Abstract—We view the index coding problem with an arbitrary number of source bits and potentially overlapping demands as an instance of rate-distortion with multiple receivers, each with distinct side information. By applying network-information-theoretic tools, we find the optimal rate for the special case in which each source bit is present at either all of the decoders, none of the decoders, all but one of the decoders, or all but two of the decoders.

I. INTRODUCTION

Consider the problem in which a single transmitter encodes a source, which consists of an i.i.d. vector of uniform bits at each time, into a single message that is broadcast losslessly to $n$ decoders. Each decoder has (a possibly empty) subset of the source component bits as side information, and wishes to losslessly reconstruct another (possibly empty) subset of the source components. This problem is called index coding [1] and it has been widely studied (e.g., [2], [3], [4]) since it was introduced [5].

Most work on index coding has applied tools from network coding or graph theory (e.g., [1]). But this problem can evidently be viewed as a special case of rate-distortion with multiple receivers, each with its own side information [6]. The distortion constraints take the form that a given function of the source, which simply extracts a subset of the bits, must be reproduced losslessly. Taking this viewpoint allows one to apply tools from network information theory to develop both coding schemes and optimality results.

In earlier work [7], we used this viewpoint to solve the index coding problem with three receivers, an arbitrary number of source components, and possibly overlapping demands. Note that some work on index coding assumes that if a particular source component must be reproduced by one decoder, then no other decoder must reproduce it (including a recent work that also uses network information theory tools [8]). If this is the case then we say that there are no overlapping demands. More commonly, it is assumed that each receiver demands exactly one bit; if this were not the case, then the receiver could be divided into multiple virtual receivers, each with the same side information as the original receiver and each demanding a single bit. Our earlier work allowed for both overlapping demands and multiple bits to be demanded by each receiver, although the number of receivers was limited to at most three, as noted above. Here we generalize that result to solve the $n$-receiver case in which each source information bit is present in the side information of none of the receivers, all of the receivers, $n-1$ of the receivers, or $n-2$ of the receivers. Note that this recovers our earlier result since, when there are at most three receivers, every source bit must fall into one of these four cases.

In following sections, first, we will provide the problem formulation. We will continue with an achievable scheme and find an explicit expression for the rate. Lastly, we will prove the converse result to show the optimality of our scheme.

II. PROBLEM FORMULATION AND MAIN RESULT

For the $n$ user general index coding problem, each decoder $\alpha$ wants to reconstruct $f_\alpha$, which is an arbitrary subset of the source $S$ that is a collection of i.i.d. Bernoulli(1/2) bits at the encoder. Therefore, there may be overlapping demands, i.e., more than one decoder may demand the same given bit. Also, each decoder $\alpha$ has side information $Y_\alpha$ consisting of an arbitrary subset of the source. We assume that decoders do not demand their own side information since they already have it.

Let $S_J$ denote the part of the source which each decoder in a subset $J$ of $[n] \equiv \{1, \ldots, n\}$ does not have and all decoders in $[n] \setminus J$ have as side information. If $J = \{\alpha\}$, i.e., a singleton, then for ease of notation we use $S_\alpha$ instead of $S_{\{\alpha\}}$. Since there are $n$ decoders, we group the elements of $S$ into $2^n$ disjoint sets such that $S = \bigcup_{J \subseteq \{0, \ldots, n\}} S_J$. Note that each $S_J$ may be empty, may consist of a single bit or, may consist of multiple bits.

Let $G_n = S_{[n]}$ denote the elements of the source that none of the decoders have, $G_n = S_0$ denote the elements all decoders have, $G_{n-1} = \bigcup_{\alpha \in [n]} S_\alpha$ denote elements that $n-1$ of decoders have, $G_{n-2} = \bigcup_{(\alpha, \beta)} S_{\{\alpha, \beta\}}$ denote elements that $n-2$ of decoders have and so on. To ease the notation for the rest of the paper, whenever we write a set $\{\alpha, \beta\}$, we assume $\alpha \neq \beta$ unless otherwise stated. Then source $S$ can be represented as $S = \{G_0, G_n, G_{n-1}, \ldots, G_1\}$, as shown Fig. 1.

The demand $f_\alpha$ at decoder $\alpha$ can be written in terms of components $S_J$ of $S$. For this, we introduce the following notation.

Let $f_{J, I}$ denote the demand that is a subset of source $S_J$ and is required by each decoder in the subset $I$ of $[n]$ and by no decoders in $[n] \setminus I$. If $I = \{\alpha\}$, then for ease of notation...
we use \( f_{\alpha n} \) instead of \( f_{\{\alpha\} n} \). We will generally assume that \( I \subseteq J \) since only decoders in \( J \) may have a demand about \( S_J \) and decoders in \( [n] \setminus J \) already have \( S_J \) as side information. If \( I \nsubseteq J \), \( f_{IJ} \) is empty. Also, \( f_{IJ} \) and \( f_{KJ} \) are independent (i.e., \( f_{IJ} \perp f_{KJ} \)) for all possible choices of \( I, K \) and \( J \) with \( I \neq K \) since \( f_{IJ} \cap f_{KJ} = \emptyset \) unless \( I = K \). Lastly, each \( f_{IJ} \) may be empty, a single bit or may consist of multiple bits.

We have denoted the source \( S = \{G_0, G_n, G_{n-1}, \ldots, G_1\} \) and demands in terms of \( S_J \)'s. From now on, we consider an ordered set structure on \( S \) which naturally induces orders on \( S_J \)'s. Then, each demand \( f_{IJ} \) is also an ordered set which can be seen as a vector. We use the same operational definitions as in [7].

**Definition 1:** An \((m, M)\) code consists of mappings

\[
\begin{align*}
  f &: \{0,1\}^{km} \rightarrow \{1, \ldots, M\} \\
  g_1 &: \{1, \ldots, M\} \times \{0,1\}^{k_1 m} \rightarrow \{0,1\}^{l_1 m} \\
  g_2 &: \{1, \ldots, M\} \times \{0,1\}^{k_2 m} \rightarrow \{0,1\}^{l_2 m} \\
  &\vdots \\
  g_n &: \{1, \ldots, M\} \times \{0,1\}^{k_n m} \rightarrow \{0,1\}^{l_n m},
\end{align*}
\]

for some \( k \) where we call \( f \) the encoding function at the encoder and \( g_a \) the decoding function at decoder \( \alpha \) where \( \alpha \in [n] \).

**Definition 2:** The probability of error for a given code is defined as

\[
P_e = P\{g_1(f(S^m), Y_1^m) \neq f_1^m(S^m), \ldots, g_n(f(S^m), Y_n^m) \neq f_n^m(S^m)\}.
\]

**Definition 3:** The rate \( R \) is achievable if there exists a sequence of \((m, M)\) codes with rate \( m^{-1} \log M \leq R \) such that the probability of error, \( P_e \), tends to zero as \( m \) tends to infinity.

**Definition 4:** The optimal rate \( R_{opt} \) is defined as

\[
R_{opt} = \inf \{R|R \text{ is achievable}\}.
\]

As our main result, we find the optimal rate for the general index coding problem when the source \( S \) consists only of \( \{G_0, G_n, G_{n-1}, G_{n-2}\} \). In other words, the source components can only be present at none of the decoders, all decoders, \( n - 1 \), or \( n - 2 \) of the decoders. The optimal rate is stated in the following theorem.

**Theorem 1:** The optimal rate, \( R_{opt} \), for the general index coding problem where \( S = \{G_0, G_n, G_{n-1}, G_{n-2}\} \) is

\[
R_{opt} = \max\{R_1, \ldots, R_n\}
\]

where

\[
R_i = H\{f_i\{\bigcup_{1 \leq j \leq n} f_{\{\alpha\}j}(I_{\{\alpha\},j}^{(1)})\} \ldots, f_{i}(I_{\{\alpha\},j}^{(n)})\} + \max_{j,\in\{n\}\setminus i} f_{\{\beta\}j}(I_{\{\beta\},j})
\]

In the next section, we find an achievable scheme giving an upper bound on \( R_{opt} \).

### III. Achievable Scheme

As we did similarly in [7], we utilize the achievable scheme in [9] which is based on random coding.

**Proposition 1:** Let \( L_I = [n]\setminus I \). Then optimal rate \( R_{opt} \) is upper bounded by

\[
R_{opt} \leq \min_{U_{L_{(1)}}, U_{L_{(2)}} \in \{1, \ldots, n\}} \left\{ \max_{i \in L_0} \left\{ I(S; U_{L_{(i)}}, Y_i) \right\} + \sum_{j=1}^{n} \max_{i \in L_0} \left\{ I(S; U_{L_{(j)}}, U_{L_{(i)}}, Y_i) \right\} \right\}
\]

where the minimization is over the set of all random variables \((U_{L_{(1)}}, U_{L_{(1)}}, \ldots, U_{L_{(n)}})\) jointly distributed with \((S)\) such that

1) There exist functions \( g_1(U_{L_{(1)}}, U_{L_{(2)}}, \ldots, U_{L_{(n)}}, Y_1), \ldots, g_n(U_{L_{(1)}}, U_{L_{(1)}}, \ldots, U_{L_{(n)}}, Y_n) \) where

\[
g_1(U_{L_{(1)}}, U_{L_{(1)}}, \ldots, U_{L_{(n)}}, Y_1) = f_1(S),
\]

\[
\ldots
\]

\[
g_n(U_{L_{(1)}}, U_{L_{(1)}}, \ldots, U_{L_{(n)}}, Y_n) = f_n(S).
\]

2) \( U_{L_{(1)}} \perp U_{L_{(1)}}(U_{L_{(2)}} - Y_k), \forall i, j, k \in [n] \) and \( i \neq j \).

Here, random variable \( U_{L_{(i)}} \) can be considered as the message sent only to \([n]\setminus I \) set of decoders. In general, one can have \( U_{L_{(i)}} \) for every \( I \subset [n] \), i.e., there is a possible message \( U_{L_{(i)}} \) for each \([n]\setminus I \) set of decoders. Here we select only \( n + 1 \) of these to be nontrivial. Notice that these \( n + 1 \) variables should be specified in the minimization to explicitly characterize the achievable rate and this is a nontrivial task in general. Therefore, we introduce the following procedure to select \( U_{L_{(i)}} \) to get an explicit expression. Note that there is no demand related to \( G_n \) since all decoders have it as side information. We place all demands of all decoders related to \( G_0, G_{n-1} \) into \( U_{L_{0}} \). The remaining demands are subsets of \( G_{n-2} \), i.e., components that \( n - 2 \) of decoders have. For any \( S_J \in G_{n-2}, \) where \( J = \{\alpha, \beta\}, \alpha \neq \beta \) decoder \( \alpha \) and decoder \( \beta \) are the two decoders that do not have \( S_J \) as side information. Then we put \( f_{\{\alpha\}j} \) into \( U_{L_{0}} \). For \( f_{\alpha n} \) and \( f_{\beta n} \), we have the following procedure.

Note that there are \( n(n-1) \) different non-overlapping pairs of demands related to \( G_{n-2} \) since \( |G_{n-2}| = \frac{n(n-1)}{2} \) and there
are two demands \( f_{\alpha,A} ) \) and \( f_{\beta,A} \) for each \( S(\alpha,\beta) \in G_{n-2} \). Also, note that for each decoder \( \alpha \) there are \( n-1 \) non overlapping demands, \( f_{\alpha,A} \) related to \( G_{n-2} \). Therefore, we can put all those non overlapping demands into a matrix \( A \) with \( n \) rows and \( n-1 \) columns in the following way, \( \alpha^b \) row, denoted by \( A_\alpha \), consists of demands \( f_{\beta(\alpha,\beta)} \) where \( \beta \) runs over the set \([n]\backslash\{\alpha\}\).

Note that \( A_\alpha \) does not contain any demand of the decoder \( \alpha \). Also, for each \( f_{\beta(\alpha,\beta)} \), all entries of \( A_\alpha \) other than \( f_{\beta(\alpha,\beta)} \) exist as side information at decoder \( \beta \). Lastly, as the size of each demand in \( A_\alpha \) can be different, we arrange the entries in \( A_\alpha \) in an increasing order with respect to their sizes. If two demands are in equal size, which one is put first does not matter. This completes the construction of the matrix \( A \). For each \( A_\alpha \) in \( A \) we apply the following \( \vee \) operation. A motivation for this definition can be obtained from our previous work \[7\] on the three decoder case.

**Definition 5:** Let \( a_i \), \( i \in \{1, \ldots, n-1\} \) be a vector. Assume without loss of generality that \( l_1 \leq l_2 \leq \ldots \leq l_{n-1} \) where \( l_i = |a_i| \) denotes the number of elements in \( a_i \). Then,

\[
(a_1, a_2, \ldots, a_{n-1})^\oplus = (a_1^\oplus, \ldots, a_{n-1}^\oplus)
\]

where

\[
(a_1^\oplus, \ldots, a_{n-1}^\oplus) = (a_1 \oplus (a_2)_1 \oplus \ldots \oplus (a_n)_1),
\]

\[
(a_2)_2 - l_1 \oplus \ldots \oplus (a_{n-1})_2 - l_1,
\]

\[
\ldots,
\]

\[
(a_{n-1})_{l_{n-1} - l_{n-2}}.
\]

Before rearranging the terms in \( R_{ach} \) further, we would like to make the following remarks.

**Remark 1:** From the Definition \[5\] we know that \( \forall k \in L_\alpha, \alpha \in [n] \) \( H(A_{a_1}^\oplus | Y_k) \) is equal to

\[
\begin{align*}
H(A_{a_1} \oplus (A_{a_2})_{a_1} \oplus \ldots \oplus (A_{a_{n-1}})_{a_1} | Y_k) \\
\leq H(A_{a_1}, (A_{a_2})_{a_1}, \ldots, (A_{a_{n-1}})_{a_1} | Y_k) \\
\leq H((A_{a_1})_{a_1} | Y_k)
\end{align*}
\]

\[
H((A_{a_1})_{a_1} | Y_k) = |A_{a_1}| = \min_{\ell \in \{1, \ldots, n-1\}} |A_{a_\ell}| \text{ bits}.
\]

a: Since \( n - 2 \) of the \( \oplus \)ed terms in \( A_{a_1}^\oplus \) are contained in \( Y_k \).

b: Where \( A_{a_\ell} \) is the demand at decoder \( k \) related to \( G_{n-2} \). Hence,

\[
H(U_{L_\alpha}(\tau(Y_k)) = H(U_{L_\alpha}(\tau(Y_j)) = \min_{\ell \in \{1, \ldots, n-1\}} |A_{a_\ell}| \text{ bits}, \forall k, j \in L_\ell.
\]

Also, from \(4\) and \(5\), we deduce that selecting \( U_{L_\ell}(\tau) = A_{a_1} \) or \( U_{L_\ell}(\tau) = A_{a_2} \) from \( A_{a_1}, A_{a_2}, \ldots, A_{a_{n-1}} \), \( \forall \alpha \in [n] \) give the same rate. Therefore, we can select \( U_{L_\ell}(\tau) = A_{a_1}, (A_{a_2})_{a_1}, \ldots, (A_{a_{n-1}})_{a_1} \forall \alpha \in [n] \) without increasing the rate.

**Remark 2:** By Definition \[5\] \( \forall \alpha \in [n] \),

\[
H(A_{a_1}^\oplus | Y_\alpha) = H(A_{a_1}^\oplus, A_{a_2}^\oplus, \ldots, A_{a_{n-1}}^\oplus | Y_n)
\]

\[
= H(A_{a_1}^\oplus, Y_\alpha) + H(A_{a_2}^\oplus, \ldots, A_{a_{n-1}}^\oplus | Y_\alpha)
\]

\[
= H(U_{\{\alpha\}} \Xi_{[n]} f_{\{\alpha\}} | Y_\alpha) + H(A_{a_2}^\oplus, \ldots, A_{a_{n-1}}^\oplus | Y_\alpha)
\]

\[
= H(U_{\{\alpha\}} \Xi_{[n]} f_{\{\alpha\}} | Y_\alpha) + \max_{j \in \ell} \{ |A_{a_\ell}| - \min_{j \in \ell} |A_{a_\ell}| \}
\]

Using the remarks above, we can write \( R_{ach} \) as

\[
\max_{k \in L_\alpha} \left\{ H(f_1 \Xi_{[n]} f_{\{\alpha\}} | Y_1), f_2 \Xi_{[n]} f_{\{\alpha\}} | Y_2), \ldots, f_n \Xi_{[n]} f_{\{\alpha\}} | Y_n) \right\}
\]

\[+ \sum_{\alpha \in L_\ell} H(U_{L_\alpha}(\tau(Y_1)) + \max_{k \in L_\ell} \{ H(U_{L_\alpha}(\tau(Y_1)) + \sum_{\alpha \in L_{[n]} \backslash \alpha} H(U_{L_{[n]}}(\tau(Y_1)) \}
\]

\[
= \max_{k \in L_\ell} \left\{ H(f_1 \Xi_{[n]} f_{\{\alpha\}} | Y_1), f_2 \Xi_{[n]} f_{\{\alpha\}} | Y_2), \ldots, f_n \Xi_{[n]} f_{\{\alpha\}} | Y_n) \right\}
\]

\[+ \sum_{\alpha \in L_\ell} H(U_{L_{[n]}}(\tau(Y_n)) + \max_{k \in L_\ell} \{ H(U_{L_{[n]}}(\tau(Y_n)) \}
\]

\[
\ldots,
\]

\[
H(f_1 \Xi_{[n]} f_{\{\alpha\}} | Y_1), f_2 \Xi_{[n]} f_{\{\alpha\}} | Y_2), \ldots, f_n \Xi_{[n]} f_{\{\alpha\}} | Y_n) \right\}
\]

\[+ \sum_{\alpha \in L_{[n]}} H(U_{L_{[n]}}(\tau(Y_n)) + \max_{k \in L_{[n]}} \{ H(U_{L_{[n]}}(\tau(Y_n)) \}
\]

\[
\ldots,
\]

\[
H(f_1 \Xi_{[n]} f_{\{\alpha\}} | Y_1), f_2 \Xi_{[n]} f_{\{\alpha\}} | Y_2), \ldots, f_n \Xi_{[n]} f_{\{\alpha\}} | Y_n) \right\}
\]

\[+ \sum_{\alpha \in L_{[n]}} H(U_{L_{[n]}}(\tau(Y_n)) + \max_{k \in L_{[n]}} \{ H(U_{L_{[n]}}(\tau(Y_n)) \}
\]
\[
\begin{align*}
\frac{b}{c} & \max \left\{ H(f_1 \{ U(1, \beta) \subseteq [n], f(1, \beta) \} ) , H(f_2 \{ U(2, \beta) \subseteq [n], f(2, \beta) \} ) , \\
& \ldots , f_n \{ U(\alpha, \beta) \subseteq [n], f(\alpha, \beta) \} , A^\beta \{ \{ \alpha \in [n] A^\alpha \} \} | Y_1 \} \\
& + \sum_{\alpha \in L(1)} H(A^\alpha \{ \{ \alpha \in [n] A^\alpha \} \} | Y_1 ) + | A^\alpha _1 | \\
& H(f_1 \{ U(1, \beta) \subseteq [n], f(1, \beta) \} , f_2 \{ U(2, \beta) \subseteq [n], f(2, \beta) \} , \\
& \ldots , f_n \{ U(\alpha, \beta) \subseteq [n], f(\alpha, \beta) \} , A^\beta \{ \{ \alpha \in [n] A^\alpha \} \} | Y_2 \} \\
& + \sum_{\alpha \in L(2)} H(A^\alpha \{ \{ \alpha \in [n] A^\alpha \} \} | Y_2 ) + | A^\alpha _2 | \\
& \ldots , \\
& H(f_1 \{ U(1, \beta) \subseteq [n], f(1, \beta) \} , f_2 \{ U(2, \beta) \subseteq [n], f(2, \beta) \} , \\
& \ldots , f_n \{ U(\alpha, \beta) \subseteq [n], f(\alpha, \beta) \} , A^\beta \{ \{ \alpha \in [n] A^\alpha \} \} | Y_n \} \\
& + \sum_{\alpha \in L(n)} H(A^\alpha \{ \{ \alpha \in [n] A^\alpha \} \} | Y_n ) + | A^\alpha _n | \\
& \leq \max \left\{ H(f_1 \{ U(1, \beta) \subseteq [n], f(1, \beta) \} , f_2 \{ U(2, \beta) \subseteq [n], f(2, \beta) \} , \\
& \ldots , f_n \{ U(\alpha, \beta) \subseteq [n], f(\alpha, \beta) \} , A^\beta \{ \{ \alpha \in [n] A^\alpha \} \} | Y_1 \} \\
& + H(\{ U_{L(1)} A^\alpha _1 \} | Y_1 ) + | A^\alpha _1 | \\
& H(f_1 \{ U(1, \beta) \subseteq [n], f(1, \beta) \} , f_2 \{ U(2, \beta) \subseteq [n], f(2, \beta) \} , \\
& \ldots , f_n \{ U(\alpha, \beta) \subseteq [n], f(\alpha, \beta) \} , A^\beta \{ \{ \alpha \in [n] A^\alpha \} \} | Y_2 \} \\
& + H(\{ U_{L(2)} A^\alpha _2 \} | Y_2 ) + | A^\alpha _2 | \\
& \ldots , \\
& H(f_1 \{ U(1, \beta) \subseteq [n], f(1, \beta) \} , f_2 \{ U(2, \beta) \subseteq [n], f(2, \beta) \} , \\
& \ldots , f_n \{ U(\alpha, \beta) \subseteq [n], f(\alpha, \beta) \} , A^\beta \{ \{ \alpha \in [n] A^\alpha \} \} | Y_n \} \\
& + H(\{ U_{L(n)} A^\alpha _n \} | Y_n ) + | A^\alpha _n | \right\} \\
\end{align*}
\]

\( a: \) By Remark 1 (i.e. \( H(U_{L(\alpha)} | Y_k) = H(U_{L(\alpha)} | Y_j) \) \( \forall k, j \in L(\alpha) \) and expanding terms inside \( \max_{k \in L(\alpha)} \)).

b: Since \( H(U_{L(\alpha)} | Y_k) = H(A^\alpha _k | Y_k) \) \( \forall k \in L(\alpha) \) and \( H(U_{L(\alpha)} | Y_k) = | A^\alpha _k | \) \( \forall k \in L(\alpha) \) (Remark 1).

c: Chain rule and \( U_{L(1)} \perp U_{L(2)} | Y_k \) \( \forall i, j \in L(k) \).

d: Chain rule and \( U_{L(1)} \perp U_{L(2)} | Y_k \) \( \forall i, j \in L(k) \).

e: Using Remark 2 that \( | A_{\alpha|} = \min_j | A_{\alpha|} | \) and chain rule.

Therefore, \( R_{ach} \) can be stated as

\[
R_{ach} = \max \{ R_{ach1}, \ldots , R_{achn} \},
\]

where

\[
R_{ach} = \max \{ f_1 \{ U(1, \beta) \subseteq [n], f(1, \beta) \} , \ldots , f_n \{ U(\alpha, \beta) \subseteq [n], f(\alpha, \beta) \} | Y_1 \} + \max | A_{ij} |.
\]

In the next section, we find a lower bound which matches \( R_{ach} \) stated above.

IV. CONVERSE

Let \( R \) be an achievable rate. Then \( \forall \epsilon > 0 \), there exists a code with \( P_e < \epsilon \) and with sufficiently large block length \( m \).

Now, consider any such code and let \( J \) be the output of an encoding function. For any \( R \) we can write,

\[
mR \geq H(J)
\]

\[
= I(f_1^m, \ldots , f_n^m, Y_1^m, \ldots , Y_n^m, J)
\]

\[
\geq I(Y_1^m, J) + I(f_2^m, J| Y_1^m) + I(f_3^m, J| f_2^m, Y_1^m) + \ldots \\
+ I(f_4^m, J| f_3^m, f_2^m, Y_1^m) + \ldots + I(f_n^m, J| f_{n-1}^m, \ldots , f_1^m, Y_1^m)
\]

\[
\geq I(f_1^m, J| Y_1^m) + I(f_2^m, J| f_1^m, f_3^m, Y_1^m) + \ldots \\
+ I(f_n^m, J| f_{n-1}^m, \ldots , f_1^m, Y_1^m)
\]

\[
\geq m \left[ H(f_1 | Y_1) + H(f_2 | Y_2, f_1, Y_1) + \ldots \right. \\
+ \left. H(f_n | Y_n, \ldots , f_1, Y_1) \right] - m e \sum_{i=2}^n \log |f_i^m| + \sum_{i=2}^n \log |f_i^m| (f_{j-i}^m)
\]

\( a: \) Chain Rule

b: Neglecting terms of the form \( I(Y_1^m, J, \ldots) \).

c: By definition \( P_e(f_1^m) \neq g_i(J, Y_1^m) \leq P_e < \epsilon \) and Fano’s inequality.

Since \( m \) is sufficiently large and by taking \( \epsilon \) arbitrarily small, we get

\[
R \geq H(f_1 | Y_1) + H(f_2 | Y_2, f_1, Y_1) + \ldots \\
+ H(f_n | Y_n, \ldots , f_1, Y_1),
\]
where \( H(f_2|Y_2, f_1, Y_1) \) is equal to

\[
H(f_2|f_1, Y_1) = H(f_2 \setminus \{(u_{2,\beta})\in[n]f_2(u_{2,\beta})\}\{u_{2,\beta})\in[n]f_2(u_{2,\beta})\}|f_1, Y_1) \\
= H(f_2 \setminus \{(u_{2,\beta})\in[n]f_2(u_{2,\beta})\}, \{u_{2,\beta})\in[n]f_2(u_{2,\beta})\}|f_1, Y_1) \\
= H(f_2 \setminus \{(u_{2,\beta})\in[n]f_2(u_{2,\beta})\}, A_{1\beta}|f_1, Y_1),
\]

where \( A_{1\beta} = f_2(u_{2,\beta}) \)

\[
= H(f_2 \setminus \{(u_{2,\beta})\in[n]f_2(u_{2,\beta})\}|f_1, Y_1) + H(A_{1\beta}) \\
= H(f_2 \setminus \{(u_{2,\beta})\in[n]f_2(u_{2,\beta})\}|f_1, Y_1) + |A_{1\beta}|
\]

and \( H(f_3|Y_2, f_2, Y_2, f_1, Y_1) \) is equal to

\[
H(f_3|f_2, Y_2, f_1, Y_1) = H(f_3 \setminus \{(u_{3,\beta})\in[n]f_3(u_{3,\beta})\}, \{u_{3,\beta})\in[n]f_3(u_{3,\beta})\}|f_2, Y_2, f_1, Y_1) \\
= H(f_3 \setminus \{(u_{3,\beta})\in[n]f_3(u_{3,\beta})\}|f_2, Y_2, f_1, Y_1) \\
= H(f_3 \setminus \{(u_{3,\beta})\in[n]f_3(u_{3,\beta})\}|f_2 \setminus \{(u_{2,\beta})\in[n]f_2(u_{2,\beta})\}, \\
Y_2, f_1, Y_1),
\]

since \( \{u_{3,\beta})\in[n]f_3(u_{3,\beta})\} \subseteq \{Y_2, f_1, Y_1\} \)

\[
= H(f_3 \setminus \{(u_{3,\beta})\in[n]f_3(u_{3,\beta})\}|f_2 \setminus \{(u_{2,\beta})\in[n]f_2(u_{2,\beta})\}, \\
Y_2, f_1, Y_1),
\]

since \( \{u_{2,\beta})\in[n]f_2(u_{2,\beta})\} \subseteq \{Y_2, f_1, Y_1\} \)

Note that \( a \) can be written as \( f_3 = \{f_{3|n})\in[n]f_3(u_{3,\beta})\} \)

Then \( a \) can be written as \( f_3 = \{f_{3|n})\in[n]f_3(u_{3,\beta})\} \)

Hence, from (9) and (13), (8) can be written as

\[
R \geq H(f_1, f_2 \setminus \{(u_{2,\beta})\in[n]f_2(u_{2,\beta})\)}, \ldots, \\
f_n \setminus \{(u_{n,\beta})\in[n]f_n(u_{n,\beta})\}|Y_1) + |A_{1\beta}|
\]

Note that right hand side of (15) is \( R_{ach}\). Since problem is symmetric, similarly we can get all \( R_{ach}\). Then,

\[
R \geq max R_{ach} = R_{ach}
\]

proving that \( R_{ach} \) is optimal. This concludes the proof of the Theorem.

V. ACKNOWLEDGMENT

This work was supported by Intel, Cisco, and Verizon under the Video-Aware Wireless Networks (VAWN) program.

REFERENCES

[1] Z. Bar-Yossef, Y. Birn, T. S. Jayram, and T. Kol, “Index coding with side information,” in Foundations of Computer Science, 2006. FOCS ’06. 47th Annual IEEE Symposium on, 2006, pp. 197–206.

[2] A. Blasiak, R. Kleinberg, and E. Lubetzky, “Broadcasting with side information: Bounding and approximating the broadcast rate,” Information Theory, IEEE Transactions on, vol. 59, no. 9, pp. 5811–5823, 2013.

[3] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, “Broadcasting with side information,” in Foundations of Computer Science, 2008. FOCS ’08. 49th Annual IEEE Symposium on, 2008, pp. 823–832.

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

[5] Y. Birn 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, 1998, pp. 1257–1264 vol.3.

[6] C. Heegard and T. Berger, “Rate distortion when side information may be absent,” Information Theory, IEEE Transactions on, vol. 31, no. 6, pp. 727–734, 1985.

[7] S. Unal and A. B. Wagner, “General index coding with side information: Three decoder case,” in Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on, 2013, pp. 1137–1141.

[8] F. Atrashofana, B. Bander, Y.-H. Kim, E. Sasoglu, and L. Wang, “On the capacity region for index coding,” in Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on, July 2013, pp. 962–966.

[9] R. Timo, T. Chan, and A. Grant, “Rate distortion with side-information at many decoders,” Information Theory, IEEE Transactions on, vol. 57, no. 8, pp. 5240–5257, 2011.