i) \leq \delta \) for all \( i \in [m] \).

The \((m, n, X, \delta)\) broadcasting with noisy side information (BNSI) problem is to design an encoder \( \mathcal{E} : F_q^n \rightarrow \text{rowspan}\{L\} \) which maps the message vector \( x \) to a codeword \( c = \mathcal{E}(x) \) of as small a code length \( N \) as possible, such that every user \( u_i \) can decode its demanded messages \( x_{X_i} \) from the broadcast codeword \( c \) and its own noisy side information \( x_{X_i} + \epsilon_i \). Thus, \( \mathcal{E} \) is a valid encoding function for the \((m, n, X, \delta)\) BNSI problem if and only if there exist decoding functions \( \mathcal{D}_1, \ldots, \mathcal{D}_m \) such that for each \( i \in [m] \) we have \( \mathcal{D}_i(\mathcal{E}(x), x_{X_i} + \epsilon_i) = x_{X_i} \) for all \( x \in F_q^n \) and \( \epsilon_i \in F_q^{n \times 1} \) with \( wt(\epsilon_i) \leq \delta \). Note that the \( i \)-th decoding function is a map \( \mathcal{D}_i : F_q^n \times \text{rowspan}\{L\} \rightarrow \text{rowspan}\{L\} \).

In this paper we will assume that the encoding function \( \mathcal{E} \) is linear, i.e., \( \mathcal{E} = xL \), where \( L \in F_q^{n \times N} \) is the encoder matrix. We will denote by \( N_{q,\text{opt}}(m, n, X, \delta) \), or simply \( N_{q,\text{opt}} \), the minimum code length \( N \) among all valid linear encoding schemes for the \((m, n, X, \delta)\) BNSI problem. Note that encoded transmission \( c = xL \), i.e., choosing the encoding matrix \( L \) to be the identity matrix \( I_n \) is a valid linear code for any BNSI problem. Hence, we have the trivial bound \( N_{q,\text{opt}}(m, n, X, \delta) \leq n \).

In the rest of this section we will briefly recall the relevant results from our recent work [14].

Given an \((m, n, X, \delta)\) BNSI problem and finite field \( F_q \), the set \( \mathcal{I}(q, m, n, X, \delta) \) of vectors of length \( n \) is defined as

\[
\mathcal{I}(q, m, n, X, \delta) = \bigcup_{i=1}^{m}\{z \in F_q^n \mid wt(z_{X_i}) \in [2\delta] \}.
\]  

(1)

That is, a vector \( z \) belongs to \( \mathcal{I}(q, m, n, X, \delta) \), or simply \( I \), if and only if for some \( i \in [m] \), \( 1 \leq wt(z_{X_i}) \leq 2\delta \).

**Theorem 1** ([14]). For any \( N \geq 1 \), a matrix \( L \in F_q^{n \times N} \) is a valid encoder matrix for the \((m, n, X, \delta)\) BNSI problem if and only if \( zL \neq 0 \), \( \forall z \in \mathcal{I}(q, m, n, X, \delta) \).

Theorem 1 shows that \( L \) is a valid encoder matrix if and only if the rows of \( L \) satisfy certain linear independence properties determined by the set \( I \). The following corollary is a restatement of Theorem 1 which emphasizes this viewpoint. For any \( i \in [m] \), we define \( \mathcal{I}_i = \{z_i \mid x_i \in \mathcal{X} \} \), i.e., \( \mathcal{I}_i \) is the index set of all the messages that are not demanded by receiver \( u_i \).

**Corollary 1** ([14]). A matrix \( L \) is a valid encoder for the \((m, n, X, \delta)\)-BNSI problem if and only if for every \( i \in [m] \), any non-zero linear combination of any \( 2\delta \) or fewer rows of \( L_{X_i} \) does not belong to \( \text{rowspan}\{L_{X_i}\} \).

The undirected bipartite graph corresponding to the \((m, n, X, \delta)\) BNSI problem is defined as \( B = (U, P, E) \) which consists of the node-sets \( U = \{u_1, u_2, \ldots, u_m\} \) and \( P = \{x_1, x_2, \ldots, x_n\} \), and the set of undirected edges \( E = \{(u_i, x_j) \mid i \in [m] \text{ and } j \in \mathcal{X} \} \). The set \( U \) is the user-set, \( P \) is the information symbol-set, and the edges \( E \) represent the demands of the users. Given a bipartite graph \( B \) corresponding to an \((m, n, X, \delta)\) BNSI problem, let \( \Phi(B) \) be the collection of all non-empty \( C \subseteq [n] \) such that for every \( i \in [m] \), either \( |X_i \cap C| = 0 \) or \( |X_i \cap C| \geq 2\delta + 1 \), i.e.,

\[
\Phi(B) = \{C \subseteq [n] \mid |X_i \cap C| \notin [2\delta], \forall i \in [m] \}.
\]

In other words, \( C \in \Phi(B) \) if and only if in the subgraph of \( B \) induced by the information symbols \( x_C = \{x_j \mid j \in C\} \) all the user-nodes have degree at least \( 2\delta + 1 \).

**Theorem 2** ([14]). For a BNSI problem represented by the bipartite graph \( B \), \( N_{q,\text{opt}} = n \) if and only if \( \Phi(B) = \phi \).

We observe that linear coding can perform strictly better than uncoded transmission if and only if \( \Phi(B) \) is non-empty. If \( \Phi(B) \) is empty uncoded transmission is optimal and \( N_{q,\text{opt}} = n \). Also, there is a simple algorithm to determine if \( \Phi(B) \) is empty for a given BNSI problem [14].

III. LOWER BOUNDS ON OPTIMAL CODELENGTH

In this section we will describe two lower bounds on \( N_{q,\text{opt}} \). First, we provide a result that will help to obtain these lower bounds on the optimal codelength.

A. Subproblems of a BNSI problem

Consider a bipartite graph \( B = (U, P, E) \) representing an \((m, n, X, \delta)\) BNSI problem. For any subset \( \rho \subseteq [m] \), let \( x_\rho = \{x_j \mid j \in \rho\} \) and \( B' = (U' \cup P', E') \) be the subgraph of \( B \) induced by the information symbol subset \( P' = x_\rho \), i.e., \( U' = \{u_i \in U \mid u_i \cap X_i \neq \phi \} \) and \( E' = \{u_i, x_j \} \in E \mid x_j \in \mathcal{X} \} \). Then \( B' \) represents the \((m', n', X', \delta)\) BNSI
problem where \( m' = |U| \), \( n' = |P'| \) and \( X' \) is the tuple \((X'_2 = X_1 \cap P, \forall u_i \in U')\). The \((m', n', X', \delta)\) BNSI problem is derived from \((m, n, X, \delta)\) by removing the information symbols \( x[i]|_{\delta} \) and corresponding demands at the receivers.

**Lemma 1.** For the subproblem \((m', n', X', \delta)\) of \((m, n, X, \delta)\), \( N_{q, opt}(m', n', X', \delta) \leq N_{q, opt}(m, n, X, \delta) \).

**Proof sketch:** We use Corollary 1 to show that if \( L \) is an optimal encoder matrix for \((m, n, X, \delta)\), then its submatrix \( L_p \) is a valid encoder matrix for \((m', n', X', \delta)\).  

### B. Lower Bounds on the Optimal Codelength

For an \((m, n, X, \delta)\) problem let \( S = \{ i \in [m] \mid |X_i| \in \{2\delta_i\}\} \) and \( X_S = \bigcup_{i \in S} X_i \), i.e., \( S \) represents the set of all users each of whose demand size is at the most \( 2\delta_i \), and \( X_S \) represents the union of the demands of all such receivers.

**Theorem 3.** The optimal codelength \( N_{q, opt} \) over any finite field \( \mathbb{F}_q \) for the \((m, n, X, \delta)\) BNSI problem satisfies the lower bound \( N_{q, opt} \geq |X_S| + \min(2\delta_n, n - |X_S|) \).

**Proof sketch:** We let \( B \) be the union of \( X_S \) and any \( \min(2\delta_n, n - |X_S|) \) elements of \([n] \setminus X_S\). Using the property that for each \( j \in X_S \) there exists an \( i \in [m] \) such that \( |X_i| \leq 2\delta_i \), we show that for the subgraph \( B' \) induced by \( x_B \), \( \Phi(B') = \phi \) if \( B' \) is the subgraph of \( B \) induced by \( x_B \).

**Theorem 4.** The optimal codelength \( N_{q, opt} \) of the \((m, n, X, \delta)\) BNSI problem is at least \( |B_{max}| \).

**Proof sketch:** Similar to the proof of Theorem 3, this claim follows immediately from \( N_{q, opt}(B_{max}) = |B_{max}| \) (Theorem 2) and \( N_{q, opt}(B) \geq N_{q, opt}(B_{max}) \) (Lemma 1). The proof of Theorem 3 relies on the property that if \( B \) is the union of \( X_S \) and any other \( \min(2\delta_n, n - |X_S|) \) elements of \([n]\), then \( \Phi(B_{max}) = \phi \), and hence, \( N_{q, opt}(B) \geq |B| = |X_S| + \min(2\delta_n, n - |X_S|) \). Since \( B_{max} \) is a largest possible subset of \([n]\) satisfying \( \Phi(B_{max}) = \phi \), we deduce that \( |B_{max}| \geq |B| = |X_S| + \min(2\delta_n, n - |X_S|) \). Thus the lower bound of Theorem 4 is at least as good as Theorem 3. While the bound in Theorem 3 can be calculated easily, we do not yet know of an efficient method to compute \( |B_{max}| \) for a general BNSI problem.

**Example 1.** Consider a BNSI problem with \( m = 3 \) receivers, \( n = 4 \) message symbols, \( \delta_1 = 1 \), and message demands \( X_1 = \{1, 2, 3\} \), \( X_2 = \{2, 3, 4\} \), \( X_3 = \{1, 3, 4\} \). For this problem, \( |X_3| = 0 \), hence from Theorem 3 we have \( N_{q, opt} \geq \min(2\delta_n, n) = 2 \). Now it is easy to verify that any subset of \( \{1, 2, 3\} \) of size 3, i.e., \( \{1, 2, 3\} \), \( \{1, 3, 4\} \) or \( \{2, 3, 4\} \), serves as \( B_{max} \) for this problem. So, from Theorem 4, \( N_{q, opt} \geq |B_{max}| = 3 \). In fact, this last bound is tight for this BNSI problem and \( N_{q, opt} = 3 \), see Example 2.

### IV. Constructions of Linear Coding Schemes and Upper Bounds on Optimal Codelength

#### A. Construction using error correcting codes

Consider a parity check matrix \( H \in \mathbb{F}_q^{(n' - k') \times n'} \) of an \([n', k']\) linear error correcting code over \( \mathbb{F}_q \) where \( n' \) and \( k' \) denote the blocklength and the dimension of the code, respectively. Let \( d_{min} \) be the minimum distance of the code. Then any set of \( (d_{min} - 1) \) columns of \( H \) are linearly independent. For a given \((m, n, X, \delta)\) BNSI problem, define \( \eta = 2\delta + \max_{i \in [m]} |Y_i| \), where \( Y_i \) is the index set of the messages that are not demanded by \( i \)-th user. Now if \( d_{min} \geq \eta + 1 \), then we observe that any set of \( \eta \) or fewer rows of \( L \) are linearly independent. Note that for any \( i \in [m] \), all the rows of \( L X_i \) together with any set of \( 2\delta \) rows of \( L X_i \) constitute a set of at the most \( \eta \) rows of \( L \). Thus, by construction, these rows are linearly independent. Hence, from Corollary 1 we conclude that this choice of \( B \) is a valid encoder matrix with codelength \( N = n - k' \). Let \( k(q, n, \eta + 1) \) be the largest possible dimension among all linear error correcting codes over \( \mathbb{F}_q \) with blocklength \( n \) and minimum distance at least \( \eta + 1 \). Then from our choice of \( B \) as the transpose of the parity-check matrix of the \([n, k(q, n, \eta + 1)]\) code with minimum distance at least \( \eta + 1 \), we conclude that \( N_{q, opt} \leq n - k(q, n, \eta + 1) \).

**Example 2.** If \( |X_i| \geq 2\delta_i + 1 \) for all \( i \in [m] \), i.e., if \( n \in \Phi(B) \), then \( \eta \leq (n - 1) \). The \([n, 1]\) repetition code over \( \mathbb{F}_q \) satisfies \( d_{min} = n \geq \eta + 1 \). The parity-check matrix of this code is a valid encoder with codelength \( N = n - 1 \), and we conclude \( N_{q, opt} \leq n - 1 \). For instance, for the BNSI problem of Example 1, we have \( N_{q, opt} \leq 3 \).

If a maximum distance separable (MDS) code of length \( n \) and minimum distance \( \eta + 1 \) exists over \( \mathbb{F}_q \), such as when \( q \geq n \), then the dimension of this code \( k' = (n - \eta + 1) \) such as when \( q \geq n \). In terms of the sizes of the demanded message sets, we have \( k' = \min_{i \in [m]} \{|X_i| - 2\delta_i| \} \). Using the transpose of the parity-check matrix of this MDS code as the encoder matrix \( L \), we arrive at

**Theorem 5.** If \( \min_{i \in [m]} \{|X_i| - 2\delta_i| \} \) MDS code exists over \( \mathbb{F}_q \) then \( N_{q, opt}(m, n, X, \delta) \leq n - \min_{i \in [m]} \{|X_i| - 2\delta_i| \} \).

The coding techniques of both Example 2 and Theorem 5 guarantee \( N < n \) if and only if \( |X_i| \geq 2\delta_i + 1 \) for all \( i \in [m] \). While the technique in Example 2 yields \( N = n - 1 \) and holds for any \( \mathbb{F}_q \), Theorem 5 may provide further reductions in the value of \( N \) but may require a large finite field \( \mathbb{F}_q \), so that the appropriate MDS code required for this construction exists.

#### B. Partition Based Coding Schemes

We can improve upon the techniques of Section IV-A by partitioning the bipartite graph into induced subgraphs and applying the coding techniques of Example 2 or Theorem 5 on each of these subgraphs, such as with the cycle covering [6], [9] and the partition multicast [1], [4] techniques proposed for index coding.
1) Partitioning using the repetition code technique: For each element $C \in \Phi(B)$, denote the subgraph of $B$ induced by $x_C$ as $B_C = (U_C, x_C, E_C)$. Then $B_C$ represents the $(m_C, n_C, X_C, \delta_C)$ BNSI problem, where $m_C = |U_C| = \{| u_i \in U | x_{C,i} \cap X_C \neq \phi \}$, $n_C = |C|$, $X_C = (X_C,i, X_C,\ i = X_C \cap C, \forall u_i \in U_C)$ and $E_C = \{| u_i, x_j \} \in E | u_i \in U_C, j \in C\}$. Now since $C \in \Phi(B)$, for every $u_i \in U_C, | x_{C,i} | = |X_C \cap C| \geq 2\delta_C + 1$. Therefore if we apply the coding technique of Example 2 for $(m_C, n_C, X_C, \delta_C)$, i.e., the encoder matrix is the transpose of the parity-check matrix of $[n_C, 1]$ repetition code, we can save one transmission compared to uncoded scheme for the $(m_C, n_C, X_C, \delta_C)$ BNSI subproblem.

For some integer $K$, let $C_1, C_2, \ldots, C_K$ be pairwise disjoint elements of $\Phi(B)$, and let $R = [\varnothing] \cup (C_1 \cup C_2 \cup \cdots \cup C_K)$. We apply the coding scheme of Example 2 on each subset $C_1, C_2, \ldots, C_K$ and transmit the information symbols indexed by the set $R$ uncoded. The codelength for this scheme is

$$N = \sum_{k=1}^{K} (|C_k| - 1) + |R| = \sum_{k=1}^{K} |C_k| + |R| - K = n - K.$$ 

At a receiver $i \in [m]$, any information symbol $x_j$, $j \in X_{C,i}$, is decoded using any valid decoding scheme for the $(m_C, n_C, X_C, \delta_C)$ BNSI problem. It can be shown that choosing $C_1, \ldots, C_K$ as disjoint, rather than allowing intersection, does not affect the minimum possible broadcast length $N$ achieved through this technique [15, Lemma 9]. We summarize the results of this subsection as

**Theorem 6.** Let $E$ be a largest collection of disjoint elements of $\Phi(B)$ for an $(n, m, X, \delta)$ BNSI problem, then $N_{q, opt}(m,n,X,\delta) \leq n - |E|$ for any finite field $\mathbb{F}_q$.

2) Partitioning using MDS codes: For any $(m,n,X,\delta)$ BNSI problem with the associated bipartite graph $B = (U, P, E)$, we have the following partition based technique for an $(m,n,X,\delta)$ problem.

**Theorem 7.** The collection of sets $\Phi(B)$ contains a unique maximal element $C_{\text{max}}$ and the subgraph $B'$ of $B$ induced by the information symbols $P \setminus x_{\text{max}}$ satisfies $\Phi(B') = \phi$.

Hence $C_{\text{max}}$ is the unique largest set in the collection $\Phi(B)$, and for the BNSI subproblem obtained by removing the information symbols $x_{\text{max}}$ uncoded scheme is optimal. We remark that this maximum element $C_{\text{max}}$ of $\Phi(B)$, can be determined using a simple algorithm (Lemma 11 of [15]). Given these properties of $C_{\text{max}}$ we can use the following partition based technique for an $(m,n,X,\delta)$ problem.

We transmit the information symbols $P \setminus x_{\text{max}}$ uncoded using $n - |C_{\text{max}}|$ channel uses. For some choice of integer $K$, we partition the maximum element $C_{\text{max}}$ of $\Phi(B)$ into $K$ disjoint subsets $S_1, S_2, \ldots, S_K \subset C_{\text{max}}$. Note that for each $a \in [K]$, the subgraph $B_a = (U_a, x_{S_a}, E_a)$ induced by $x_{S_a}$ denotes the $(m_a, n_a, X_a, \delta_a)$ BNSI problem where $m_a = |U_a| = \{| u_i \in U | S_a \cap X_a \neq \phi \}$, $n_a = |S_a|$, $X_a = (X_a, i, X_a, i = X_a \cap S_a, \forall u_i \in U_a)$ and $E_a = \{| u_i, x_j \} \in E | u_i \in U_a, j \in S_a\}$. For $a \in [K]$, let $d_a = \min_{u_i \in U_a} (|X_r \cap S_a| - 2\delta_a)^+$. We now use the MDS based coding technique of Theorem 5 independently on each subproblem $(m_a, n_a, X_a, \delta_a)$, $a \in [K]$. Using this technique, while encoding the information symbols indexed by $S_a$, we save $d_a$ transmissions compared to the uncoded scheme. Therefore the total number of transmissions that can be saved, compared to uncoded transmission, is $d_{\text{sum}} = \sum_{a=1}^{K} d_a$. The optimal partitioning is the solution to the following optimization problem.

**Optimization 1.**

Maximize

$$\sum_{a=1}^{K} d_a, \text{ where } d_a = \min_{u_i \in U_a} (|X_r \cap S_a| - 2\delta_a)^+$$

subject to $1 \leq K \leq n$, $S_1, S_2, \ldots, S_K \subset C_{\text{max}}$ such that for any $a, a' \in [K], a \neq a'$, $S_a \cap S_{a'} = \phi, \sum_{a=1}^{K} S_a = C_{\text{max}}$.

**Theorem 8.** Let $D_{\text{sum}}$ be the solution to the optimization problem Optimization 1. Then $N_{q, opt} \leq n - D_{\text{sum}}$ for every large enough finite field $\mathbb{F}_q$.

Theorem 8 provides a better upper bound on $N_{q, opt}$ than Theorem 6. This is evident by considering the choice $S_1 = C_1, \ldots, S_K = C_K$ for the MDS based partitioning. However, Theorem 8, in general, requires a large field size to utilize MDS codes, while Theorem 6 is valid for any finite field $\mathbb{F}_q$.

V. BNSI PROBLEM AND INDEX CODING

A. Brief Review of Index Coding

**Index Coding** [1], [2] is the problem of broadcasting a vector of $n_{IC}$ symbols $x_{IC} = (x'_1, x'_2, \ldots, x'_n) \in \mathbb{F}^{n_{IC}}$ to $m_{IC}$ users $u'_1, u'_2, \ldots, u'_m$ over a noiseless broadcast channel. The $i^{th}$ user $u'_i$ has prior knowledge of a part of the message vector as side information denoted as $x'_{IC,i}$. $x_{IC} \subset [n_{IC}]$, and demands the message $x'_{IC,i}$, for the function $f : [n_{IC}] \rightarrow [n_{IC}]$ denotes the demands at the receivers and is such that $f(i) \neq x'_{IC,i}$, for all $i \in [m_{IC}]$. Note that unlike BNSI problems, the side information in Index Coding problems is noise free and do not include the demanded message symbol. Defining $X_{IC} = (x_{IC,1}, x_{IC,2}, \ldots, x_{IC,m_{IC}})$, we describe this Index Coding problem as the $(m_{IC}, n_{IC}, X_{IC, f})$ Index Coding problem. The design objective is to construct a code with as small a codelength $N_{IC}$ as possible such that each user can recover its demanded message using the transmitted codeword and its own side information. For a scalar linear Index Code of length $N_{IC}$ for the $(m_{IC}, n_{IC}, X_{IC, f})$ Index Coding problem the transmitted codeword is $x_{IC}L_{IC}$, where $L_{IC} \in \mathbb{F}^{n_{IC}} \times X_{IC}$ is the corresponding encoder matrix. The minimum codelength $N_{IC}$ among all valid scalar linear coding schemes for the $(m_{IC}, n_{IC}, X_{IC, f})$ Index Coding problem over the field $\mathbb{F}_q$ will be denoted as $N_{q, opt, IC}(m_{IC}, n_{IC}, X_{IC, f})$. From Corollary 3.20 in [12] it follows that that $L_{IC}$ is a valid encoder matrix for the $(m_{IC}, n_{IC}, X_{IC, f})$ scalar linear Index Coding problem if and only if
\[ z_{IC} \neq 0, \quad \forall z \in I_{IC}(q, m_{IC}, n_{IC}, \chi_{IC}, f), \]

where \( I_{IC}(q, m_{IC}, n_{IC}, \chi_{IC}, f) \) is defined as the set

\[ \bigcup_{i=1}^{n_{IC}} \{ z \in F_{q}^{m_{IC}} | z_{\chi_{i,IC}} = 0 \} \cup \{ z \in F_{q}^{m_{IC}} | z_{\chi_{i,IC}} \neq 0 \} \].

B. Index Coding problem equivalent to a given BNSI problem

Given an \((m, n, \chi, \delta)\) BNSI problem, we now construct an Index Coding problem such that the two problems have identical (scalar) linear coding solutions. For each \( i \in [m] \), define

\[ \hat{\mu}_i = \begin{cases} |\chi_i| \times \left( \frac{|\chi_i| - 1}{2\delta_i - 1} \right) & \text{if } |\chi_i| \geq 2\delta_i \\ \text{otherwise} & \end{cases} \]

and let \( \hat{\mu} = \sum_{i=1}^{m} \hat{\mu}_i \). The number of information symbols in the constructed Index Coding problem is \( n \) and the number of receivers is \( \hat{\mu} \). The demands and the side information of the \( \hat{\mu} \) receivers are as follows. For each \( i \in [m] \), for each possible choice of an element \( p \in \chi_i \) and each possible choice of \( Q \subseteq \chi_i \setminus \{ p \} \) with \( |Q| = \min(|\chi_i| - 1, 2\delta_i - 1) \), we define a user \( u'_{i,Q} \) in the Index Coding problem with demand \( f(j) = p \) and side information \( X_{i,IC} = \chi_i \setminus (Q \cup \{ p \}) \). The number of users in the constructed Index Coding problem is the total number of ways that the elements \( i, p \) and \( Q \) can be chosen and is equal to \( \hat{\mu} \). We will represent this Index Coding problem using the tuple \((\hat{\mu}, n, \chi, f)\) where \( \chi = (\chi_1,IC, \ldots, \chi_{\hat{\mu},IC}) \).

Theorem 9. For any given \((m, n, \chi, \delta)\) BNSI problem and the corresponding \((\hat{\mu}, n, \chi, f)\) Index Coding problem \( I(q, m, n, \chi, \delta) = I_{IC}(q, \hat{\mu}, n, \chi, f) \).

Proof sketch: A vector \( z \in I \) only if there exists an \( i \in [m] \) such that \( wt(z_{\chi_i}) \in 2[\delta_i] \), i.e., there exists a \( p \in \chi_i \) such that \( z_p \neq 0 \) and there exists a \( Q \subset \chi_i \setminus \{ p \} \) with \( |Q| = \min(|\chi_i| - 1, 2\delta_i - 1) \) such that \( z_Q = 0 \). Thus, from construction, there exists a \( j \in [\hat{\mu}] \) such that \( z_{\hat{\mu},j} \neq 0 \) and \( z_{\hat{\mu},j,IC} = 0 \), and hence, \( z \in I_{IC} \), see (3). Reverse is similar. ■

From Theorem 9, and the code design criteria for BNSI problem (Theorem 1) and Index Coding problem (2), we conclude that a matrix \( L \) is a valid encoder for the BNSI problem if and only if it is valid for the constructed Index Coding problem. Thus, the family of linear BNSI problems is a subset of scalar linear index coding problems. Further, this subset relationship is strict as shown by the following example, which is a scalar linear Index Coding problem with no equivalent linear BNSI problem.

Example 3. Consider the scalar linear Index Coding problem with field size \( q = 2 \), \( m_{IC} = n_{IC} = 3 \), \( f(i) = i \) and \( \chi_{1,IC} = \{ 1, 2, 3 \} \setminus \{ i \} \) for \( i \in [m_{IC}] \). For this problem, the set \( I_{IC} = \{ z \in F_{2}^{m_{IC}} | wt(z) = 1 \} \). It is straightforward to verify that it is impossible to construct a BNSI problem over \( F_{2} \) with \( n = 3 \) symbols, for any choice of \( \delta_i \geq 1 \) and \( m \geq 1 \), such that \( I(m, n, \chi, \delta) = I_{IC}(m_{IC}, n_{IC}, \chi_{IC}, f) \). ■

The equivalence guaranteed by Theorem 9 allows us to use any known scalar linear Index Coding technique to solve BNSI problems. However, the number of users \( \hat{\mu} \) in the equivalent Index Coding problem is considerably larger than that of the BNSI problem, resulting in equivalent Index Coding problems that are more complex than the corresponding BNSI problems. In contrast, our proposed techniques (Sections III and IV) exploit the structure inherent in BNSI networks, and provide a direct and elegant approach to solve BNSI problems.

VI. CONCLUSION AND DISCUSSION

We considered the problem of transmission in a broadcast channel where each user owns an erroneous version of its own demanded message as side information. We derived bounds on the optimal codelength, provided code constructions based on the parity-check matrices of error correcting codes, and a relation between linear BNSI problems and scalar linear Index Coding problems. It will be interesting to derive tighter bounds on the optimal codelength, and to consider a probabilistic model of noise in the side information.

REFERENCES

[1] Y. Birk and T. Kol, “Informed-source coding-on-demand (iscod) over broadcast channels,” in INFOCOM ’98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 3, Mar 1998, pp. 1257–1264 vol.3.
[2] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, “Index coding with side information,” IEEE Trans. Inf. Theory, vol. 57, no. 3, pp. 1479–1494, Mar. 2011.
[3] H. Maleki, V. R. Cadambe, and S. A. Jafar, “Index coding – an interference alignment perspective,” IEEE Trans. Inf. Theory, vol. 60, no. 9, pp. 5402–5432, Sept 2014.
[4] A. S. Tehrani, A. G. Dimakis, and M. J. Neely, “Bipartite index coding,” in Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, IEEE, 2012, pp. 2246–2250.
[5] A. Blasiak, R. Kleinberg, and E. Lubetzky, “Broadcasting with side information: Bounding and approximating the broadcast rate,” IEEE Trans. Inf. Theory, vol. 59, no. 9, pp. 5811–5823, Sept 2013.
[6] 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, July 2011, pp. 244–248.
[7] K. Shankugan, A. G. Dimakis, and M. Langberg, “Local graph coloring and index coding,” in Proc. 2013 IEEE Int. Symp. Inf. Theory, July 2013, pp. 1125–1156.
[8] L. Ong, C. K. Ho, and F. Lim, “The single-uniprior index-coding problem: The single-sender case and the multi-sender extension,” IEEE Transactions on Information Theory, vol. 62, no. 6, pp. 3165–3182, June 2016.
[9] M. J. Neely, A. S. Tehrani, and Z. Zhang, “Dynamic index coding for wireless broadcast networks,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7525–7540, Nov 2013.
[10] J. W. Kim and J. S. No, “Index coding with erroneous side information,” IEEE Transactions on Information Theory, vol. 63, no. 12, pp. 7687–7697, Dec 2017.
[11] ——, “Equivalences between network codes with link errors and index codes with side information errors,” CoRR, vol.abs/1705.09429v1, 2017. [Online]. Available: https://arxiv.org/abs/1705.09429v1.
[12] S. H. Dau, V. Skachek, and Y. M. Chee, “Error correction for index coding with side information,” IEEE Trans. Inf. Theory, vol. 59, no. 3, pp. 1517–1531, March 2013.
[13] S. Sr. Karat and B. S. Rajan, “Optimal linear error correcting index codes for some index coding problems,” in 2017 IEEE Wireless Communications and Networking Conference (WCNC), March 2017, pp. 1–6.
[14] S. Ghosh and L. Natarajan, “On linear codes for broadcasting with noisy side information,” in 2018 Twenty Fourth National Conference on Communications (NCC), to appear, India, Feb. 2018.
[15] S. Ghosh and Natarajan, “Linear codes for broadcasting with noisy side information,” CoRR, vol.abs/1801.02868, 2018. [Online]. Available: http://arxiv.org/abs/1801.02868.