Distortion Bounds for Source Broadcast Problem

Lei Yu, Houqiang Li, Senior Member, IEEE, and Weiping Li, Fellow, IEEE

Abstract

This paper investigates the joint source-channel coding problem of sending a memoryless source over a memoryless broadcast channel. An inner bound and an outer bound on the achievable distortion region are derived, which respectively generalize and unify several existing bounds. As a consequence, we also obtain an inner bound and an outer bound for degraded broadcast channel case. When specialized to Gaussian source broadcast or binary source broadcast, the inner bound and outer bound not only can recover the best known inner bound and outer bound in the literature, but also can be used to generate some new results. Besides, we also extend the inner bound and outer bound to Wyner-Ziv source broadcast problem, i.e., source broadcast with side information available at decoders. Some new bounds are obtained when specialized to Wyner-Ziv Gaussian case and Wyner-Ziv binary case. In addition, when specialized to lossless transmission of a source with independent components, the bounds for source broadcast problem (without side information) is also used to achieve an inner bound and an outer bound on capacity region of general broadcast channel with common messages, which respectively generalize Marton’s inner bound and Nair-El Gamal outer bound to $K$-user broadcast channel case.

Index Terms

Joint source-channel coding (JSCC), hybrid coding, hybrid digital-analog (HDA), broadcast, Wyner-Ziv, side information, multivariate covering/packing, network information theory.

I. INTRODUCTION

As stated in Shannon’s source-channel separation theorem [2], cascading source coding and channel coding does not lose the optimality for the point-to-point communication systems. This separation theorem does not only suggest a simple system architecture in which source coding and channel coding are separated by a universal digital interface, but also guarantees that such architecture does not incur any asymptotic performance loss. Consequently, it forms the basis of the architecture of today’s communication systems. However, for many multi-user communication systems, the optimality of such a separation does not hold any more [3], [4]. Therefore, an increasing amount of literature focus on joint source-channel coding (JSCC) in multi-user setting.

The authors are all with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: yulei@ustc.edu.cn, lihq@ustc.edu.cn, wpli@ustc.edu.cn). The material in this paper will be presented in part at IEEE ISIT 2016 [1].
One of the most classical problems in this area is JSCC of transmitting a Gaussian source over average power constrained $K$-user Gaussian broadcast channel. Goblick [3] observed that when the source bandwidth and the channel bandwidth are matched (i.e., one channel use per source sample) linear uncoded transmission (symbol-by-symbol mapping) is optimal. However, the optimality of such a simple linear scheme cannot be extended to the case of the bandwidth mismatch. One way to approximately characterize the achievable distortion region is finding its inner bound and outer bound. For inner bound, analog coding schemes or hybrid coding schemes have been studied in a vast body of literature [4], [5], [6], [7]. For 2-user Gaussian broadcast communication, Prabhakaran et al. [7] gave the tightest inner bound so far, which is achieved by hybrid digital-analog (HDA) scheme. On the other hand, Reznic et al. [8] derived a nontrivial outer bound for 2-user Gaussian broadcast problem with bandwidth expansion (i.e., more than one channel uses per source sample) by introducing an auxiliary random variable (or remote source). Tian et al. [9] extended this outer bound to $K$-user case by introducing more than one auxiliary random variables. Similar to the results of Reznic et al., the outer bound given by Tian et al. is also nontrivial only for bandwidth expansion case [10]. Beyond broadcast communication, Minero et al. [15] considered sending memoryless correlated source transmitted over memoryless multi-access channel and derived an inner bound using a unified framework of hybrid coding, and also Lee et al. [20] derived a unified achievability result for memoryless network communication.

Besides, in [6], [17], [18] Wyner-Ziv source communication problem was investigated, in which side information of the source is available at decoder(s). Shamai et al. [6] studied the problem of sending Wyner-Ziv source over point-to-point channel, and proved that for such communication system separate coding (which combines Wyner-Ziv coding with channel coding) does not incur any loss of optimality. Nayak et al. [17] and Gao et al. [18] investigated Wyner-Ziv source broadcast problem, and obtained an outer bound by simply applying cut-set bound (the minimum distortion achieved in point-to-point setting) for each receiver.

In this paper, we consider JSCC of transmitting a memoryless source over $K$-user memoryless broadcast channel, and give an inner bound and an outer bound on the achievable distortion region. The inner bound is derived by a unified framework of hybrid coding inspired by [15], and the outer bound is derived by introducing auxiliary random variables (or remote sources) as in [8] and [9]. Both these two bounds are generalizations and unifications of several existing bounds in the literature. Besides, as a consequence, we also obtain an inner bound and an outer bound for degraded broadcast channel case. Owing to the generalization of our results, when specialized to Gaussian source broadcast and binary source broadcast, our inner bound could recover best known performance achieved by hybrid coding, and our outer bound could recover the best known outer bounds given by Tian et al. [9] and Khezeli et al. [11]. Moreover, for these cases, our bounds can also be used to generate some new results. Besides, we also extend the inner bound and outer bound to Wyner-Ziv source broadcast problem, i.e., source broadcast with side information at decoders. When specialized to Wyner-Ziv Gaussian case and Wyner-Ziv binary case, our bounds reduce to some new bounds. In addition, when specialized to lossless transmission of a source with independent components, the bounds for source broadcast problem (without side information) is also used to achieve an inner bound and an outer bound on capacity region of general broadcast channel with common messages,
which respectively generalize Marton’s inner bound and Nair-El Gamal outer bound to $K$-user broadcast channel case.

The rest of this paper is organized as follows. Section II summarizes basic notations, definitions and preliminaries, and formulates the problem. Section III gives the main results for source broadcast problem, including general, degraded, Gaussian and binary cases. Section IV extends the results to Wyner-Ziv source broadcast problem. Finally, Section V gives the concluding remarks.

II. Problem Formulation and Preliminaries

A. Notation

Throughout this paper, we follow the notation in [16]. For example, for discrete random variable $X \sim p_X$ on alphabet $\mathcal{X}$ and $\epsilon \in (0,1)$, the set of $\epsilon$-typical $n$-sequences $x^n$ (or the typical set in short) is defined as $T^{(n)}_\epsilon(X) = \{x^n : \left| \frac{|\{i : x_i = x\}|}{n} - p_X(x) \right| \leq \epsilon p_X(x) \text{ for all } x \in \mathcal{X}\}$. When it is clear from the context, we will use $T^{(n)}_\epsilon$ instead of $T^{(n)}_\epsilon(X)$.

In addition, we use $X_A$ to denote the vector $(X_j : j \in A)$, use $[i : j]$ to denote the set $\{\lfloor i \rfloor, \lfloor i \rfloor + 1, \cdots, \lfloor j \rfloor\}$, and use $\mathbf{1}$ to denote an all-one vector (similarly, use $\mathbf{2}$ to denote an all-2 vector). We say vector $m_{[1:N]}$ is smaller than vector $m'_{[1:N]}$ if $m_j = m'_j$, $k < j \leq K$ and $m_k < m'_k$ for some $k$. For two vectors $m_\mathcal{I}$ and $m'_\mathcal{I}$, we say $m_\mathcal{I}$ is component-wise unequal to $m'_\mathcal{I}$, if $m_i \neq m'_i$ for all $i \in \mathcal{I}$, and denote it as $m_\mathcal{I} \Leftrightarrow m'_\mathcal{I}$. Besides, we use $\mathbf{1}\{A\}$ to denote indicator function of event $A$, i.e.,

$$\mathbf{1}\{A\} = \begin{cases} 1, & \text{if } A \text{ is true;} \\ 0, & \text{if } A \text{ is false.} \end{cases}$$

B. Problem Formulation

Consider the source broadcast system shown in Fig. 1. The discrete memoryless source (DMS) $S^n$ is first coded into $X^n$ using a source-channel code, then transmitted to $K$ receivers through a discrete memoryless broadcast channel (DM-BC) $p_{Y_{[1:K]}|X}$, and finally, the receiver $k$ produces source reconstruction $\hat{S}_k^n$ from the received signal $Y_k^n$.
Definition 1 (Source). A discrete memoryless source (DMS) is specified by a probability mass function (pmf) $p_S$ on a finite alphabet $S$. The DMS $p_S$ generates an i.i.d. random process $\{S_t\}$ with $S_t \sim p_S$.

Definition 2 (Broadcast Channel). A $K$-user discrete memoryless broadcast channel (DM-BC) is specified by a collection of conditional pmfs $p_{Y[1:K]|X}$ on finite output alphabet $Y_1 \times \cdots \times Y_K$ for each $x$ in finite input alphabet $X$.

Definition 3 (Degraded Broadcast Channel). A DM-BC $p_{Y[1:K]|X}$ is stochastically degraded (or simply degraded) if there exist a random vector $\hat{Y}_{[1:K]}$ such that $\hat{Y}_k | \{X = x\} \sim p_{Y_k|X}(\hat{y}_k|x), 1 \leq k \leq K$, i.e., $\hat{Y}_{[1:K]}$ has the same conditional marginal pmfs as $Y_{[1:K]}$ (given $X$), and $X \rightarrow \hat{Y}_K \rightarrow \hat{Y}_{K-1} \rightarrow \cdots \rightarrow \hat{Y}_1$ form a Markov chain. In addition, as a special case, if $X \rightarrow Y_K \rightarrow Y_{K-1} \rightarrow \cdots \rightarrow Y_1$, i.e., $\hat{y}_k = y_k, 1 \leq k \leq K$, then $p_{Y[1:K]|X}$ is physically degraded.

Definition 4. An $n$-length source-channel code is defined by the encoding function $x^n : S^n \mapsto X^n$ and a sequence of decoding functions $\hat{s}_k : Y^n_k \mapsto \hat{S}_k^n, 1 \leq k \leq K$, where $\hat{S}_k$ is the alphabet of source reconstruction at receiver $k$.

For any $n$-length source-channel code, the induced distortion is defined as
\[
\mathbb{E} d_k \left( S^n, \hat{S}_k^n \right) = \frac{1}{n} \sum_{t=1}^n \mathbb{E} d_k \left( S_t, \hat{S}_{k,t} \right),
\]
for $1 \leq k \leq K$, where $d_k(s, \hat{s}_k) : S \times \hat{S}_k \mapsto [0, +\infty]$ is a distortion measure function for receiver $k$.

Definition 5. For transmitting source $S$ over channel $p_{Y[1:K]|X}$, if there exists a sequence of source-channel codes such that
\[
\limsup_{n \to \infty} \mathbb{E} d_k \left( S^n, \hat{S}_k^n \right) \leq D_k,
\]
then we say that the distortion tuple $D_{[1:K]}$ is achievable.

Definition 6. For transmitting source $S$ over channel $p_{Y[1:K]|X}$, the admissible distortion region is defined as
\[
\mathcal{R} \triangleq \{ D_{[1:K]} : D_{[1:K]} \text{ is achievable} \}.
\]

The admissible distortion region $\mathcal{R}$ only depends on the marginal distributions of $p_{Y[1:K]|X}$, hence for source broadcast over stochastically degraded channel it suffices to only consider the broadcast over physically degraded channel.

In addition, Shannon’s source-channel separation theorem shows that the minimum distortion for transmitting source over point-to-point channel satisfies
\[
R_k (D_k) = C_k,
\]
\[1\]To simplify notation, the Markov chain is assumed in this direction. Note that this differs from that in the conference version [1].
where \( R_k (\cdot) \) is the rate-distortion function of the source with distortion measure \( d_k \), and \( C_k \) is the capacity of the channel of the receiver \( k \). Therefore, the optimal distortion (Shannon limit) is

\[
D_k^* = R_k^{-1} (C_k).
\]

Obviously,

\[
\mathcal{R} \subseteq \mathcal{R}^* \triangleq \{ D_{[1:K]} : D_k \geq D_k^*, 1 \leq k \leq K \},
\]

where \( \mathcal{R}^* \) is named trivial outer bound.

In the system above, source bandwidth and channel bandwidth are matched. In this paper, we also consider the communication system with bandwidth mismatch, whereby \( m \) samples of a DMS are transmitted through \( n \) uses of a DM-BC. For this case, bandwidth mismatch factor is defined as \( b = \frac{n}{m} \).

### C. Multivariate Covering/Packing Lemma

Two important results we need to prove the achievability part in this work are the following lemmas, both of which are generalized versions of the existing covering/packing lemmas.

Let \( (U, V_{[0:k]}) \sim p_{U,V_{[0:k]}} \), and let \( (U^n, V^n_0) \sim p_{U^n,V^n_0} \) be a random vector sequence. For each \( j \in [1 : k] \), let \( \mathcal{A}_j \subseteq [1 : j-1] \). Assume \( \mathcal{A}_j \) satisfies if \( i \in \mathcal{A}_j \), then \( \mathcal{A}_i \subseteq \mathcal{A}_j \). For each \( j \in [1 : k] \) and each \( m, \mathcal{A}_j \subseteq \prod_{i \in \mathcal{A}_j} [1 : 2^{n r_i}] \), let \( V^n_j (m, \mathcal{A}_j, m_j) \), \( m_j \in [1 : 2^{n r_j}] \), be pairwise conditionally independent random sequences, each distributed according to \( \prod_{i=1}^n p_{V_j | V_{\mathcal{A}_j}, V_0} (v_{j,i} | v_{\mathcal{A}_j}, i(m_{\mathcal{A}_j}), v_0, i) \). Hence for each \( j \in [1 : k] \), \( \mathcal{A}_j \cup \{0\} \) denotes the index set of the random variables on which the codeword \( V^n_j \) is superposed. Based on the notations above, we have the following generalized Multivariate Covering Lemma and generalized Multivariate Packing Lemma.

**Lemma 1** (Multivariate Covering Lemma). Let \( \epsilon' < \epsilon \). If \( \lim_{n \to \infty} \mathbb{P} \left( (U^n, V^n_0) \in \mathcal{T}^{(n)}_{\epsilon'} \right) = 1 \), then there exists \( \delta(\epsilon) \) that tends to zero as \( \epsilon \to 0 \) such that

\[
\lim_{n \to \infty} \mathbb{P} \left( (U^n, V^n_0, V^n_{[1:k]} (m_{[1:k]})) \in \mathcal{T}^{(n)}_{\epsilon} \text{ for some } m_{[1:k]} \right) = 1,
\]

if \( \sum_{j \in \mathcal{J}} r_j > \sum_{j \in \mathcal{J}} H (V_j | V_{\mathcal{A}_j}, V_0) - H (V_j | V_0 U) + \delta(\epsilon) \) for all \( \mathcal{J} \subseteq [1 : k] \) such that \( \mathcal{J} \neq \emptyset \) and if \( j \in \mathcal{J} \), then \( \mathcal{A}_j \subseteq \mathcal{J} \).

**Lemma 2** (Multivariate Packing Lemma). There exists \( \delta(\epsilon) \) that tends to zero as \( \epsilon \to 0 \) such that

\[
\lim_{n \to \infty} \mathbb{P} \left( (U^n, V^n_0, V^n_{[1:k]} (m_{[1:k]})) \in \mathcal{T}^{(n)}_{\epsilon} \text{ for some } m_{[1:k]} \right) = 0,
\]

if \( \sum_{j \in \mathcal{J}} r_j < \sum_{j \in \mathcal{J}} H (V_j | V_{\mathcal{A}_j}, V_0) - H (V_j | V_0 U) - \delta(\epsilon) \) for some \( \mathcal{J} \subseteq [1 : k] \) such that \( \mathcal{J} \neq \emptyset \) and if \( j \in \mathcal{J} \), then \( \mathcal{A}_j \subseteq \mathcal{J} \).

Note that all the existing covering and packing lemmas such as [16, Lem. 8.2] and [19, Lem. 4], only involve single-layer codebook. Our Multivariate Covering and Packing Lemmas generalize them to the case of multilayer codebook, and certainly our Covering/Packing Lemmas could recover all of them.
III. Source Broadcast

A. Source Broadcast

Now, we bound the distortion region for source broadcast communication. To write the inner bound, we first introduce an auxiliary random variable $V_j, 1 \leq j \leq N \triangleq 2^K - 1$ for each of the $2^K - 1$ nonempty subsets $\mathcal{G}_j \subseteq [1 : K]$, and let $V_j$ denote a common message transmitted from sender to all the receivers in $\mathcal{G}_j$. The $V_j$ corresponds to a subset $\mathcal{G}_j$ by the following one-to-one mapping.

Sort all the nonempty subsets $\mathcal{G} \subseteq [1 : K]$ in the decreasing order\(^2\). Map the $j$th subset in the resulting sequence to $j$. Obviously this mapping is one-to-one corresponding. For example, if $K = 3$, then $\mathcal{G}_1 = \{1, 2, 3\}, \mathcal{G}_2 = \{2, 3\}, \mathcal{G}_3 = \{1, 3\}, \mathcal{G}_4 = \{1, 2\}, \mathcal{G}_5 = \{3\}, \mathcal{G}_6 = \{2\}, \mathcal{G}_7 = \{1\}$.

Besides, let
\[
\mathcal{A}_j \triangleq \{i \in [1 : N] : \mathcal{G}_j \subseteq \mathcal{G}_i\}, 1 \leq j \leq N, \tag{9}
\]
\[
\mathcal{D}_k \triangleq \{i \in [1 : N] : k \in \mathcal{G}_i\}, 1 \leq k \leq K. \tag{10}
\]

Later we will show that they respectively correspond to the index set of the random variables on which the codeword $V_j^n$ is superposed, and the index set of decodable codewords $V_j^n$’s for receiver $k$ in the proposed hybrid coding scheme; see Appendix C-A. Decoder $k$ is able to recover correctly the $V_j^n$, designated by the encoder with probability approaching 1 as $n \to \infty$ if $j \in \mathcal{D}_k$. In addition, it is easy to verify that if $j \in \mathcal{D}_k$, then $\mathcal{A}_j \subseteq \mathcal{D}_k$. It means that the proposed codebook satisfies that if information $V_j^n$ can be recovered correctly by receiver $k$, then $V_j^n_{\mathcal{A}_j}$ can also be recovered correctly by it.

Based on the notations above, we define distortion region
\[
\mathcal{R}_{(i)} = \bigg\{D_{[1 : K]} : \text{There exist some pmf } p_{V_{[1 : N]}|S}, \text{ vector } r_{[1 : N]}, \text{ and functions } x \left( v_{[1 : N]}, s \right), \hat{s}_k \left( v_{D_k}, y_k \right), 1 \leq k \leq K \text{ such that}
\]
\[
\mathbb{E}d_k \left( S, \hat{S}_k \right) \leq D_k, 1 \leq k \leq K,
\]
\[
\sum_{j \in \mathcal{J}} r_j > \sum_{j \in \mathcal{J}^c} H \left( V_j | V_{\mathcal{A}_j} \right) - H \left( V_{\mathcal{J}} | S \right)
\]
for all $\mathcal{J} \subseteq [1 : N]$ such that $\mathcal{J} \neq \emptyset$ and if $j \in \mathcal{J}$, then $\mathcal{A}_j \subseteq \mathcal{J}$,
\[
\sum_{j \in \mathcal{J}} r_j < \sum_{j \in \mathcal{J}^c} H \left( V_j | V_{\mathcal{A}_j} \right) - H \left( V_{\mathcal{J}} | Y_k V_{\mathcal{J}} \right)
\]
for all $1 \leq k \leq K$ and for all $\mathcal{J} \subseteq \mathcal{D}_k$ such that $\mathcal{J}^c \triangleq \mathcal{D}_k \setminus \mathcal{J} \neq \emptyset$ and if $j \in \mathcal{J}$, then $\mathcal{A}_j \subseteq \mathcal{J}$ \} \bigg\}. \tag{11}

\(^2\)We say a set $\mathcal{G}$ is larger than another $\mathcal{H}$ if $|\mathcal{G}| > |\mathcal{H}|$, or $|\mathcal{G}| = |\mathcal{H}|$ and there exists some $1 \leq i \leq |\mathcal{G}|$ such that $\mathcal{G}[i] > \mathcal{H}[i]$ and $\mathcal{G}[i] = \mathcal{H}[i] \forall l \leq i - 1$, where $\mathcal{G}[i]$ (or $\mathcal{H}[i]$) denotes the $i$th largest element in $\mathcal{G}$ (or $\mathcal{H}$).
Besides, define another distortion region

\[
R^{(o)} = \left\{ D_{[1:K]} : \text{There exists some pmf } p_{\hat{S}_{[1:K]}|S} \text{ such that } \right. \\
\left. \mathbb{E} d_k \left( S, \hat{S}_k \right) \leq D_k, 1 \leq k \leq K, \right. \\
\text{and for any pmf } p_{U_{[1:L]}|S}, \text{ one can find } p_{X,U_{[1:L]},W_{[1:K]}|W_{[1:K]}} \text{ satisfying } \\
\sum_{i=1}^{m} I \left( \hat{S}_{A_i} ; U_{B_i} | U_{j=0}^{i-1} B_j \right) \leq \sum_{i=1}^{m} I \left( Y_{A_i} ; \hat{U}_{B_i}, W_{A_{i+1}} | \hat{U}_{j=0}^{i-1} B_j, W_{A_i}, \hat{W}_{A_{i-1}} \right), \\
\text{for any } m \geq 1, A_i \subseteq [1 : K], B_i \subseteq [1 : L], 0 \leq i \leq m, A_0, A_{m+1} = \emptyset, \\
\text{and } \hat{W}_{A_i} \triangleq W_{A_i}, \text{ if } i \text{ is odd}; W_{A_i}' \text{, otherwise, or } \hat{W}_{A_i} \triangleq W_{A_i}, \text{ if } i \text{ is odd}; W_{A_i}, \text{ otherwise } \right\}. 
\]

(12)

Then we have the following theorem. The proof is given in Appendix C.

**Theorem 1.** *For transmitting source* \( S \) *over general broadcast channel* \( p_{Y_{[1:K]}|X} \),

\[
R^{(i)} \subseteq R \subseteq R^{(o)}. 
\]

(13)

**Remark 1.** The inner bound in Theorem 1 can be easily extended to Gaussian or any other well-behaved continuous-alphabet source-channel pair by standard discretization method \cite[Thm. 3.3]{16}, and moreover for this case the outer bound still holds. Theorem 1 can be also extended to the case of source-channel bandwidth mismatch, where \( m \) samples of a DMS are transmitted through \( n \) uses of a DM-BC. This can be accomplished by replacing the source and channel symbols in Theorem 1 by supersymbols of lengths \( m \) and \( n \), respectively. Besides, Theorem 1 could be also extended to the problem of broadcasting correlated sources (by modifying the distortion measure) or source broadcast with channel input cost (by adding channel input constraint).

The inner bound \( R^{(i)} \) in Theorem 1 is achieved by a unified hybrid coding scheme depicted in Fig. 2. In this scheme, the codebook has a layered (or superposition) structure, and consists of randomly and independently generated codewords \( V_{[1:N]}^{n_{1:N}}(m_{[1:N]}), m_{[1:N]} \in \prod_{i=1}^{N} [1 : 2^{n_{[1:N]}}], \) where \( r_{[1:N]} \) satisfies \( (11) \). At encoder side, upon source sequence \( S^n \), the encoder produces digital messages \( M_{[1:N]} \) with \( M_i \) meant for all the receivers \( k \) satisfying
\( i \in D_k \). Then, the codeword \( V^n_{\{1:N\}}(M_{\{1:N\}}) \) and the source sequence \( S^n \) are used to generate channel input \( X^n \) by symbol-by-symbol mapping \( x(v^n_{\{1:N\}}, s) \). At decoder sides, upon received signal \( Y^n_k \), decoder \( k \) could reconstruct \( M_{D_k} \) (and also \( V^n_{D_k}(M_{D_k}) \)) losslessly, and then \( \hat{S}_k^n \) is produced by symbol-by-symbol mapping \( \hat{s}_k(v_{D_k}, y_k) \). Such a scheme could achieve any \( D_{\{1:K\}} \) in the inner bound \( \mathcal{R}^{(i)} \).

To reveal essence of such hybrid coding, the digital transmission part of this hybrid coding can be roughly understood as cascade of a \( K \)-user Gray-Wyner source-coding and a \( K \)-user Marton’s broadcast channel-coding, which share a common codebook. According to [16, Thm. 13.3], the encoding operation of Gray-Wyner source-coding with rates \( r_{\{1:N\}} \) is successful if \( \sum_{j \in J} r_j > \sum_{j \in J'} H(V_j | V_{A_j}) - H(V_j | S) \) for all \( J \subseteq \{1 : N\} \) such that \( J \neq \emptyset \) and if \( j \in J \), then \( A_j \subseteq J \), and according to [16, Thm. 5.2] the decoding operation of Marton’s broadcast channel-coding with rates \( r_{\{1:N\}} \) is successful if \( \sum_{j \in J} r_j < \sum_{j \in J'} H(V_j | V_{A_j}) - H(V_j | Y_k V_{J'}) \) for all \( 1 \leq k \leq K \) and for all \( J \subseteq D_k \) such that \( J^c \neq \emptyset \) and if \( j \in J \), then \( A_j \subseteq J \). Note that in Marton’s broadcast channel-coding, \( r_{\{1:N\}} \) does not correspond to the regular broadcast-rates (i.e., the rates of subcodebooks in multicoding), but is the rates of the whole codebook. Since the proposed hybrid coding satisfies the two sufficient conditions above, \( V^n_{D_k}(M_{D_k}) \) could be losslessly transmitted to receiver \( k \). Note that such informal understanding is inaccurate owing to the use of symbol-by-symbol mapping, but it provides the rationale for our scheme. Besides, the design of such unified hybrid coding is inspired by the hybrid coding scheme for sending correlated sources over multi-access channel in [15].

The outer bound \( \mathcal{R}^{(o)} \) in Theorem 1 are derived by introducing auxiliary random variables (or remote sources) \( U^n_{\{1:L\}} \). This proof method could also be found in [9], [11, Thm. 2] and [12, Lem. 1]. In [9] it is used to derive the outer bound for Gaussian source broadcast, and in [11, Thm. 2] and [12, Lem. 1] it is used to derive the outer bounds for sending source over 2-user general broadcast channel. A deeper understanding of this proof method has been given by Khezeli et al. in [12]. \( p_{\hat{S}_{\{1:K\}}|S} \) can be considered as a virtual broadcast channel realized over physical broadcast channel \( p_{Y_{\{1:K\}}|X} \), and hence certain measurements based on \( p_{\hat{S}_{\{1:K\}}|S} \) are less than or equal to those based on \( p_{Y_{\{1:K\}}|X} \). This leads to the necessary conditions on the communication. Note that this proof method generalizes the one used to derive trivial outer bound, but it does not always result in a tighter outer bound than the trivial one [10].

For 2-user broadcast case, the inner bound in Theorem 1 reduces to

\[
\mathcal{R}^{(i)} = \left\{ (D_1, D_2) : \text{There exist some pmf } p_{V_0, V_1, V_2 | S}, \right. \\
\text{and functions } x(v_0, v_1, v_2, s), \hat{s}_k(v_0, v_k, y_k), k = 1, 2, \text{ such that} \\
E d_k \left( S, \hat{S}_k \right) \leq D_k, \\
I(V_0 V_k; S) < I(V_0 V_k; Y_k), j = 1, 2, \\
I(V_0 V_1 V_2; S) + I(V_1 V_2|V_0) < \min \{ I(V_0; Y_1), I(V_0; Y_2) \} + I(V_1; Y_1 V_0) + I(V_2; Y_2|V_0), \\
I(V_0 V_1; S) + I(V_0 V_2; S) + I(V_1 V_2|V_0 S) < I(V_0 V_1; Y_1) + I(V_0 V_2; Y_2) \right\}.
\]

(14)

This inner bound was first given in by Yassaee et. al [21]. On the other hand, for 2-user broadcast case, letting...
Let $L = 3$, the outer bound in Theorem 1 reduces to
\[
R^{(o)} = \{(D_1, D_2) : \text{There exists some pmf } p_{\hat{S}_1, \hat{S}_2|S} \text{ such that}
\]
\[
\mathbb{E}d_k \left( S, \hat{S}_k \right) \leq D_k, k = 1, 2,
\]
and for any pmf $p_{U_1, U_2, U_3|S}$, one can find $p_{X, \tilde{U}_1, \tilde{U}_2, \tilde{U}_3, W_1, W_2, W_3} \text{ satisfying}
\]
\[
I \left( \hat{S}_{A_1}; U_{B_1}|U_{B_0} \right) \leq I \left( Y_{A_1}; \tilde{U}_{B_1}|\tilde{U}_{B_0}\tilde{W}_{A_1} \right),
\]
\[
I \left( \hat{S}_{A_1}; U_{B_1}|U_{B_0} \right) + I \left( \hat{S}_{A_2}; U_{B_2}|U_{B_0}\tilde{U}_{B_1}\tilde{W}_{A_1} \right) \leq I \left( Y_{A_1}; \tilde{U}_{B_1}|\tilde{U}_{B_0}\tilde{W}_{A_1}\tilde{W}_{A_2} \right) + I \left( Y_{A_2}; \tilde{U}_{B_2}|\tilde{U}_{B_0}\tilde{W}_{A_1}\tilde{W}_{A_2} \right),
\]
\[
I \left( \hat{S}_{A_1}; U_{B_1}|U_{B_0} \right) + I \left( \hat{S}_{A_2}; U_{B_2}|U_{B_0}\tilde{U}_{B_1}\tilde{W}_{A_1} \right) + I \left( \hat{S}_{A_3}; U_{B_3}|U_{B_0}\tilde{U}_{B_1}\tilde{U}_{B_2}\tilde{W}_{A_1}\tilde{W}_{A_2}\tilde{W}_{A_3} \right),
\]
\[
+ I \left( Y_{A_2}; \tilde{U}_{B_2}\tilde{W}_{A_3}|\tilde{U}_{B_3}\tilde{U}_{B_1}\tilde{U}_{B_2}\tilde{W}_{A_1}\tilde{W}_{A_2} \right) + I \left( Y_{A_3}; \tilde{U}_{B_3}|\tilde{U}_{B_0}\tilde{U}_{B_1}\tilde{U}_{B_2}\tilde{W}_{A_3} \right),
\]
\[
\text{for any } A_i \subseteq [1:2], B_i \subseteq [1:3], 0 \leq i \leq 3,
\]
and $\tilde{W}_{A_i} \triangleq W_{A_i}$, if $i$ is odd; $W'_{A_i}$, otherwise, or $\tilde{W}_{A_i} \triangleq W'_{A_i}$, if $i$ is odd; $W_{A_i}$, otherwise.
\[
(15)
\]
This outer bound is as tight as if not tighter than that given by Khezeli et al. in [12, Thm. 1]. This is because the outer bound in [12, Lem 1] is just $R^{(o)}$ with $A_i$ restricted to be an element (not a subset) of $[1:2]$. When consider lossless transmission of independent source, Theorem 1 can be used to achieve bounds on capacity region of general broadcast channel with common messages. In this case, $S = (M_j : 1 \leq j \leq N)$ and all $M_j$'s are independent with each other. For each $1 \leq j \leq N$, let $R_j$ denote the rate of the common message $M_j$ that is to be transmitted losslessly from sender to all the receivers in $G_j \subseteq [1:K]$. The correspondence between $M_j$ and $G_j$ is kept same to that between $V_j$ and $G_j$; see the beginning of this section. Then such achievable rates $R_{[1:N]}$ constitute the capacity region $C$.

Now, define rate region
\[
C^{(i)} = \{ R_{[1:N]} : \text{There exist some pmf } p_{V_{[1:N]}|S}, \text{ vector } r_{[1:N]}
\]
and function $x(v_{[1:N]})$ such that
\[
\sum_{j \in \mathcal{J}} r_j > \sum_{j \in \mathcal{J}} R_j + \sum_{j \in \mathcal{J}} H \left( V_j|V_{A_j} \right) - H \left( V_{A_j}|S \right)
\]
for all $\mathcal{J} \subseteq [1:N]$ such that $\mathcal{J} \neq \emptyset$ and if $j \in \mathcal{J}$, then $A_j \subseteq \mathcal{J}$,
\[
\sum_{j \in \mathcal{J}^C} r_j < \sum_{j \in \mathcal{J}^C} H \left( V_j|V_{A_j} \right) - H \left( V_{A_j}|V_{\mathcal{J}^C} \right)
\]
for all $1 \leq k \leq K$ and for all $\mathcal{J} \subseteq D_k$ such that $\mathcal{J}^c \triangleq D_k \setminus \mathcal{J} \neq \emptyset$ and if $j \in \mathcal{J}$, then $A_j \subseteq \mathcal{J}$.
\[
(16)
\]
and another rate region

\[ C^{(o)} = \{ R_{[1:N]} : \text{There exists some pmf } \prod_{i=1}^{N} p_{\hat{U}_i} p_{W_i|U_i} \text{ and function } x \left( \hat{u}_{[1:N]} \right) \text{ such that} \]

\[ \sum_{i=1}^{m} \sum_{j \in (U_{k \in A_i}, \mathcal{P}) \cap B_i} R_j \leq \sum_{i=1}^{m} I \left( Y_{A_i} ; \hat{U}_{B_i}, \hat{W}_{A_{i+1}} ; \hat{U}_{j=0} B_j \hat{W}_{A_{i-1}} \right), \]

for any \( m \geq 1, A_i \subseteq [1 : K], B_i \subseteq [1 : N], 0 \leq i \leq m, A_0, A_{m+1} \neq \emptyset, \)

and \( \hat{W}_{A_i} \triangleq W_{A_i}, \) if \( i \) is odd; \( \hat{W}_{A_i} \), otherwise, or \( \hat{W}_{A_i}' \), if \( i \) is odd; \( \hat{W}_{A_i} \), otherwise \}. \quad (17)

Then as a consequence of Theorem 1, we can establish the following bounds on the capacity region of general broadcast channel. The proof is omitted.

**Theorem 2.** For general broadcast channel \( p_{Y_{[1:K]}|X} \), the capacity region \( C \) with common messages satisfies

\[ C^{(i)} \subseteq C \subseteq C^{(o)}. \quad (18) \]

**Remark 2.** The inner bound and the outer bound in Theorem 2 respectively generalize Marton’s inner bound and Nair-El Gamal outer bound to the case of \( K \)-user broadcast channel.

### B. Source Broadcast over Degraded Channel

If the channel is degraded, define

\[ \mathcal{R}_{DBC}^{(i)} = \left\{ D_{[1:K]} : \text{There exist some pmf } p_{Y_{k}|SP_{V_{K-1}}|V_{k}} \cdots p_{V_{i}|V_{2}}, \right. \]

and functions \( x(v_k, s), \hat{s}_k(v_k, y_k), 1 \leq k \leq K \) such that

\[ \mathbb{E} d_k \left( S, \hat{S}_k \right) \leq D_k, \]

\[ I \left( S ; V_k \right) \leq \sum_{j=1}^{k} I \left( Y_j ; V_j | V_{j-1} \right), 1 \leq k \leq K, \text{ where } V_0 \triangleq \emptyset \}, \quad (19) \]

and

\[ \mathcal{R}_{DBC}^{(o)} = \left\{ D_{[1:K]} : \text{There exists some pmf } p_{Y_{[1:K]}|S} \text{ such that} \right. \]

\[ \mathbb{E} d_k \left( S, \hat{S}_k \right) \leq D_k, 1 \leq k \leq K, \]

\[ \left( I \left( \hat{S}_{[1:k]} ; S \right) : k \in [1 : K] \right) \in \mathcal{C}_{DBC} \left( p_{Y_{[1:K]}|X} \right), \]

and for any pmf \( p_{U_{K-1}|SP_{U_{K-2}}|U_{K-1}} \cdots p_{U_1|U_2}, U_0 \triangleq \emptyset, U_K \triangleq S, \)

\[ \left( I \left( \hat{S}_{[1:k]} ; U_k | U_{k-1} \right) : k \in [1 : K] \right) \in \mathcal{C}_{DBC} \left( p_{Y_{[1:K]}|X} \right) \text{ holds} \}, \quad (20) \]

where

\[ \mathcal{C}_{DBC} \left( p_{Y_{[1:K]}|X} \right) = \left\{ R_{[1:K]} : 0 \leq R_k \leq I \left( X ; Y_k \right), 1 \leq k \leq K \text{ for some pmf } p_X \right\} \quad (21) \]
and

\[
C_{DBC} \left( p_{Y_1\mid K}\mid X \right) = \left\{ R_{1\mid K} : \text{There exists some pmf } p_X p_{V_{K-1}} p_{V_{K-2}} \ldots p_{V_1} p_{V_2} \text{ such that} \right. \\
0 \leq R_k \leq I (Y_k; V_k \mid V_{k-1}) , 1 \leq k \leq K, \text{ where } V_0 \triangleq \emptyset, V_K \triangleq X \right\}.
\]  
(22)

Note that \( C_{DBC} \left( p_{Y_1\mid K}\mid X \right) \) is the capacity of degraded broadcast channel \( p_{Y_1\mid K}\mid X \), and by rate-splitting argument, it can be also expressed as

\[
C_{DBC} \left( p_{Y_1\mid K}\mid X \right) = \left\{ R_{1\mid K} : \text{There exists some pmf } p_X p_{V_{K-1}} p_{V_{K-2}} \ldots p_{V_1} p_{V_2} \text{ such that} \right. \\
R_k \geq 0, \sum_{j=1}^{k} R_j \leq \sum_{j=1}^{k} I (Y_j; V_j \mid V_{j-1}) , 1 \leq k \leq K, \text{ where } V_0 \triangleq \emptyset, V_K \triangleq X \right\}.
\]  
(23)

Then as a consequence of Theorem 1, the following theorem holds. The proof is given in Appendix D.

**Theorem 3.** For transmitting source \( S \) over degraded broadcast channel \( p_{Y_1\mid K}\mid X \),

\[
R^{(i)}_{DBC} \subseteq \mathcal{R} \subseteq R^{(o)}_{DBC}.
\]  
(24)

**C. Quadratic Gaussian Source Broadcast**

Consider sending Gaussian source \( S \sim \mathcal{N}(0, N_S) \) with quadratic distortion measure \( d_k(s, \hat{s}) = d(s, \hat{s}) \equiv (s - \hat{s})^2 \) over Gaussian broadcast channel \( Y_k = X + W_k, 1 \leq k \leq K \) with \( W_k \sim \mathcal{N}(0, N_k) \), \( N_1 \geq N_2 \geq \cdots \geq N_K \). Assume bandwidth mismatch factor is \( b \). Then the inner bound \( R^{(i)}_{DBC} \) in Theorem 3 could recover the best known inner bound so far [7, Thm. 5] by setting suitable random variables and symbol-by-symbol mappings.

**Corollary 1.** [7, Thm. 5] For transmitting Gaussian source \( S \) with quadratic distortion measure over 2-user Gaussian broadcast channel with bandwidth mismatch factor \( b \), \( R^{(i)}_{DBC} \subseteq \mathcal{R} \).

- For \( b < 1 \) (bandwidth compression)

\[
R^{(i)}_{DBC} = \{ (D_1(\lambda, \gamma), D_2(\lambda, \gamma)) : 0 \leq \lambda \leq 1, 0 \leq \gamma \leq 1 \},
\]  
(25)

where

\[
D_1(\lambda, \gamma) = \frac{bN_S}{\lambda^2 \gamma^2 + \lambda^2 P + N_1} + \frac{(1 - b)N_S}{\left( \frac{P + N_1}{\lambda P + N_1} \right)^{\frac{1}{\gamma}}},
\]  
(26)

\[
D_2(\lambda, \gamma) = \frac{bN_S}{\lambda^2 \gamma^2 + \lambda^2 P + N_1} + \frac{(1 - b)N_S}{\left( \frac{P + N_1}{\lambda P + N_1} \right)^{\frac{1}{\gamma}}},
\]  
(27)

- For \( b > 1 \) (bandwidth expansion)

\[
R^{(i)}_{DBC} = \{ (D_1(\lambda, \gamma), D_2(\lambda, \gamma)) : 0 \leq \lambda \leq 1, 0 \leq \gamma \leq 1 \},
\]  
(28)
where
\[
D_1(\lambda, \gamma) = \frac{N_S}{\left( \frac{\alpha N_s}{\alpha^2 N_s + N_1} \right)^{b-1}} \left( \frac{\lambda N_1}{\lambda^2 N_1 + N_1} \right)^{b-1},
\]
\[
D_2(\lambda, \gamma) = \frac{N_S}{\left( \frac{\alpha N_s}{\alpha^2 N_s + N_2} \right)^{b-1}} \left( \frac{\lambda N_2}{\lambda^2 N_2 + N_2} \right)^{b-1}.
\]

Proof: For \( b = \frac{n}{m} < 1 \) (bandwidth compression), the source \( S^n = (S^n, S^n_{n+1}) \) is transmitted over channel \( Y^n_k = X^n + W^n_k, k = 1, 2 \). Define a set of random variables \( (U_1^{m-n}, E_1^{m-n}, U_2^{m-n}, E_2^{m-n}, X_1^n, X_2^n, X_2^n, V_1, V_2) \) such that

\[
S^n_{n+1} = U_1^{m-n} + E_1^{m-n},
\]
\[
E_1^{m-n} = U_2^{m-n} + E_2^{m-n},
\]
\[
X_2^n = X_2^n + \beta \alpha S^n,
\]
\[
V_1 = (U_1^{m-n}, X_1^n),
\]
\[
V_2 = (V_1, U_2^{m-n}, X_2^n),
\]

where \( U_1^{m-n}, U_2^{m-n}, E_2^{m-n} \) are mutually independent Gaussian variables, \( X_2^n \) and \( X_1^n \) are Gaussian variables independent of all the other variables, \( \text{Var}(X_2^n) = \lambda \gamma P, \text{Var}(X_1^n) = (1 - \lambda) P, \text{Var}(E_1^n) = \frac{N_s}{\left( \frac{\alpha N_s}{\alpha^2 N_s + N_2} \right)^{b-1}}, \text{Var}(E_2^n) = \frac{N_s}{\left( \frac{\alpha N_s}{\alpha^2 N_s + N_1} \right)^{b-1}} \), and \( \alpha = \sqrt{\frac{\lambda (1 - \gamma) P}{N_s}}, \beta = \frac{\lambda \gamma P}{\alpha N_s} + N_2 \). Define a set of functions

\[
x^n(v_2, s^n) = x^n_2 + \alpha s^n + x^n_2 - \beta \alpha s^n = x^n_1 + \alpha s^n + x^n_2,
\]
\[
s^n_1(v_1, y^n_1) = \frac{\alpha N_s}{\alpha^2 N_s + N_1} (y^n_1 - x^n_1), u_1^{m-n},
\]
\[
s^n_2(v_2, y^n_2) = \frac{\alpha N_s}{\alpha^2 N_s + N_2} (y^n_2 - x^n_1), u_1^{m-n} + u_2^{m-n}).
\]

Substitute these variables and functions into the inner bound \( R^{(i)}_{DBC} \) in Theorem 3, then the \( b < 1 \) case in Corollary 1 is recovered.

For \( b = \frac{n}{m} > 1 \) (bandwidth expansion), the source \( S^n \) is transmitted over channel \( Y^n_k = U^n_k + W^n_k, k = 1, 2 \). Define a set of random variables \( (U_1^n, E_1^n, U_2^n, E_2^n, X_1^{n-m}, X_2^{n-m}, V_1, V_2) \) such that

\[
S^n = U_1^n + E_1^n,
\]
\[
E_1^n = U_2^n + E_2^n,
\]
\[
V_1 = (U_1^n, X_1^{n-m}),
\]
\[
V_2 = (V_1, U_2^n, X_2^{n-m}),
\]

where \( U_1^n, U_2^n, E_2^n \) are mutually independent Gaussian variables, \( X_1^{n-m} \) and \( X_2^{n-m} \) are two Gaussian variables independent of all the other random variables, \( \text{Var}(X_2) = \frac{(1 - \lambda)(1 - \gamma) P}{b-1}, \text{Var}(X_2) = \frac{\lambda (1 - \gamma) P}{b-1}, \text{Var}(E_1^n) = \frac{N_s}{\left( \frac{\alpha N_s}{\alpha^2 N_s + N_2} \right)^{b-1}} \), and \( \alpha = \sqrt{\frac{\lambda (1 - \gamma) P}{N_s}}, \beta = \frac{\lambda \gamma P}{\alpha N_s} + N_2 \). Define a set of functions

\[
x^n(v_2, s^n) = x^n_2 + \alpha s^n + x^n_2 - \beta \alpha s^n = x^n_1 + \alpha s^n + x^n_2,
\]
\[
s^n_1(v_1, y^n_1) = \frac{\alpha N_s}{\alpha^2 N_s + N_1} (y^n_1 - x^n_1), u_1^{m-n},
\]
\[
s^n_2(v_2, y^n_2) = \frac{\alpha N_s}{\alpha^2 N_s + N_2} (y^n_2 - x^n_1), u_1^{m-n} + u_2^{m-n}).
\]

Substitute these variables and functions into the inner bound \( R^{(i)}_{DBC} \) in Theorem 3, then the \( b < 1 \) case in Corollary 1 is recovered.
It can be proved that the Gaussian broadcast capacity region shrinks as the bandwidth increases. Additionally, the broadcast channel consisting of subchannels that achieves the optimal point-to-point distortion for each receiver. Assume the comparison of capacity regions for two Gaussian broadcast channels Gaussian broadcast channel of transformed into a forward Gaussian channel with the same probability distribution, assume Theorem 4. [9, Thm. 2] For transmitting Gaussian source \( S \) with quadratic distortion measure over \( K \)-user Gaussian broadcast channel with bandwidth mismatch factor \( b \),

\[
\mathcal{R} \subseteq \mathcal{R}_{DBC}^{(o)} \triangleq \left\{ D_{[1:K]} : \text{For any variables } + \infty = \tau_0 \geq \tau_1 \geq \cdots \geq \tau_K = 0, \frac{1}{b} \left\{ \frac{1}{2} \log \left( \frac{N_S + \tau_k}{D_k + \tau_k} \right) \left( \frac{D_k + \tau_{k-1}}{N_S + \tau_{k-1}} \right) : k \in [1:K] \right\} \in \mathcal{C}_{GBC} \right\},
\]

where \( \mathcal{C}_{GBC} \) denotes the capacity of Gaussian broadcast channel given by

\[
\mathcal{C}_{GBC} = \left\{ R_{[1:K]} : R_k \geq 0, 1 \leq k \leq K, N_{K+1} = 0, \sum_{k=1}^{K} (N_k - N_{k+1}) \exp \left( \frac{k}{2} \sum_{j=1}^{k} R_j \right) \leq P + N_1 \right\}.
\]

To compare \( \mathcal{R}_{DBC}^{(o)} \) with the trivial distortion bound, we consider a set of Gaussian test channels \( p_{S_{[1:K]}^*|S} \), \( 1 \leq k \leq K \) that achieves the optimal point-to-point distortion for each receiver. Assume \( p_{S_{[1:K]}|S}^* \) is the backward Gaussian broadcast channel consisting of subchannels \( p_{S_{[1:K]}|S} \), \( 1 \leq k \leq K \). Since any backward Gaussian channel can be transformed into a forward Gaussian channel with the same probability distribution, assume \( p_{V_{[1:K]}|U} \) is the forward Gaussian broadcast channel of \( p_{S_{[1:K]}|S} \). Then for Gaussian source broadcast, the outer bound is in form of the comparison of capacity regions for two Gaussian broadcast channels \( p_{V_{[1:K]}|U} \) and \( p_{V_{[1:K]}|X} \). Note that \( p_{V_{[1:K]}|X} \) and \( p_{V_{[1:K]}|U} \) have different bandwidth (the bandwidth ratio is \( b \)) but the same point-to-point capacity for each receiver. It can be proved that the Gaussian broadcast capacity region shrinks as the bandwidth increases. Additionally, \( \mathcal{C}_{DBC} \left( p_{V_{[1:K]}|U} \right) = \mathcal{C}_{DBC} \left( p_{V_{[1:K]}|X} \right) \) when bandwidth matched. Hence \( \mathcal{C}_{DBC} \left( p_{V_{[1:K]}|U} \right) \subseteq \mathcal{C}_{DBC} \left( p_{V_{[1:K]}|X} \right) \) always holds for bandwidth compression case. This is the reason why the outer bound in [9] is nontrivial only for bandwidth expansion. The details can be found in [10].

The bounds in Corollary 1 and Theorem 4 are illustrated in Fig. 3.

D. Hamming Binary Source Broadcast

Consider sending binary source \( S \sim \text{Bern} \left( \frac{1}{2} \right) \) with Hamming distortion measure \( d_k(s, \hat{s}) = d(s, \hat{s}) \triangleq 0 \), if \( s = \hat{s}; 1 \), otherwise, over binary broadcast channel \( Y_k = X \oplus W_k, 1 \leq k \leq K \) with \( W_k \sim \text{Bern} \left( p_k \right) \), \( \frac{1}{2} \geq p_1 \geq p_2 \geq \cdots \geq p_K \geq 0 \). Assume bandwidth mismatch factor is \( b \).
Fig. 3. Distortion bounds for sending Gaussian source over Gaussian broadcast channel with $b = 2, N_S = 1, P = 50, N_1 = 10, N_2 = 1$. Outer Bounds 1 and 2 and Inner Bounds 1 and 2 respectively correspond to the outer bound in Theorem 4, the outer bound in Theorem 8, the inner bound in Corollary 1, and the inner bound achieved by Wyner-Ziv separate coding (uncoded systematic code) [17, Lem. 3]. Trivial Outer Bound corresponds to the trivial outer bound (6). Besides, Outer Bound 2 and Inner Bound 2 could be considered as the outer bound and inner bound for Wyner-Ziv source broadcast problem with $b = 1, \beta_1 = \frac{N_S N_1}{P + N_1}, \beta_2 = \frac{N_S N_2}{P + N_2}$, and in this case Trivial Outer Bound corresponds to the Wyner-Ziv outer bound (55).

We first consider the inner bound. For bandwidth expansion ($b > 1$), as a special case of hybrid coding systematic source-channel coding (Uncoded Systematic Coding) has been first investigated in [6]. For any point-to-point lossless communication system, such systematic coding does not loss the optimality; however, for some lossy cases such as Hamming binary source communication, it is not optimal any more. To retain the optimality, we can first quantize the source $S$, and then transmit the quantized signal using Uncoded Systematic Coding. The performance of this code could be obtained directly from Theorem 3.

Specifically, let $U_2 = S \oplus E_2, U_1 = U_2 \oplus E_1$ with $E_2 \sim \text{Bern}(D_2), E_1 \sim \text{Bern}(d_1)$. Let $V_2 = (U_2, X^{b-1}), V_1 = (U_1, X_1^{b-1}), X_1^{b-1} = X^{b-1} \oplus B^{b-1}$, where $X_1^{b-1}$ and $X^{b-1}$ are independent of $U_2$ and $U_1$, and $X^{b-1}$ and $B^{b-1}$ follow $b - 1$ dimensional $\text{Bern}(\frac{1}{2})$ and $\text{Bern}(\theta)$, respectively. Let $x^b(v_2, s) = (u_2, x_2^{b-1}), \hat{s}_2(v_2, y_2^b) = u_2$ and $\hat{s}_1(v_1, y_1^b) = u_1$ if $d_1 < p_1; y_1$, otherwise. Substitute these variables and functions into the inner bound $R_D^{(i)}$, in Theorem 3, then we get the following corollary.

**Corollary 2** (Coded Systematic Coding). For transmitting binary source $S$ with Hamming distortion measure over
2-user binary broadcast channel with bandwidth mismatch factor $b$,
\[
\mathcal{R} \supseteq \mathcal{R}^{(i)}_{CSC} \triangleq \text{convex hull}\{(D_1, D_2) : 0 \leq \theta_d, d_1 \leq \frac{1}{2}\}
\]
\[
D_1 \geq \min\{d_1 \ast D_2, p_1 \ast D_2\},
\]
\[
r_1 = 1 - H_2(d_1 \ast p_1) + (b - 1) [1 - H_2(\theta \ast p_1)],
\]
\[
r_2 = H_2(d_1 \ast p_2) - H_2(p_2) + (b - 1) [H_2(\theta \ast p_2) - H_2(p_2)],
\]
\[
1 - H_2(d_1 \ast D_2) \leq r_1,
\]
\[
1 - H_2(D_2) \leq r_1 + r_2\}
\]
(48)

where $\ast$ denotes the binary convolution, i.e.,
\[
x \ast y = (1 - x)y + x(1 - y),
\]
(49)

and $H_2$ denotes the binary entropy function, i.e.,
\[
H_2(p) = -p \log p - (1 - p) \log(1 - p).
\]
(50)

Remark 3. Coded Systematic Coding without timesharing does not always lead to a convex distortion region, hence a timesharing mechanism is needed to improve performance and achieve $\mathcal{R}^{(i)}_{CSC}$. It is equivalent to adding a timesharing variable $Q$ into $V_2$ and $V_1$, before substitute them into the inner bound $\mathcal{R}^{(i)}_{DBC}$. Besides, note that unlike Uncoded Systematic Coding, the Coded Systematic Coding could always achieve the optimal distortion for at least one of the receivers. Moreover, unlike separate coding the Coded Systematic Coding could weaken the cliff effect, and result in slope-cliff effect.

In addition, the outer bound in Theorem 3 reduces to the following outer bound for Hamming binary source broadcast problem. This outer bound was first given in [11, Equation (41)] for 2-user case. The proof is similar to that of [11, Equation (41)], hence it is omitted here.

Theorem 5. For transmitting binary source $S$ with Hamming distortion measure over $K$-user binary broadcast channel with bandwidth mismatch factor $b$,
\[
\mathcal{R} \subseteq \mathcal{R}^{(o)}_{DBC} \triangleq \left\{D_{[1:K]} : \text{For any variables } \frac{1}{2} = \tau_0 \geq \tau_1 \geq \tau_2 \geq \cdots \geq \tau_K = 0,
\]
\[
\frac{1}{b} (H_2(\tau_{k-1} \ast D_k) - H_2(\tau_k \ast D_k) : k \in [1 : K]) \in \mathcal{C}_{BBC}\right\},
\]
(51)

where $\mathcal{C}_{BBC}$ denotes the capacity of binary broadcast channel given by
\[
\mathcal{C}_{BBC} = \left\{R_{[1:K]} : \text{There exist some variables } \frac{1}{2} = \theta_0 \geq \theta_1 \geq \theta_2 \geq \cdots \geq \theta_K = 0 \text{ such that}
\]
\[
0 \leq R_k \leq H_2(\theta_{k-1} \ast p_k) - H_2(\theta_k \ast p_k), 1 \leq k \leq K\right\}.
\]
(52)

The bounds in Corollary 2 and Theorem 5 are illustrated in Fig. 4.
Fig. 4. Distortion bounds for sending binary source over binary broadcast channel with \( b = 2, p_1 = 0.18, p_2 = 0.12 \). Outer Bounds 1 and 2 respectively correspond to the outer bound in Theorem 5 and the outer bound in Theorem 9. Separate Coding, Uncoded Systematic Coding and Coded Systematic Coding respectively correspond to the separate scheme combining successive-refinement code [16, Example 13.3] with superposition code [16, Example 5.3], the inner bound in Corollary 3, and the inner bound in Corollary 2. Trivial Outer Bounds 1 and 2 correspond to the trivial outer bound (6) and the Wyner-Ziv outer bound (55), respectively. Besides, Trivial Outer Bound 2, Outer Bound 2 and Uncoded Systematic Coding could be considered as the outer bounds and inner bound for Wyner-Ziv source broadcast problem with \( b = 1, \beta_1 = p_1, \beta_2 = p_2 \).

Fig. 5. Wyner-Ziv source broadcast system: broadcast communication system with side information at decoders.

IV. WYNER-ZIV SOURCE BROADCAST: SOURCE BROADCAST WITH SIDE INFORMATION

We now extend the problem by allowing decoders to access side information correlated with the source. As depicted in Fig. 5, receiver \( k \) observes memoryless side information \( Z^n_k \), and it produces source reconstruction \( \hat{S}^n_k \) from received signal \( Y^n_k \) and side information \( Z^n_k \).

**Definition 7.** An \( n \)-length Wyner-Ziv source-channel code is defined by the encoding function \( x^n : S^n \mapsto X^n \) and a sequence of decoding functions \( \hat{s}^n_k : Y^n_k \times Z^n_k \mapsto \hat{S}^n_k, 1 \leq k \leq K \).
Definition 8. If there exists a sequence of Wyner-Ziv source-channel codes satisfying

\[
\limsup_{n \to \infty} \mathbb{E}d_k \left( S^n, \hat{S}^n_k \right) \leq D_k,
\]

then we say that the distortion tuple \( D_{[1:K]} \) is achievable.

Definition 9. The admissible distortion region for Wyner-Ziv broadcast problem is defined as

\[
\mathcal{R}_{SI} \triangleq \left\{ D_{[1:K]} : D_{[1:K]} \text{ is achievable} \right\}.
\]

In addition, Shamai et al. [6, Thm. 2.1] showed that for transmitting source over point-to-point channel \( p_{Y|X} \) with side information \( Z_k \) available at decoder, the minimum achievable distortion satisfies \( R_{S|Z_k} (D_k) = C_k \), where \( R_{S|Z_k} (\cdot) \) is the Wyner-Ziv rate-distortion function of the source \( S \) given that the decoder observes \( Z_k \) [16]. Therefore, the optimal distortion is \( D_{SI,k}^* = R_{S|Z_k}^{-1}(C_k) \). Obviously,

\[
\mathcal{R}_{SI} \subseteq \mathcal{R}_{SI}^* \triangleq \left\{ D_{[1:K]} : D_k \geq D_{SI,k}^*, 1 \leq k \leq K \right\},
\]

where \( \mathcal{R}_{SI}^* \) is named Wyner-Ziv outer bound.

Besides, we also consider the communication system with bandwidth mismatch, whereby \( m \) samples of a DMS are transmitted through \( n \) uses of a DM-BC with \( l \) samples of side information available at each decoder. For simplicity, we let \( m = l \), and for this case, bandwidth mismatch factor is defined as \( b = \frac{n}{m} \).

A. Wyner-Ziv Source Broadcast

If consider \( Z_{[1:K]} \) to be transmitted from sender to the receivers over a virtual broadcast channel \( p_{Z_{[1:K]}|S} \), and define \( X' = (S,X) \) and \( Y_k' = (Z_k,Y_k), 1 \leq k \leq K \), then the Wyner-Ziv source broadcast problem is equivalent to the problem of sending \( p_S \) over \( p_{V_{[1:K]}|X'} \) with \( S \) restricted to be the input of subchannel \( p_{Z_{[1:K]}|S} \). Hence Wyner-Ziv source broadcast problem could be considered as the problem of (uncoded) systematic source-channel coding. If set \( x'(v_{[1:N]}, s) = (s, x'(v_{[1:N]}, s)) \), then from Theorem 1, we obtain the following inner bound for such systematic source-channel coding problem. It is therefore also an inner bound for the Wyner-Ziv source broadcast problem.

\[
\mathcal{R}_{SI}^{(i)} \triangleq \left\{ D_{[1:K]} : \text{There exist some pmf } p_{V_{[1:N]}|S}, \text{ vector } r_{[1:N]}, \right. \\
\text{ and functions } x'(v_{[1:N]}, s), \hat{s}_k (v_{D_k}, y_k, z_k), 1 \leq k \leq K \text{ such that } \\
\mathbb{E}d_k \left( S, \hat{S}_k \right) \leq D_k, 1 \leq k \leq K, \\
\sum_{j \in J} r_j > \sum_{j \in J} H(V_j|V_{A_j}) - H(V_J|S) \\
\text{ for all } J \subseteq [1 : N] \text{ such that } J \neq \emptyset \text{ and if } j \in J, \text{ then } A_j \subseteq J, \\
\sum_{j \in J} r_j < \sum_{j \in J^c} H(V_j|V_{A_j}) - H(V_{J^c}|Y_kZ_kV_J) \\
\text{ for all } 1 \leq k \leq K \text{ and for all } J \subseteq D_k \text{ such that } J^c \triangleq D_k \setminus J \neq \emptyset \text{ and if } j \in J, \text{ then } A_j \subseteq J \right\}.
\]
On the other hand, the distortion region of systematic source-channel coding problem belongs to that of the source broadcast problem without input constraint for subchannel \( p_{Z_{[1:K]}}|S \). Hence \( R_{SI} \subseteq R( p_S, p_{Y_{[1:K]}}|X' ) \),

where \( R( p_S, p_{Y_{[1:K]}}|X' ) \) denotes the admissible distortion region for sending \( S \) over broadcast channel \( p_{Y_{[1:K]}}|X' \).

Then from Theorem 1, we obtain the following outer bound on \( R( p_S, p_{Y_{[1:K]}}|X' ) \), which therefore is also an outer bound on \( R_{SI} \).

\[
R_{SI,1}^{(o)} = \left\{ D_{[1:K]} : \text{There exists some pmf } p_{V_{[1:K]}}|S \text{ and functions } \hat{s}_k(v_k, z_k), 1 \leq k \leq K \text{ such that} \right. \\
\left. \mathbb{E} d_k \left( S, \hat{S}_k \right) \leq D_k, 1 \leq k \leq K, \right. \\
\left. \text{and for any pmf } p_{U_{[1:L]}}|S, \text{ one can find } p_{S,X,U_{[1:L]},W_{[1:K]},W'_{[1:K]}} \right. \\
\left. \text{satisfying} \right. \\
\left. \sum_{i=1}^{m} I \left( V_{A_i}Z_{A_i}; U_{B_i}|U_{j=0}^{i-1}B_j \right) \leq \sum_{i=1}^{m} I \left( Y_{A_i}Z_{A_i}; \hat{U}_{B_i}\hat{W}_{A_{i+1}}|\hat{U}_{j=0}^{i-1}B_j \hat{W}_{A_i} \right), \right. \\
\left. \text{for any } m \geq 1, A_i \subseteq [1:K], B_i \subseteq [1:L], 0 \leq i \leq m, A_0, A_{m+1} \triangleq \emptyset, \right. \\
\left. \text{and } \hat{W}_{A_i} \triangleq \hat{W}_{A_i}, \text{if } i \text{ is odd; } W'_{A_i}, \text{otherwise, or } \hat{W}_{A_i} \triangleq W'_{A_i}, \text{if } i \text{ is odd; } W_{A_i}, \text{otherwise} \right\}. 
\]

(57)

In addition, regard \( (U_{[1:L]}, Z_{[1:K]}) \) as auxiliary random variables following \( p_{U_{[1:L]}|SP_{Z_{[1:K]}}|S} \), then following a similar proof to that of the outer bound in Theorem 1, we can achieve another outer bound on \( R_{SI} \).

\[
R_{SI,2}^{(o)} = \left\{ D_{[1:K]} : \text{There exists some pmf } p_{V_{[1:K]}}|S \text{ and functions } \hat{s}_k(v_k, z_k), 1 \leq k \leq K \text{ such that} \right. \\
\left. \mathbb{E} d_k \left( S, \hat{S}_k \right) \leq D_k, 1 \leq k \leq K, \right. \\
\left. \text{and for any pmf } p_{U_{[1:L]}}|S, \text{ one can find } p_{X,U_{[1:L]},Z_{[1:K]},W_{[1:K]},W'_{[1:K]}} \right. \\
\left. \text{satisfying} \right. \\
\left. \sum_{i=1}^{m} I \left( V_{A_i}Z_{A_{i+1}}|U_{j=0}^{i-1}B_j \right) \leq \sum_{i=1}^{m} I \left( Y_{A_i}Z_{A_{i+1}}|\hat{U}_{j=0}^{i-1}B_j \hat{Z}_{j=1} A_i \hat{W}_{A_i} \right), \right. \\
\left. \text{for any } m \geq 1, A_i \subseteq [1:K], B_i \subseteq [1:L], 0 \leq i \leq m, A_0, A_{m+1} \triangleq \emptyset, \right. \\
\left. \text{and } \hat{W}_{A_i} \triangleq \hat{W}_{A_i}, \text{if } i \text{ is odd; } W'_{A_i}, \text{otherwise, or } \hat{W}_{A_i} \triangleq W'_{A_i}, \text{if } i \text{ is odd; } W_{A_i}, \text{otherwise} \right\}. 
\]

(58)

Therefore, the following theorem holds. The proof is omitted.

**Theorem 6.** For transmitting source \( S \) over broadcast channel \( p_{Y_{[1:K]}}|X \) with side information \( Z_k \) at decoder \( k \),

\[
R_{SI}^{(o)} \subseteq R_{SI} \subseteq R_{SI,1}^{(o)} \cap R_{SI,2}^{(o)}. 
\]

(59)

**Remark 4.** Similar to Theorem 1, Theorem 6 could also be extended to Gaussian or any other well-behaved continuous-alphabet source-channel pair, the problem of broadcasting Wyner-Ziv correlated sources, or Wyner-Ziv source broadcast with channel input cost.


B. Wyner-Ziv Source Broadcast over Degraded Channel with Degraded Side Information

Theorem 6 can be used to derive the inner bound and outer bound for the case of degraded channel and degraded side information. Define

\[
\mathcal{R}^{(i)}_{\text{SI-D}} = \left\{ D_{[1:K]} : \text{There exists some pmf } p_{V_K|S} p_{V_{K-1}|V_K} \cdots p_{V_1|V_2}, \right. \\
\text{and functions } x(v_K, s), \hat{s}_k(v_k, y_k, z_k), 1 \leq k \leq K \text{ such that} \\
\mathbb{E}d_k(S, \hat{S}_k) \leq D_k, \\
(I(S; V_k) \leq \sum_{j=1}^{k} I(Y_j Z_j; V_j|V_{j-1}), 1 \leq k \leq K, \text{ where } V_0 \triangleq \emptyset \right\},
\]

and

\[
\mathcal{R}^{(o)}_{\text{SI-D}} = \left\{ D_{[1:K]} : \text{There exists some pmf } p_{V_K|S} p_{V_{K-1}|V_K} \cdots p_{V_1|V_2}, \right. \\
\text{and functions } \hat{s}_k(v_k, z_k), 1 \leq k \leq K \text{ such that} \\
\mathbb{E}d_k(S, \hat{S}_k) \leq D_k, \\
(I(V_k; S|Z_k) : k \in [1:K]) \in C_{DBC}'(p_{Y_1|K}|X), \\
\text{and for any pmf } p_{U_{K-1} | S} p_{U_{K-2} | U_{K-1}} \cdots p_{U_1 | U_2}, U_0 \triangleq \emptyset, U_K \triangleq S, \\
(I(V_k; U_k|U_{k-1} Z_k) : k \in [1:K]) \in C_{DBC}(p_{Y_1|K}|X) \text{ holds} \right\},
\]

where \( C_{DBC}'(p_{Y_1|K}|X) \) and \( C_{DBC}(p_{Y_1|K}|X) \) are given in (21) and (22), respectively. Then we have the following theorem. The proof is analogous to that of Theorem 3, and is therefore omitted.

**Theorem 7.** For transmitting source \( S \) over degraded broadcast channel \( p_{Y_1|K}|X \) (\( X \to Y_K \to Y_{K-1} \to \cdots \to Y_1 \)) with degraded side information \( Z_k \) (\( S \to Z_K \to Z_{K-1} \to \cdots \to Z_1 \)) at decoder \( k \),

\[
\mathcal{R}^{(i)}_{\text{SI-D}} \subseteq \mathcal{R}_{\text{SI}} \subseteq \mathcal{R}^{(o)}_{\text{SI-D}},
\]

C. Wyner-Ziv Gaussian Source Broadcast

Consider sending Gaussian source \( S \sim \mathcal{N}(0, N_S) \) with quadratic distortion measure \( d_k(s, \hat{s}) = d(s, \hat{s}) = (s - \hat{s})^2 \) over Gaussian broadcast channel \( Y_k = X + W_k, 1 \leq k \leq K \) with \( W_k \sim \mathcal{N}(0, N_k) \), \( N_1 \geq N_2 \geq \cdots \geq N_K \). Assume the side information \( Z_k \) observed by receiver \( k \) satisfies \( S = Z_k + B_k \) with independent Gaussian variables \( Z_k \sim \mathcal{N}(0, N_S - \beta_k) \) and \( B_k \sim \mathcal{N}(0, \beta_k) \). Assume bandwidth mismatch factor is \( b \). Then the inner bound in Theorem 7 could recover the existing results in the literature [17], [18], and the outer bound in Theorem 7 could be used to prove the following outer bound for Wyner-Ziv Gaussian source broadcast problem. The proof is given in Appendix E.
Theorem 8. For transmitting Gaussian source $S$ over Gaussian broadcast channel with degraded side information $Z_k (\beta_1 \geq \beta_2 \geq \cdots \geq \beta_K)$ at decoder $k$,

$$R_{SI} \subseteq R_{SI, D}^{(i)} \triangleq \left\{ D_{[1:K]} : \text{For any variables } + \infty = \tau_0 \geq \tau_1 \geq \cdots \geq \tau_K = 0, \frac{1}{b} \left( \frac{1}{2} \log \left( \frac{\beta_k + \tau_k}{D_k + \tau_k} \right) \right), k \in [1 : K] \right\} \in C_{GBC},$$

where $C_{GBC}$ denotes the capacity region of the Gaussian broadcast channel given in (47).

The bound in Theorem 8 is shown in Fig. 3.

D. Wyner-Ziv Binary Source Broadcast

Consider sending binary source $S \sim \text{Bern} \left( \frac{1}{2} \right)$ with Hamming distortion measure $d_k(s, \hat{s}) = d(s, \hat{s}) \triangleq 0$, if $s = \hat{s}; 1$, otherwise, over binary broadcast channel $Y_k = X \oplus W_k, 1 \leq k \leq K$ with $W_k \sim \text{Bern} (p_k), \frac{1}{2} \geq p_1 \geq p_2 \geq \cdots \geq p_K \geq 0$. Assume the side information $Z_k$ observed by receiver $k$ satisfies $S = Z_k \oplus B_k$ with independent variables $Z_k \sim \text{Bern} \left( \frac{1}{2} \right)$ and $B_k \sim \text{Bern} (\beta_k)$. Assume bandwidth mismatch factor is $b$.

Let $V_1 = (U_1, X_b^k), V_2 = (U_2, X_b^k), V_3 = \emptyset$, and $U_1$ and $U_2$ are independent of $X_b^k$ and $X_b$. $S, U_2$ and $U_1$ satisfy the distribution $p_{U_2|S} p_{U_1|U_2}$, where

$$p_{U_2|S} = \begin{pmatrix} 0 & 1 & 2 \\ 1 & q_2 \bar{\alpha}_2 & q_2 \alpha_2 \bar{\alpha}_2 \bar{\alpha}_2 \\ 0 & 1 & 2 \\ 2 & 0 & 0 & 1 \end{pmatrix},$$

$$p_{U_1|U_2} = \begin{pmatrix} 0 & q_1' \bar{\alpha}_1' & q_1' \alpha_1' \bar{\alpha}_1' \bar{\alpha}_1' \\ 1 & q_1' \alpha_1' & q_1' \bar{\alpha}_1' \bar{\alpha}_1' \bar{\alpha}_1' \end{pmatrix}.$$  

with $0 \leq q_2, q_1' \leq 1, 0 \leq \alpha_2, \alpha_1' \leq \frac{1}{2}$. $X^b$ and $X_b^k$ satisfy $X^b = X \oplus B^b$, $X^b \sim b$ dimensional Bern($\frac{1}{2}$), and $B^b \sim b$ dimensional Bern($\theta$) with $0 \leq \theta \leq \frac{1}{2}$. Denote $\alpha_1 = \alpha_2 * \alpha_1', q_1 = q_2 q_1'$, and set $x^b(v_2, s) = x^b$ and for $i = 1, 2$,

$$\hat{s}_i(v_i, y_i^b, z_i) = \begin{cases} z_i, & \text{if } \alpha_i \geq \beta_i \text{ or } \alpha_i < \beta_i, u_i = 2; \\ u_i, & \text{if } \alpha_i < \beta_i, u_i = 0, 1. \end{cases}$$  

Substitute these random variables and functions into $R_{SI}^{(i)}$ in Theorem 6, then we get the following performance (the hybrid coding reduces to a layered digital coding), which is tighter than that of the Layered Description Scheme (LDS) [17, Lem. 4].

Corollary 3 (Layered Digital Coding). For transmitting binary source $S$ with Hamming distortion measure over
2-user binary broadcast channel with side information \( Z_k \) at decoder \( k \),
\[
\mathcal{R}_{SI} \supseteq \mathcal{R}_{LDC}^{(i)} \triangleq \left\{(D_1, D_2) : 0 \leq q_1 \leq q_2 \leq 1, 0 \leq \alpha_2 \leq \alpha_1 \leq \frac{1}{2}, 0 \leq \theta \leq \frac{1}{2},
\begin{align*}
q_1 r(\alpha_1, \beta_1) &\leq b(1 - H_2(\theta * p_1)), \\
q_1 r(\alpha_1, \beta_2) &\leq b(1 - H_2(\theta * p_2)), \\
q_2 r(\alpha_2, \beta_2) &\leq b(1 - H_2(p_2)), \\
q_1 r(\alpha_1, \beta_1) + (q_2 r(\alpha_2, \beta_2) - q_1 r(\alpha_1, \beta_2)) &\leq b(1 - H_2(\theta * p_1)) + b(H_2(\theta * p_2) - H_2(p_2)), \\
D_i &\leq q_i \min \{\alpha_i, \beta_i\} + (1 - q_i) \beta_i, i = 1, 2 \right\},
\end{align*}
\]
where
\[
r(\alpha, \beta) = H_2(\alpha * \beta) - H_2(\alpha),
\]
\( * \) denotes the binary convolution given in (49), and \( H_2 \) denotes the binary entropy function given in (50).

In addition, the outer bound in Theorem 7 reduces to the following one for Wyner-Ziv binary case. The proof is given in Appendix F.

**Theorem 9.** For transmitting binary source \( S \) with Hamming distortion measure over \( K \)-user binary broadcast channel with degraded side information \( Z_k \) \((\frac{1}{2} \geq \beta_1 \geq \beta_2 \geq \cdots \geq \beta_K \geq 0) \) at decoder \( k \),
\[
\mathcal{R}_{SI} \subseteq \mathcal{R}_{SL,D}^{(o)} \triangleq \left\{D_{[1:K]} : \text{There exists some variables } 0 \leq \alpha_1, \alpha_2, \cdots, \alpha_K \leq \frac{1}{2}
\right. \\
\left. \text{such that } \alpha_k \leq D_k^i \triangleq \min \{D_k, \beta_k\}, 1 \leq k \leq K, \\
\right. \\
\text{and for any variables } \frac{1}{2} = \tau_0 \geq \tau_1 \geq \tau_2 \geq \cdots \geq \tau_K = 0,
\left. \frac{1}{b} \left(\eta_k (H_2(\beta_k * \tau_k) - H_2(\beta_k * \tau_{k-1})) - (H_4(\alpha_k, \beta_k, \tau_k) - H_4(\alpha_k, \beta_k, \tau_{k-1})) : k \in [1:K] \right) \right\} \subseteq \mathcal{C}_{BBC},
\]
where \( \mathcal{C}_{BBC} \) denotes the capacity region of the binary broadcast channel given in (52),
\[
\eta_k \triangleq \begin{cases} 
\frac{\beta_k - D_k^i}{\beta_k - \alpha_k}, & \text{if } \alpha_k < \beta_k, \\
0, & \text{otherwise},
\end{cases}
\]
\[
H_4(x, y, z) \triangleq -(x y z + x \overline{y} \overline{z}) \log (x y z + x \overline{y} \overline{z}) - (x \overline{y} \overline{z} + x y \overline{z}) \log (x \overline{y} \overline{z} + x y \overline{z})
\]
\[
- (x y \overline{z} + x \overline{y} z) \log (x y \overline{z} + x \overline{y} z) - (x \overline{y} z + x y \overline{z}) \log (x \overline{y} z + x y \overline{z}),
\]
and \( \overline{x} \triangleq 1 - x \).

The bounds in Corollary 3 and Theorem 9 are shown in Fig. 4.
V. CONCLUDING REMARKS

In this paper, we focused on the joint source-channel coding problem of sending a memoryless source over memoryless broadcast channel, and developed an outer bound and an inner bound for this problem. The inner bound is achieved by a unified hybrid coding scheme, and it can recover the best known performance of existing hybrid coding. Similarly, our outer bound can also recover the best known outer bound in the literature. Besides, we also extend the results to Wyner-Ziv source broadcast problem. All these bounds are also used to generate some new results, including the bounds on capacity region of general broadcast channel with common messages which respectively generalize Marton’s inner bound and Nair-El Gamal outer bound.

The inner bound achieved by proposed hybrid coding is established by using generalized Multivariate Covering Lemma and generalized Multivariate Packing Lemma, and the outer bound is derived by introducing auxiliary random variables (or remote sources) and exploiting Csiszár sum identity as in [12]. These lemmas and tools are expected to be exploited to derive more and stronger achievability and converse results for network information theory.

APPENDIX A

PROOF OF LEMMA 1

We follow similar steps to the proof of mutual covering lemma [15]. Let

\[ B = \left\{ m_{[1:k]} \in \prod_{i=1}^{k} [1 : 2^{n_{ri}}] : (U^n, V^n, V^n_{[1:k]}(m_{[1:k]})) \in T^{(n)}_{e} \right\}. \]  

(72)

Then we only need to show

\[ \lim_{n \to \infty} P(|B| = 0) = 0. \]

On the other hand,

\[ \lim_{n \to \infty} P(|B| = 0) = \lim_{n \to \infty} \sum_{u^n, v^n_0} p_{U^n, V^n_0} (u^n, v^n_0) P(|B| = 0 | u^n, v^n_0) \leq \lim_{n \to \infty} \sum_{(u^n, v^n_0) \in T^{(n)}_{e}} p_{U^n, V^n_0} (u^n, v^n_0) P(|B| = 0 | u^n, v^n_0) + \lim_{n \to \infty} P((u^n, v^n_0) \notin T^{(n)}_{e}) \]

(73)

To prove \( \lim_{n \to \infty} P(|B| = 0) = 0 \), it is sufficient to show \( \lim_{n \to \infty} P(|B| = 0 | u^n, v^n_0) = 0 \) for any \((u^n, v^n_0) \in T^{(n)}_{e}\). Using Chebyshev’s inequality, we can bound the probability as

\[ P(|B| = 0 | u^n, v^n_0) \leq P((|B| - E|B|)^2 \geq (E|B|)^2 | u^n, v^n_0) \leq \frac{\text{Var}(|B| | u^n, v^n_0)}{(E(|B| | u^n, v^n_0))^2}. \]

(76)

Next we prove the upper bound \( \frac{\text{Var}(|B| | u^n, v^n_0)}{(E(|B| | u^n, v^n_0))^2} \) tends to zero as \( n \to \infty \). Define

\[ E\left(m_{[1:k]} \right) \triangleq \begin{cases} 1, & \text{if } (u^n, v^n_0, V^n_{[1:k]}(m_{[1:k]})) \in T^{(n)}_{e} \\ 0, & \text{otherwise} \end{cases}. \]

(77)

May 22, 2016 DRAFT
for each $m_{[1:k]} \in \prod_{i=1}^{k}[1:2^{n_{r_i}}]$, then $|\mathcal{B}|$ can be expressed as

$$|\mathcal{B}| = \sum_{m_{[1:k]} \in \prod_{i=1}^{k}[1:2^{n_{r_i}}]} E(m_{[1:k]}).$$  \hspace{1cm} (78)

Denote

$$p_0 = \mathbb{P}\left((u^n, v^n_0, V^n_{[1:k]}(m_{[1:k]})) \in T^{(n)}_e | u^n, v^n_0\right),$$  \hspace{1cm} (79)

$$p_{\mathcal{I}} = \mathbb{P}\left((u^n, v^n_0, V^n_{[1:k]}(m_{[1:k]})) \in T^{(n)}_e, (u^n, v^n_0, V^n_{[1:k]}(m_{[1:k]}) \in T^{(n)}_e | u^n, v^n_0\right),$$  \hspace{1cm} (80)

for $m_{[1:k]} = 1$, and $m_{[1:k]} = 2$. Obviously, $p_{[1:k]} = p_0$. Then

$$\mathbb{E}(|\mathcal{B}| | u^n, v^n_0) = \sum_{m_{[1:k]}} \mathbb{P}\left((u^n, v^n_0, V^n_{[1:k]}(m_{[1:k]})) \in T^{(n)}_e | u^n, v^n_0\right) = 2^n \sum_{j=1}^{r_j} p_0,$$  \hspace{1cm} (81)

and

$$\mathbb{E}(|\mathcal{B}|^2 | u^n, v^n_0) = \sum_{\mathcal{I} \subseteq [1:k]} \sum_{m_{[1:k]}} \sum_{m_{I^c}^e : m_{I^c}^e \neq m_{I^c}} \mathbb{P}\left((u^n, v^n_0, V^n_{[1:k]}(m_{[1:k]})) \in T^{(n)}_e, (u^n, v^n_0, V^n_{[1:k]}(m_{I^c}) \in T^{(n)}_e | u^n, v^n_0\right)$$

$$= 2^n \sum_{j=1}^{r_j} p_0 + \sum_{\mathcal{I} \subseteq [1:k]} \sum_{m_{[1:k]}} \sum_{m_{I^c}^e : m_{I^c}^e \neq m_{I^c}} \mathbb{P}\left((u^n, v^n_0, V^n_{[1:k]}(m_{[1:k]})) \in T^{(n)}_e, (u^n, v^n_0, V^n_{[1:k]}(m_{I^c}) \in T^{(n)}_e | u^n, v^n_0\right).$$  \hspace{1cm} (82)

Define

$$\mathcal{J} \triangleq \{ \mathcal{J} \subseteq [1:k] : \text{if } j \in \mathcal{J}, \text{ then } A_j \subseteq \mathcal{J} \}. \hspace{1cm} (84)$$

Then any set $\mathcal{I} \subseteq [1:k]$ can transform into a $\mathcal{J}(\mathcal{I}) \in \mathcal{J}$ by removing all the elements $j$'s such that $A_j \not\subseteq \mathcal{I}$. According to generation of random codebook, we can observe that $p_{\mathcal{I}} = p_{\mathcal{J}(\mathcal{I})}$. Therefore,

$$\sum_{\mathcal{I} \subseteq [1:k]} \sum_{m_{[1:k]}} \sum_{m_{I^c}^e : m_{I^c}^e \neq m_{I^c}} \mathbb{P}\left((u^n, v^n_0, V^n_{[1:k]}(m_{[1:k]})) \in T^{(n)}_e, (u^n, v^n_0, V^n_{[1:k]}(m_{I^c}) \in T^{(n)}_e | u^n, v^n_0\right)$$

$$= \sum_{\mathcal{I} \subseteq [1:k]} \sum_{m_{[1:k]}} \sum_{m_{I^c}^e : m_{I^c}^e \neq m_{I^c}} p_{\mathcal{J}(\mathcal{I})} \hspace{1cm} (85)$$

$$\leq \sum_{\mathcal{I} \subseteq [1:k]} 2^n (\sum_{j=1}^{r_j} p_{\mathcal{J}(\mathcal{I})} \hspace{1cm} (86)$$

$$\leq \sum_{\mathcal{I} \subseteq [1:k]} 2^n (\sum_{j=1}^{r_j} p_{\mathcal{J}(\mathcal{I})} \hspace{1cm} (87)$$

$$\leq \sum_{\mathcal{J} \in \mathcal{J}} 2^n (\sum_{j=1}^{r_j} p_{\mathcal{J}(\mathcal{I})} \hspace{1cm} (88)$$

$$\leq \sum_{\mathcal{J} \in \mathcal{J}} 2^n (\sum_{j=1}^{r_j} p_{\mathcal{J}(\mathcal{I})} \hspace{1cm} (89)$$
where (87) follows from \(\mathcal{J}(\mathcal{I}) \subseteq \mathcal{I}\), (88) follows from that for each \(\mathcal{J} \subseteq \mathcal{I}\), there are at most \(2^k - |\mathcal{J}|\) of \(\mathcal{I}\)'s that could transform into \(\mathcal{J}\), and \(o(1)\) denotes a term that vanishes as \(n \to \infty\). Hence

\[
\text{Var}(B|u^n, v^n_0) \leq E[|B|^2|u^n, v^n_0]
\]

\[
\leq 2^n \sum_{j=1}^k r_j p_0 + \sum_{\mathcal{J} \subseteq \mathcal{I}} 2^n (\sum_{j=1}^k r_j + |\mathcal{J}| p_{J}) + o(1) p_{J}. \tag{90}
\]

Furthermore we have

\[
\frac{\text{Var}(B|u^n, v^n_0)}{(E[|B|^2|u^n, v^n_0]^2)} \leq \frac{2^n \sum_{j=1}^k r_j p_0 + \sum_{\mathcal{J} \subseteq \mathcal{I}} 2^n (\sum_{j=1}^k r_j + |\mathcal{J}| p_{J}) + o(1) p_{J}}{(2^n \sum_{j=1}^k r_j p_0)^2}
\]

\[
= 2^{-n} \sum_{j=1}^k r_j \frac{1}{p_0} + \sum_{\mathcal{J} \subseteq \mathcal{I}} 2^n (\sum_{j \in J} r_j + o(1)) \frac{p_{J}}{p_0}. \tag{91}
\]

According to generation process of random codebook, we can observe that

\[
p_0 = \sum_{v^n_{1:k} : (u^n, v^n_0, v^n_{1:k}) \in T_e^{(n)}} P(V^n_{1:k}|m_{1:k}) = v^n_{1:k}|u^n, v^n_0) \tag{92}
\]

\[
\geq 2^{-n(\sum_{j=1}^k H(V_j|V_{A_j}V_0) - H(V_{1:k}|U_{1:k}) + 2\delta(\epsilon))}, \tag{93}
\]

where (95) follows from that for any \((u^n, v^n_0, v^n_{1:k}) \in T_e^{(n)}\),

\[
P(V^n_{1:k}|m_{1:k}) = v^n_{1:k}|u^n, v^n_0) \geq 2^{-n(\sum_{j=1}^k H(V_j|V_{A_j}V_0) + \delta(\epsilon))}, \tag{94}
\]

and for any \((u^n, v^n_0) \in T_e^{(n)}\),

\[
\left| \{ v^n_{1:k} : (u^n, v^n_0, v^n_{1:k}) \in T_e^{(n)} \} \right| \geq 2^n(H(V_{1:k}|U_{1:k}) + \delta(\epsilon)). \tag{95}
\]

Similarly, we also can get

\[
p_{J} \leq 2^{-n(\sum_{j=1}^k H(V_j|V_{A_j}V_0) + \sum_{j \in J} H(V_j|V_{A_j}V_0) - H(V_{1:k}|U_{1:k}) - H(V_{J}||U_{1:k,V_0}V_j) - 4\delta(\epsilon))}. \tag{96}
\]

Substitute (95) and (98) into (93), then we have

\[
\text{Var}(B|u^n, v^n_0) \leq 2^{-n(\sum_{j=1}^k r_j - (\sum_{j=1}^k H(V_j|V_{A_j}V_0) - H(V_{1:k}|U_{1:k}) + 2\delta(\epsilon)))}
\]

\[
+ \sum_{\mathcal{J} \subseteq \mathcal{I}} 2^{-n(\sum_{j \in J} r_j - (\sum_{j \in J} H(V_j|V_{A_j}V_0) - H(V_{J}||U_{1:k,V_0}V_j) + 6\delta(\epsilon) + o(1))}). \tag{97}
\]

(99) tends to zero if

\[
\sum_{j=1}^k r_j > \sum_{j=1}^k H(V_j|V_{A_j}V_0) - H(V_{1:k}|U_{1:k}) + 2\delta(\epsilon) \tag{98}
\]

\[
\sum_{j \in J} r_j > \sum_{j \in J} H(V_j|V_{A_j}V_0) - H(V_{J}||U_{1:k,V_0}) + 6\delta(\epsilon) + o(1), \tag{99}
\]

i.e., \(\sum_{j \in J} r_j > \sum_{j \in J} H(V_j|V_{A_j}V_0) - H(V_{J}||U_{1:k,V_0}) + \delta'(\epsilon)\) for some \(\delta'(\epsilon)\) that tends to zero as \(\epsilon \to 0\). This completes the proof.
APPENDIX B
PROOF OF LEMMA 2

For any $J$ such that $J \neq \emptyset$ and if $j \in J$ then $A_j \subseteq J$,

$$\mathbb{P} \left( (U^n, V_0^n, V_{[1:k]}^n(m_{[1:k]})) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_{[1:k]} \right)$$

$$\leq \mathbb{P} \left( (U^n, V_0^n, V_J^n(m_J)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_J \right)$$

$$= \sum_{u^n, v_0^n} p_{U^n, V_0^n} (u^n, v_0^n) \mathbb{P} \left( (u^n, v_0^n, V_J^n(m_J)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_J|u^n, v_0^n \right)$$

$$\leq \sum_{u^n, v_0^n} p_{U^n, V_0^n} (u^n, v_0^n) \sum_{m_J} \mathbb{P} \left( (u^n, v_0^n, V_J^n(m_J)) \in \mathcal{T}_{\epsilon}^{(n)}|u^n, v_0^n \right). \tag{102}$$

Similar to (95), we can obtain that

$$\mathbb{P} \left( (u^n, v_0^n, V_J^n(m_J)) \in \mathcal{T}_{\epsilon}^{(n)}|u^n, v_0^n \right) \leq 2^{-n \left( \sum_{j \in \mathcal{J}} H(V_j|V_{A_j}V_0) - H(V_J|UV_0) - 2\delta(\epsilon) \right)}. \tag{103}$$

Substitute it into (104), then we have

$$\mathbb{P} \left( (U^n, V_0^n, V_J^n(m_J)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_{[1:k]} \right)$$

$$\leq 2^{n \left( \sum_{j \in \mathcal{J}} r_j - \left( \sum_{j \in \mathcal{J}} H(V_j|V_{A_j}V_0) - H(V_J|UV_0) - 2\delta(\epsilon) \right) \right)}. \tag{104}$$

(106) tends to zero if

$$\sum_{j \in J} r_j < \sum_{j \in J} H(V_j|V_{A_j}V_0) - H(V_J|UV_0) - 2\delta(\epsilon). \tag{107}$$

This completes the proof.

APPENDIX C
PROOF OF THEOREM 1

A. Inner Bound

Actually the inner bound can be seen as a corollary to [20, Thm. 1] by choosing proper network topology, transit probability and symbol-by-symbol functions. In the following, we provide a direct description of the proposed hybrid coding scheme and a direct proof for it.

**Codebook Generation:** Fix conditional pmf $p_{V_{[1:N]}|S}$, vector $r_{[1:N]}$, encoding function $x(v_{[1:N]}, s)$ and decoding
functions $\hat{s}_k(v_{D_k}, y_k)$ that satisfy

$$E_d \left(S, \hat{S} \right) \leq D_k, 1 \leq k \leq K,$$

(108)

$$\sum_{j \in J} r_j > \sum_{j \in J} H(V_j | V_{A_j}) - H(V_j | S)$$

for all $J \subseteq [1 : N]$ such that $J \neq \emptyset$ and if $j \in J$, then $A_j \subseteq J$,

(109)

$$\sum_{j \in J^c} r_j < \sum_{j \in J^c} H(V_j | V_{A_j}) - H(V_j | Y)$$

for all $1 \leq k \leq K$ and for all $J \subseteq D_k$ such that $J^c \triangleq D_k \backslash J \neq \emptyset$ and if $j \in J$, then $A_j \subseteq J$.

(110)

For each $j \in [1 : N]$ and each $m_{A_j} \in \prod_{i \in A_j} [1 : 2^{n_{r_j}}]$, randomly and independently generate a set of sequences $v_j^n(m_{A_j}, m_j), m_j \in [1 : 2^{n_{r_j}}]$, with each distributed according to $\prod_{i=1}^n p_{V_j | V_{A_j}}(v_j,i | v_{A_j,i}(m_{A_j}))$. The codebook

$$C = \left\{ v^n_{[1:N]} \left( m_{[1:N]} \right) : m_{[1:N]} \in \prod_{i=1}^N [1 : 2^{n_{r_i}}] \right\}.$$

(111)

is revealed to the encoder and all the decoders.

**Encoding:** We use joint typicality encoding. Given $s^n$, encoder finds the smallest index vector $m_{[1:N]}$ such that

$$\left( s^n, v^n_{[1:N]} \left( m_{[1:N]} \right) \right) \in T_e^{(n)}.$$

If there is no such index vector, let $m_{[1:N]} = 1$. Then the encoder transmits the signal

$$x_i = x \left( v^n_{[1:N],i} \left( m_{[1:N]} \right), s_i \right), 1 \leq i \leq n.$$

(112)

**Decoding:** We use joint typicality decoding. Let $\epsilon' > \epsilon$. Upon receiving signal $y^n_k$, the decoder of the receiver $k$ finds the smallest index vector $\hat{m}_{D_k}^{(k)}$ such that

$$(v_{D_k}^n(\hat{m}_{D_k}^{(k)}), y^n_k) \in T_{e'}^{(n)}.$$

(113)

If there is no such index vector, let $\hat{m}_{D_k}^{(k)} = 1$. The decoder reconstructs the source as

$$\hat{s}_{k,i} = \hat{s}_k(v_{D_k,i}(\hat{m}_{D_k}^{(k)}), y_{k,i}), 1 \leq i \leq n.$$

(114)

**Analysis of Expected Distortion:** We bound the distortion averaged over $S^n$, and the random choice of the codebook $C$. Define the “error” event

$$\mathcal{E} = \mathcal{E}_1 \cup \left( \bigcup_k \mathcal{E}_{2,k} \right) \cup \left( \bigcup_k \mathcal{E}_{3,k} \right),$$

(115)

where

$$\mathcal{E}_1 = \left\{ \left( s^n, V^n_{[1:N]} \left( m_{[1:N]} \right) \right) \notin T_e^{(n)} \text{ for all } m_{[1:N]} \right\},$$

(116)

$$\mathcal{E}_{2,k} = \left\{ \left( s^n, V^n_{[1:N]} \left( m_{[1:N]} \right), Y^n_k \right) \notin T_{e'}^{(n)} \right\},$$

(117)

$$\mathcal{E}_{3,k} = \left\{ \left( V^n_{D_k} (m'_{D_k}), Y^n_k \right) \in T_{e'}^{(n)} \text{ for some } m'_{D_k} \neq M_{D_k} \right\},$$

(118)

for $1 \leq k \leq K$. Using union bound, we have
\[
\mathbb{P}(E) \leq \mathbb{P}(E_1) + \sum_{k=1}^{K} \mathbb{P}(E_1^c \cap E_{2,k}) + \sum_{k=1}^{K} \mathbb{P}(E_{3,k}).
\] (119)

Now we claim that if (109) and (110) hold, then \(\mathbb{P}(E)\) tends to zero as \(n \to \infty\). Before proving it, we show that this claim implies the inner bound of Theorem 1.

Define
\[
E_{4,k} = \left\{ (S^n, V^n_{D_k}(\hat{M}_{I_k}(k)), Y^n_{k} \notin T'_k) \right\}.
\] (120)
then we have \(E^c \subseteq E^c_{4,k}\), i.e., \(E_{4,k} \subseteq E\). This implies that \(\mathbb{P}(E_{4,k}) \leq \mathbb{P}(E) \to 0\) as \(n \to \infty\). Then utilizing typical average lemma [16], we have
\[
\limsup_{n \to \infty} E_{d_k} \left( S^n, \hat{S}_k^n \right) = \limsup_{n \to \infty} \left( \mathbb{P}(E_{4,k}) \mathbb{E} \left[ d_k \left( S^n, \hat{S}_k^n \right) | E_{4,k} \right] + \mathbb{P}(E^c_{4,k}) \mathbb{E} \left[ d_k \left( S^n, \hat{S}_k^n \right) | E^c_{4,k} \right] \right)
\] (121)
\[
= \limsup_{n \to \infty} \mathbb{E} \left[ d_k \left( S^n, \hat{S}_k^n \right) | E^c_{4,k} \right] \leq (1 + \epsilon') E_{d_k} \left( S, \hat{S}_k \right)
\] (122)
\[
\leq (1 + \epsilon') D_k.
\] (123)

Therefore, the desired distortions are achieved for sufficiently small \(\epsilon'\).

Next we turn back to prove the claim above. Following from Multivariate Covering Lemma (Lemma 1), the first term of (119), \(\mathbb{P}(E_1)\), vanishes as \(n \to \infty\), and according to conditional typicality lemma [16, Sec. 3.7], the second item tends to zero as \(n \to \infty\).

Now we focus on the third term of (119), \(E_{3,k}\) can be written as
\[
E_{3,k} = \bigcup_{I \subseteq D_k} E_{3,k}^I,
\] (125)
where
\[
E_{3,k}^I = \left\{ (V^n_{D_k}(M_{I'}, m'_{I'}), Y^n_{k}) \in T'_k \text{ for some } m'_{I'}, m'_{I'}, \leftrightarrow M_{I'} \right\},
\] (126)
with \(I^c \triangleq D_k \\setminus I\). Using union bound we have
\[
\mathbb{P}(E_{3,k}) \leq \sum_{I \subseteq D_k} \mathbb{P}(E_{3,k}^I).
\] (127)

Each \(D_k\) has finite number of subsets, hence we only need to show for each \(I \subseteq D_k\), \(\mathbb{P}(E_{3,k}^I)\) vanishes as \(n \to \infty\). To show this, it is needed to analyze the correlation between coding index \(M_{[1:N]}\) and nonchosen codewords. Specifically, \(M_{[1:N]}\) depends on source sequence and the entire codebook, and hence standard packing lemma cannot be applied directly. This problem has been resolved by the technique developed in [15], [20].
where (129) follows from the union bound.

Define a sub-codebook as

\[
\mathcal{C}(m_x, m_{xe}) = \left\{ V^n_{y_k} (m_x, m_{xe}, \mathcal{D}_k) : \forall (m_{xe}', m_{xe}) \wedge m_{xe}' \right\}. \tag{130}
\]

Define another coding index as \( \bar{M}_{[1:N]} \) which is generated by performing the same coding process as \( M_{[1:N]} \) but on codebook \( \mathcal{C}(m_x, m_{xe}) \), i.e., given source sequence \( s^n \), encoder finds the smallest index vector \( \bar{m}_{[1:N]} \) such that \( \left( s^n, v^n_{[1:N]} (\bar{m}_{[1:N]}) \right) \in \mathcal{T}^{(n)} \); if there is no such index vector, let \( \bar{m}_{[1:N]} = 1 \). Then according to the generation process of \( M_{[1:N]} \) and \( \bar{M}_{[1:N]} \), we have if \( M_{[1:N]} = m_{[1:N]} \), then \( \bar{M}_{[1:N]} = m_{[1:N]} \). Now continuing with (129), we have

\[
P \left( M_{[1:N]} = m_{[1:N]}, Y^n_k = y^n_k, (V^n_{y_k} (m_x, m_{xe}), y^n_k) \in \mathcal{T}^{(n)} \right)
= \sum_{v^n_{[1:N]} \in \mathcal{T}^{(n)}} P \left( M_{[1:N]} = m_{[1:N]}, \bar{M}_{[1:N]} = m_{[1:N]}, \mathcal{C}(m_x, m_{xe}) = c, S^n = s^n, V^n_{D_k} (m_x, m_{xe}) = v^n_{D_k} \right)
\]

\[
\prod_{i=1}^{n} P_{Y_k|x} (y_{k,i} | x (v^n_{[1:N]}, i (m_{[1:N]}), s_i)) \right\{ (v^n_{D_k}, y^n_k) \in \mathcal{T}^{(n)} \} \right) \tag{131}
\]

\[
\prod_{i=1}^{n} P_{Y_k|x} (y_{k,i} | x (v^n_{[1:N]}, i (m_{[1:N]}), s_i)) \right\{ (v^n_{D_k}, y^n_k) \in \mathcal{T}^{(n)} \} \right) \tag{132}
\]

\[
\prod_{i=1}^{n} P_{Y_k|x} (y_{k,i} | x (v^n_{[1:N]}, i (m_{[1:N]}), s_i)) \right\{ (v^n_{D_k}, y^n_k) \in \mathcal{T}^{(n)} \} \right) \tag{133}
\]

\[
\prod_{i=1}^{n} P_{Y_k|x} (y_{k,i} | x (v^n_{[1:N]}, i (m_{[1:N]}), s_i)) \right\{ (v^n_{D_k}, y^n_k) \in \mathcal{T}^{(n)} \} \right) \tag{134}
\]

where \( c = \left\{ v^n_{[1:N]} (m_x, m_{xe}', m_{xe}') : \forall (m_{xe}', m_{xe}) \wedge m_{xe}' \right\} \), and (134) follows from the fact that \( V^n_{D_k} (m_x, m_{xe}) \rightarrow \mathcal{C}(m_x, m_{xe}) \rightarrow (S^n, \bar{M}_{[1:N]}) \) forms a Markov chain.

Define

\[
J \triangleq \{ J \subseteq \mathcal{D}_k : \text{if } j \in J, \text{then } A_j \subseteq J \}. \tag{135}
\]
Then any set $\mathcal{I} \subseteq \mathcal{D}_k$ can transform into a $\mathcal{J}$ ($\mathcal{I}$) $\in \mathcal{J}$ by removing all the elements $j$'s such that $A_j \not\subseteq \mathcal{I}$. Denote $\mathcal{J}^c \triangleq \mathcal{D}_k \setminus \mathcal{J}$. Then according to the generation process of the codebook, continuing with (134), we have

$$\sum_{v_{n_{\mathcal{D}_k}}} \mathbb{P} \left( V_{n_{\mathcal{D}_k}} (m_\mathcal{I}, m_\mathcal{I}'_c) = v_{n_{\mathcal{D}_k}} | C(m_\mathcal{I}, m_\mathcal{I}'_c) = c \right) 1 \left\{ (v_{n_{\mathcal{D}_k}}, y_k^n) \in T^{(n)}_c \right\}$$

$$= \sum_{v_{n_{\mathcal{J}^c}}} \mathbb{P} \left( V_{n_{\mathcal{J}^c}} (m_\mathcal{I}, m_\mathcal{I}'_c) = v_{n_{\mathcal{J}^c}} | C(m_\mathcal{I}, m_\mathcal{I}'_c) = c \right) 1 \left\{ (v_{n_{\mathcal{J}^c}}, y_k^n) \in T^{(n)}_c \right\}$$

$$= \sum_{v_{n_{\mathcal{J}^c}}} \prod_{j \in \mathcal{J}^c} \mathbb{P}_{v_{n_{\mathcal{A}_j}}} \left( v_{J_j, i} | v_{A_j, \mathcal{J}^c, j} (m_\mathcal{J}), v'_{A_j, \mathcal{J}^c, j} \right) 1 \left\{ (v_{n_{\mathcal{J}^c}}, y_k^n) \in T^{(n)}_c \right\}$$

$$\leq 2^n (H(V_{\mathcal{J}^c}|Y_k V_{\mathcal{J}^c}) - \sum_{j \in \mathcal{J}^c} H(V_j|V_{\mathcal{A}_j}) + (|\mathcal{J}^c| + 1) \delta(\epsilon')),$$

where $\delta(\epsilon')$ is a term that tends to zero as $\epsilon' \to 0$, and (138) follows from the fact that $\prod_{i=1}^n \mathbb{P}_{V_{n_{\mathcal{A}_j}}} \left( v_{j, i} | v_{A_j, i} \right) \leq 2^{-n(H(V_j|V_{A_j})-\delta(\epsilon'))}$ for any $(v_{j, n_{\mathcal{A}_j}}) \in T^{(n)}_c$ and $\left\{ (v_{n_{\mathcal{J}^c}}, y_k^n) \in T^{(n)}_c \right\}$ for any $(y_k^n, v_{n_{\mathcal{J}^c}})$.

Combining (129), (134) and (138) gives

$$\mathbb{P} \left( (V_{n_{\mathcal{D}_k}} (M_\mathcal{I}, m_\mathcal{I}'_c), Y_k^n) \in T^{(n)}_c \text{ for some } m_\mathcal{I}'_c, m_\mathcal{I}' \not\subset M_\mathcal{I} \right) \leq 2^n \left( \sum_{j \in \mathcal{J}^c} r_j - (\sum_{j \in \mathcal{J}^c} H(V_j|V_{\mathcal{A}_j}) - H(V_{\mathcal{J}^c}|Y_k V_{\mathcal{J}^c}) - (|\mathcal{J}^c| + 1) \delta(\epsilon')) \right).$$

Hence if $\sum_{j \in \mathcal{J}^c} r_j < \sum_{j \in \mathcal{J}^c} H(V_j|V_{\mathcal{A}_j}) - H(V_{\mathcal{J}^c}|Y_k V_{\mathcal{J}^c}) - (|\mathcal{J}^c| + 1) \delta(\epsilon')$ for all $\mathcal{J} \in \mathcal{J}$, then the third term of (119) tends to zero as $n \to \infty$. Letting $\epsilon'$ small enough, this completes the proof of the lower bound.

Besides, it is worth noting that although Multivariate Packing Lemma (Lemma 2) has not been employed directly in the proof, the derivation after (137) is essentially the same as that of Multivariate Packing Lemma.

### B. Outer Bound

For fixed $p_{u_{[1:L]}}|S$, we first introduce a set of auxiliary random variables $U_{[1:L]}^n$ that follow distribution $\prod_{i=1}^n p_{u_{[1:L]}}|s_i (u_{[1:L], |s_i)}}$. Hence the Markov chains $U_{[1:L]}^n \rightarrow S^n \rightarrow X^n \rightarrow Y_k^n \rightarrow S_k^n, 1 \leq k \leq K$ hold. We first derive a lower bound for

$$\sum_{i=1}^m I \left( Y_{A_i}^n; U_{B_i}^n | U_{[\cup_{j=1:B_i}] j}^n \right).$$
Hence (149) holds for $m$. We will show for any $m \geq 1$,

$$
\sum_{i=1}^{m} I \left( Y_{A_i}^{\eta}; U_B^m | U_{j=0}^{m-1} B_j \right)
= \sum_{i=1}^{m} \sum_{t=1}^{n} I \left( Y_{A_i}^{\eta}; U_{B_i,t} | U_{j=0}^{m-1} B_j U_B^{t-1} \right)
= \sum_{i=1}^{m} \sum_{t=1}^{n} H \left( U_{B_i,t} | U_{j=0}^{m-1} B_j U_B^{t-1} \right) - H \left( U_{B_i,t} | U_{j=0}^{m-1} B_j U_B^{t-1} Y_{A_i}^{\eta} \right)
= \sum_{i=1}^{m} \sum_{t=1}^{n} H \left( U_{B_i,t} | U_{j=0}^{m-1} B_j U_B^{t-1} \right) - H \left( U_{B_i,t} | U_{j=0}^{m-1} B_j U_B^{t-1} Y_{A_i}^{\eta} \right)
= \sum_{i=1}^{m} \sum_{t=1}^{n} I \left( U_{B_i,t}; Y_{A_i}^{\eta} | U_{j=0}^{m-1} B_j U_B^{t-1} \right)
= \sum_{i=1}^{m} \sum_{t=1}^{n} I \left( U_{B_i,t}; \hat{S}_{A_i,t} | U_{j=0}^{m-1} B_j \right)
= n \sum_{i=1}^{m} I \left( U_{B_i,Q}; \hat{S}_{A_i,Q} | U_{j=0}^{m-1} B_j, Q \right)
= n \sum_{i=1}^{m} I \left( U_{B_i,Q}; \hat{S}_{A_i,Q} | U_{j=0}^{m-1} B_j, Q \right)
\geq n \sum_{i=1}^{m} I \left( U_{B_i,Q}; \hat{S}_{A_i,Q} | U_{j=0}^{m-1} B_j, Q \right)
= n \sum_{i=1}^{m} I \left( U_{B_i}; \hat{S}_{A_i} | U_{j=0}^{m-1} B_j \right),
$$

where the time-sharing random variable $Q$ is defined to be uniformly distributed $[1 : n]$ and independent of all other random variables, and in (148), $U_l \triangleq U_{l,Q} \triangleq \hat{S}_{B,Q}, 1 \leq l \leq L, 1 \leq k \leq K$.

Now, we turn to upper-bounding $\sum_{i=1}^{m} I \left( Y_{A_i}^{\eta}; U_B^m | U_{j=0}^{m-1} B_j \right)$. We will show for any $m \geq 1$,

$$
\sum_{i=1}^{m} I \left( Y_{A_i}^{\eta}; U_B^m | U_{j=0}^{m-1} B_j \right) \leq \sum_{i=1}^{m} \sum_{t=1}^{n} I \left( Y_{A_i,t}; U_{B_i}^{\eta} | U_{j=0}^{m-1} B_j, \hat{Y}_{A_i}^{t-1} \hat{Y}_{A_{i+1}}^{t-1} \hat{Y}_{A_{i+2}}^{t-1} A_i \right) - \sum_{t=1}^{n} I \left( Y_{A_m,t}; \hat{Y}_{A_{m+1}}^{t-1} | U_{j=0}^{m} B_j, \hat{Y}_{A_m}^{t-1} \right)
$$

(149)

by induction method, where

$$
\hat{Y}_{A_i}^{t-1} \triangleq \begin{cases} 
Y_{A_i}^{t-1} & \text{if } i \text{ is odd;} \\
Y_{A_i,t+1}^{t} & \text{if } i \text{ is even.}
\end{cases}
$$

For $m = 1$,

$$
I \left( Y_{A_1}^{\eta}; U_B^1 | U_{B_0}^0 \right)
= \sum_{i=1}^{n} I \left( Y_{A_1}; U_B^1 | U_{B_0}^0, \hat{Y}_{A_1}^{t-1} \right)
= \sum_{i=1}^{n} I \left( Y_{A_1}; U_B^1 \hat{Y}_{A_2}^{t-1} | U_{B_0}^{t-1} Y_{A_1}^{t-1} \right) - I \left( Y_{A_1}; \hat{Y}_{A_2}^{t-1} | U_{j=0}^{m} B_j, \hat{Y}_{A_1}^{t-1} \right).
$$

Hence (149) holds for $m = 1$. 

May 22, 2016

DRAFT
Assume (149) holds for $m - 1$, then we have

$$
\sum_{i=1}^{m} I \left( Y_{A_i}^n; U_{B_i}^n | U_{\cup j=0}^{n-1} B_j \right)
= I \left( Y_{A_m}^n; U_{B_m}^n | U_{\cup j=0}^{n-1} B_j \right) + \sum_{i=1}^{m-1} \sum_{t=1}^{n} I \left( Y_{A_i,t}; U_{B_i}^{t-1} \tilde{Y}_{A_{i+1}}^{t+1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_i}^{t-1} \tilde{Y}_{A_{i+1}}^{t+1} \right)
- \sum_{t=1}^{n} I \left( Y_{A_{m-1},t}; \tilde{Y}_{A_{m}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right). \tag{153}
$$

Considering the first term of (153), we have

$$
I \left( Y_{A_m}^n; U_{B_m}^n | U_{\cup j=0}^{n-1} B_j \right)
= \sum_{t=1}^{n} I \left( Y_{A_m,t}; U_{B_m}^{t-1} \tilde{Y}_{A_{m-1}}^{t} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) \tag{154}
= \sum_{i=1}^{n} I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) - I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) \tag{155}
= \sum_{t=1}^{n} I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) + I \left( Y_{A_m,t}; U_{B_m}^{t-1} \tilde{Y}_{A_{m}}^{t-1} \tilde{Y}_{A_{m-1}}^{t-1} \right)
- I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) \tag{156}
= \sum_{t=1}^{n} I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) + I \left( Y_{A_m,t}; U_{B_m}^{t-1} \tilde{Y}_{A_{m}}^{t-1} \tilde{Y}_{A_{m-1}}^{t-1} \right) - I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) \tag{157}
= \sum_{t=1}^{n} I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) + I \left( Y_{A_m,t}; U_{B_m}^{t-1} \tilde{Y}_{A_{m}}^{t-1} \tilde{Y}_{A_{m-1}}^{t-1} \right)
- I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) \tag{158}
\leq \sum_{t=1}^{n} I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) + I \left( Y_{A_m,t}; U_{B_m}^{t-1} \tilde{Y}_{A_{m}}^{t-1} \tilde{Y}_{A_{m-1}}^{t-1} \right)
- I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right). \tag{159}
$$

Combine (153) and (159), and utilize the following identity

$$
\sum_{t=1}^{n} I \left( Y_{A_m,t}; \tilde{Y}_{A_{m-1}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m-1}}^{t-1} \right) = \sum_{t=1}^{n} I \left( Y_{A_{m-1},t}; \tilde{Y}_{A_{m}}^{t-1} | U_{\cup j=0}^{n-1} B_j \tilde{Y}_{A_{m}}^{t-1} \right), \tag{160}
$$

which follows from Csiszár sum identity [16, p. 25], then we have (149) holds for $m$. Hence (149) holds for any $m \geq 1$. 

From (149), we have
\[
\sum_{i=1}^{m} I \left( Y_{A_i}^n ; U_{B_i}^n | U_{j=0}^{i-1} B_j \right) \leq \sum_{i=1}^{m} \sum_{t=1}^{n} I \left( Y_{A_i,t}^n ; U_{B_i}^n | U_{j=0}^{i-1} B_j, \tilde{Y}_{A_i}^{t-1} \right) \ (161)
\]
\[
= n \sum_{i=1}^{m} I \left( Y_{A_i,Q}^n ; U_{B_i}^n, \tilde{Y}_{A_i}^{Q-1} | U_{j=0}^{i-1} B_j, \tilde{Y}_{A_i}^{Q-1} Q \right) \ (162)
\]
\[
= n \sum_{i=1}^{m} I \left( Y_{A_i}^n ; \hat{U}_{B_i}, \hat{W}_{A_i} | \hat{U}_{j=0}^{i-1} B_j, \hat{W}_{A_i} \right), \ (163)
\]
where the time-sharing random variable $Q$ is defined above, and $\hat{U}_i \triangleq U_{i}^n, Y_k \triangleq Y_{k,Q}, W_k \triangleq \left( Y_{k,Q-1}^n, Q \right), W'_k \triangleq \left( Y_{k,Q+1}^n, Q \right), 1 \leq l \leq L, 1 \leq k \leq K,$ and $\hat{W}_{A_i} \triangleq W_{A_i},$ if $i$ is odd; $W'_{A_i},$ otherwise.

If we redefine (150) as
\[
\tilde{Y}_{A_i}^{t-1} \triangleq \begin{cases} 
Y_{A_i,t+1}^n \text{ if } i \text{ is odd}; \\
Y_{A_i}^{t-1} \text{ if } i \text{ is even},
\end{cases}
\]
then (163) still holds for $\hat{W}_{A_i} \triangleq W'_{A_i},$ if $i$ is odd; $W'_{A_i},$ otherwise.

Combine bounds (148) and (163), then the outer bound $R^{(o)}$ in Theorem 1 holds.

**Appendix D**

**Proof of Theorem 3**

**A. Inner Bound**

For the inner bound $R^{(i)}$ in Theorem 1, retain all the random variables $V'_i$'s corresponding to the sets $G_i = [1 : K], [2 : K], \cdots, [K],\{K\},$ rename them and corresponding rates $r_i$'s to $V'_1, V'_2, \cdots, V'_K$ and $r'_1, r'_2, \cdots, r'_K,$ respectively, and set all the other random variables to empty, then $R^{(i)}$ reduces to

\[
R^{(i)} = \left\{ D_{[1 : K]} : \text{There exist some pmf } p_{V'_{[1 : K]}}, \text{ vector } r'_{[1 : K]}, \right. \\
\left. \begin{array}{l}
\text{and functions } x \left( v'_{[1 : K]} ; s \right), \hat{s}_k \left( v'_{[1 : k]} ; y_k \right), 1 \leq k \leq K \text{ such that} \\
\mathbb{E} d_k \left( S, \hat{S}_k \right) \leq D_k, 1 \leq k \leq K, \\
\sum_{j=1}^{k} r'_{j} > I \left( V'_{[1 : k]} ; S \right), 1 \leq k \leq K, \\
r'_{k} < I \left( V'_{k} ; Y_{k} | V'_{[1 : k-1]} \right), 1 \leq k \leq K, \right. \right\}, \ (165)
\]
and the coding scheme in the proof of Theorem 1 reduces to a superposition coding scheme. Define a set of random variables $V'_k \triangleq V'_{[1 : k]}, 1 \leq k \leq K.$ Substitute these into (165), then the inner bound in Theorem 3 is recovered.

**B. Outer Bound**

For the outer bound, we provide two proofs. The first follows from Theorem 1, and the second is a more simple and direct proof that does not utilize the Csiszár sum identity.
Proof method 1: Set \( L = K, A_i = [1 : i], 1 \leq i \leq m, B_0 = \emptyset \) and \( U_0 \triangleq \emptyset, U_K \triangleq S, p_{U_1}, \cdots, p_{U_{K-1}} | S = p_{U_{K-1}} | S p_{U_{K-2}} | U_{K-1} \cdots p_{U_1} | U_2 \). Substitute these into the outer bound \( \mathcal{R}^{(o)} \) in Theorem 1, and utilize the degradation of the channel, then we get

\[
\mathcal{R}^{(o)} = \left\{ D_{[1:K]} : \text{there exists some pmf } p_{S_{[1:K]} | S} \text{ such that}
\right. \\
\left. \mathbb{E} d_k \left( S, \hat{S}_k \right) \leq D_k, 1 \leq k \leq K,ight.
\]

and for any pmf \( p_{U_{[1:L]} | S} \), one can find \( p_{X, C_{[1:L]}, W_{[1:K]}, W'_{[1:K]}} \) satisfying

\[
\sum_{i=1}^{m} I \left( \hat{S}_{[1:i]}; U_i | U_{i-1} \right) \leq \sum_{i=1}^{m} I \left( Y_i; \hat{U}_{[1:i]} \tilde{W}_{[1:i+1]} | \tilde{U}_{[1:i]} \tilde{W}_{[1:i-1]} \right),
\]

for any \( m \geq 1 \), and \( \tilde{W}_{A_i} \triangleq W_{A_i}, \text{if } i \text{ is odd}; W'_{A_i}, \text{otherwise} \). (166)

Define \( V_i \triangleq \left( \hat{U}_{[1:i]}, \tilde{W}_{[1:i+1]}, \tilde{W}_{[1:i]} \right), 1 \leq i \leq m. \) Obviously, \( V_1 \rightarrow V_2 \rightarrow \cdots \rightarrow V_K \rightarrow X \rightarrow Y_k \) forms a Markov chain, hence

\[
(I \left( Y_k; V_k | V_{k-1} \right) : k \in [1 : K]) \in C_{DBC} \left( p_{Y_{[1:K]} | X} \right).
\]

According to (23) and (166),

\[
(I \left( \hat{S}_{[1:k]}; U_k | U_{k-1} \right) : k \in [1 : K]) \in C_{DBC} \left( p_{Y_{[1:K]} | X} \right).
\]

On the other hand, reset \( A_1 = [1 : i], B_1 = \{ K \}, m = 1, B_0 = \emptyset. \) Then we get

\[
I \left( \hat{S}_{[1:i]}; \hat{S} \right) \leq I \left( Y_i; \hat{U}_K \tilde{W}_{[1:i]} \right) \leq I \left( Y_i; \tilde{U}_K \tilde{W}_{[1:i]} \right), \text{i.e., } \left( I \left( \hat{S}_{[1:k]}; S \right) : k \in [1 : K] \right) \in C'_{DBC} \left( p_{Y_{[1:K]} | X} \right). \]

Combining it with (168) completes the proof.

Proof method 2: For fixed \( p_{U_{K-1}} | S p_{U_{K-2}} | U_{K-1} \cdots p_{U_1} | U_2 \), we first introduce a set of auxiliary random variables \( U_{[1:K-1]}^n \) that follow distribution \( \prod_{i=1}^{n} p_{U_{K-1}} | S (u_{K-1}, i | s_i) p_{U_{K-2}} | U_{K-1} (u_{K-2}, i | u_{K-1}, i) \cdots p_{U_1} | U_2 (u_1, i | u_2, i). \) Then \( U_1^n \rightarrow U_2^n \rightarrow \cdots \rightarrow U_{K-1}^n \rightarrow S^n \rightarrow X^n \rightarrow Y_K^n \rightarrow Y_{K-1}^n \rightarrow \cdots \rightarrow Y_1^n \) follows a Markov chain. We first derive a lower bound for \( I \left( Y_k^n; U_k | U_{k-1}^n \right). \)

May 22, 2016
\[ I \left( Y^n_k; U^n_k | U^n_{k-1} \right) \]
\[ = I \left( Y^n_{1:k}; U^n_k | U^n_{k-1} \right) \]
\[ = \sum_{i=1}^{n} I \left( Y^n_{1:k}; U^n_{k+1} | U^n_i \right) \]
\[ = \sum_{i=1}^{n} H \left( U_{k,i} | U^n_{k-1} Y^n_{1:k} \right) - H \left( U_{k,i} | U^n_{k-1} Y^n_{1:k} \right) \]
\[ = \sum_{i=1}^{n} I \left( U_{k,i}; U^n_{k-1} Y^n_{1:k} | U^n_k \right) \]
\[ \geq \sum_{i=1}^{n} I \left( U_{k,i}; \hat{S}_{1:k,i} | U^n_k \right) \]
\[ = n I \left( U_{k,Q}; \hat{S}_{1:k,Q} | U^n_k \right) \]
\[ = n I \left( U_{k,Q}; \hat{S}_{1:k,Q} | U^n_k \right) \]
\[ \geq n I \left( U_{k,Q}; \hat{S}_{1:k,Q} | U^n_k \right) \]
\[ = n I \left( U_{k,Q}; \hat{S}_{1:k,Q} | U^n_k \right) \]

where the time-sharing random variable \( Q \) is defined to be uniformly distributed \([1 : n]\) and independent of all other random variables, and in (178), \( \hat{S}_k \triangleq \hat{S}_{k,Q}, U_k \triangleq U_{k,Q}, 1 \leq k \leq K \).

Next, we turn to upper-bounding \( I \left( Y^n_k; U^n_k | U^n_{k-1} \right) \), and write the following:

\[ I \left( Y^n_k; U^n_k | U^n_{k-1} \right) \]
\[ = \sum_{i=1}^{n} I \left( Y_{k,i}; U^n_k | Y^n_{1:k} \right) \]
\[ \leq \sum_{i=1}^{n} I \left( Y_{k,i}; U^n_k Y^n_{i-1} | Y^n_{1:k} \right) \]
\[ = n \sum_{i=1}^{n} I \left( Y_{k,Q}; U^n_k Y^n_{Q-1} | U^n_{k-1} Y^n_{Q-1} \right) \]
\[ = n I \left( Y_{k,Q}; U^n_k Y^n_{Q-1} | U^n_{k-1} Y^n_{Q-1} \right) \]

where the time-sharing random variable \( Q \) is defined above, and \( Y^n_{Q-1} \triangleq X^n_{Q-1}, V^n_k \triangleq \left( U^n_k, Y^n_{k+1}, Q \right), Y_k \triangleq Y_{k,Q}, 1 \leq k \leq K \).

Obviously, \( V'_1 \rightarrow V'_2 \rightarrow \cdots \rightarrow V'_K \rightarrow X \rightarrow Y_k \) forms a Markov chain, hence

\[ \left( I \left( Y_k; V'_k | V'_{k-1} \right) : k \in [1 : K] \right) \in C_{DBC} \left( p_{Y_{[1:K]}|X} \right), \]
Combining this with (178) and (182), we have
\[
\left( I \left( U_k; S_{[1:k]}|U_{k-1} \right) : k \in [1:K] \right) \in C_{DBC} \left( p_{Y_{[1:K]}|X} \right).
\] (184)

On the other hand, it can be proved that \( nI \left( S_{[1:k]}; S \right) \leq I \left( S_{[1:k]}; S^\prime \right) \leq I \left( Y_k; X \right) \leq nI \left( Y_k; X \right) \), i.e.,
\[
\left( I \left( S_{[1:k]}; S \right) : k \in [1:K] \right) \in C_{DBC} \left( p_{Y_{[1:K]}|X} \right).
\] Combining it with (184) completes the proof.

**APPENDIX E**

**Proof of Theorem 8**

For Wyner-Ziv Gaussian broadcast with bandwidth mismatch case (bandwidth mismatch factor \( b \)), Theorem 7 states that if \( D_{[1:K]} \) is achievable, then there exists some pmf \( p_{V_K|S^{P_{V_{K-1}}|V_K} \cdots p_{V_1}|V_2} \) and functions \( \hat{s}_k (v_k, z_k) \), \( 1 \leq k \leq K \) such that
\[
\mathbb{E} d \left( S, \hat{S}_k \right) \leq D_k,
\] (185)
and for any pmf \( p_{U_{K-1}|S^{P_{U_{K-2}}|U_{K-1}} \cdots p_{U_1}|U_2} \),
\[
\frac{1}{b} \left( I \left( V_k; U_{k-1}Z_k \right) : k \in [1:K] \right) \in C_{GBC}
\] (186)
holds, where the capacity of Gaussian broadcast channel \( C_{GBC} \) is given in (47).

Choose \( U_{K-1} = S + E_{K-1} \) and \( U_k = U_{k+1} + E_k \), \( 1 \leq k \leq K - 2 \), where \( E_k \sim \mathcal{N} \left( 0, \tau_k \right) \) is independent of all the other random variables. Define \( E_k = \sum_{j=k}^{K-1} E_j' \sim \mathcal{N} \left( 0, \tau_k \right) \) with \( \tau_k = \sum_{j=k}^{K-1} \tau_j' \). Then
\[
I \left( V_1; U_1|Z_1 \right) \geq I \left( \hat{S}_1; U_1|Z_1 \right)
\] (187)
\[
= h \left( U_1|Z_1 \right) - h \left( U_1|\hat{S}_1Z_1 \right)
\] (188)
\[
= h \left( U_1|Z_1 \right) - h \left( U_1 - \hat{S}_1 | \hat{S}_1Z_1 \right)
\] (189)
\[
\geq h \left( U_1|Z_1 \right) - h \left( U_1 - \hat{S}_1 \right)
\] (190)
\[
\geq \frac{1}{2} \log \left( 2\pi e \left( \beta_1 + \tau_1 \right) \right) - \frac{1}{2} \log \left( 2\pi e \left( D_1 + \tau_1 \right) \right)
\] (191)
\[
= \frac{1}{2} \log \frac{\beta_1 + \tau_1}{D_1 + \tau_1},
\] (192)
where (191) follows from Gaussian distribution maximizes the differential entropy for a given second moment.

On the other hand,
\[
I \left( V_k; U_k|U_{k-1}Z_k \right) \geq I \left( \hat{S}_k; U_k|U_{k-1}Z_k \right)
\] (193)
\[
= I \left( \hat{S}_k; U_k|Z_k \right) - I \left( \hat{S}_k; U_{k-1}|Z_k \right)
\] (194)
\[
= h \left( U_k|Z_k \right) - h \left( U_{k-1}|Z_k \right) + h \left( U_{k-1}|Z_k\hat{S}_k \right) - h \left( U_k|Z_k\hat{S}_k \right).
\] (195)
The first two terms of (195)
\[
h \left( U_k|Z_k \right) - h \left( U_{k-1}|Z_k \right) = \frac{1}{2} \log \frac{\beta_k + \tau_k}{\beta_k + \tau_{k-1}}.
\] (196)
The last two terms of (195)
\begin{align}
  h \left( U_{k-1} | Z_k \hat{S}_k \right) - h \left( U_k | Z_k \hat{S}_k \right) &= h \left( U_{k-1} | Z_k \hat{S}_k \right) - h \left( U_k | Z_k \hat{S}_k E'_{k-1} \right) \\
  &= h \left( U_{k-1} | Z_k \hat{S}_k \right) - h \left( U_{k-1} | Z_k \hat{S}_k E'_{k-1} \right) \\
  &= I \left( U_{k-1}; E'_{k-1} | Z_k \hat{S}_k \right) \\
  &= h \left( E'_{k-1} \right) - h \left( E'_{k-1} | Z_k \hat{S}_k U_{k-1} \right) \\
  &= h \left( E'_{k-1} \right) - h \left( E'_{k-1} | Z_k, \hat{S}_k, U_{k-1} - \hat{S}_k \right) \\
  &\geq h \left( E'_{k-1} \right) - h \left( E'_{k-1} | U_{k-1} - \hat{S}_k \right) \\
  &= I \left( E'_{k-1} ; S - \hat{S}_k + E_k + E'_{k-1} \right) \\
  &\geq \frac{1}{2} \log \frac{D_k + \tau_{k-1}}{D_k + \tau_k},
\end{align}

where (204) is by applying the mutual information game result that Gaussian noise is the worst additive noise under a variance constraint [14, p. 298, Problem 9.21] and taking \( E'_{k-1} \) as channel input.

Combining (195), (196) and (204), we have

\[ I \left( V_k ; U_k | U_{k-1} Z_k \right) \geq \frac{1}{2} \log \frac{(\beta_k + \tau_k) (D_k + \tau_{k-1})}{(\beta_k + \tau_{k-1}) (D_k + \tau_k)}. \]  

(186), (192) and (205) imply Theorem 8 holds.

**APPENDIX F**

**PROOF OF THEOREM 9**

Observe that if there is no information transmitted over the channel, receiver \( k \) could produce a reconstruction within distortion \( \beta_k \). Hence we only need consider the case of \( D_{[1:K]} \) with

\[ D_k \leq \beta_k, 1 \leq k \leq K. \]

For Wyner-Ziv binary broadcast with bandwidth mismatch case (bandwidth mismatch factor \( b \)), Theorem 7 states that if \( D_{[1:K]} \) is achievable, then there exists some pmf \( p_{V_K | S P_{V_{K-1}} | V_K} \cdots p_{V_1 | V_2} \) and functions \( \hat{s}_k \left( v_k, z_k \right), 1 \leq k \leq K \) such that

\[ \mathbb{E} d \left( S, \hat{S}_k \right) = \mathbb{P} \left( \hat{s}_k \oplus S = 1 \right) \leq D_k, \]

and for any pmf \( p_{U_{K-1} | S P_{U_{K-2}} | U_{K-1} \cdots P_{U_1} | U_2}, \)

\[ \frac{1}{b} \left( I \left( V_k ; U_k | U_{k-1} Z_k \right) : k \in [1 : K] \right) \in \mathcal{C}_{BBC} \]

holds, where the capacity of binary broadcast channel \( \mathcal{C}_{BBC} \) is given in (52) [13].

Define the sets

\[ \mathcal{A}_k = \{ v_k : \hat{s}_k \left( v_k, 0 \right) = \hat{s}_k \left( v_k, 1 \right) \}, 1 \leq k \leq K, \]
so that their complements
\[ A_k^c = \{ v_k : \hat{s}_k (v_k, 0) \neq \hat{s}_k (v_k, 1) \}, \quad 1 \leq k \leq K. \] (210)

By hypothesis,
\[
\mathbb{E} d \left( S, \hat{S}_k \right) = \mathbb{P}(V_k \in A_k) \mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k \in A_k \right] + \mathbb{P}(V_k \in A_k^c) \mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k \in A_k^c \right] \\
\leq D_k.
\] (211)

We first show that
\[ \mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k \in A_k^c \right] \geq \beta_k. \] (213)

To do this, we write
\[
\mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k \in A_k^c \right] = \sum_{v_k \in A_k^c} \frac{\mathbb{P}(V_k = v_k)}{\mathbb{P}(V_k \in A_k^c)} \mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k = v_k \right].
\] (214)

If \( v_k \in A_k^c \) and \( \hat{s}_k (v_k, 0) = 0 \) then \( \hat{s}_k (v_k, 1) = 1 \). Therefore, for such \( v_k \),
\[
\mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k = v_k \right] = \mathbb{P} (\hat{Z}_k = 0, S = 1 | V_k = v_k) + \mathbb{P} (\hat{Z}_k = 1, S = 0 | V_k = v_k)
\]
\[
= \mathbb{P} (S = 1 | V_k = v_k) \mathbb{P} (\hat{Z}_k = 0 | S = 1) + \mathbb{P} (S = 0 | V_k = v_k) \mathbb{P} (\hat{Z}_k = 1 | S = 0)
\]
\[
= \beta_k,
\] (217)

where (216) follows that \( Z_k \to S \to V_k \) forms a Markov chain. If \( v_k \in A_k^c \) but \( \hat{s}_k (v_k, 0) = 1 \), then for such \( v_k \),
\[
\mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k = v_k \right] = 1 - \beta_k \geq \beta_k,
\] (218)
since \( \beta_k \leq \frac{1}{2} \). Therefore, (213) follows from (217) and (218).

Now we write
\[
\mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k \in A_k \right] = \sum_{v_k \in A_k} \frac{\mathbb{P}(V_k = v_k)}{\mathbb{P}(V_k \in A_k)} \mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k = v_k \right],
\] (219)

and define \( g_k (v_k) \triangleq \hat{s}_k (v_k, 0), \lambda_{v_k} \triangleq \frac{\mathbb{P}(V_k = v_k)}{\mathbb{P}(V_k \in A_k)}, \mu_k \triangleq \mathbb{P}(V_k \in A_k), \)
\[
d_{v_k} \triangleq \mathbb{E} \left[ d \left( S, \hat{S}_k \right) | V_k = v_k \right] = \mathbb{P} (S \neq g_k (v_k) | V_k = v_k),
\] (220)

then utilizing (212) and (213), we have
\[
d' \triangleq \mu_k \sum_{v_k \in A_k} \lambda_{v_k} d_{v_k} + (1 - \mu_k) \beta_k \leq D_k.
\] (221)

Next we will show
\[
I \left( V_k; U_k | Z_k \right) \leq \beta_k - D_k \left( H_2 (\beta_k * \tau_k) - (H_4 (\alpha_k, \beta_k, \tau_k) - H_2 (\alpha_k * \beta_k)) \right).
\] (222)

Choose \( U_{K-1} = S \oplus E'_{K-1} \) and \( U_k = U_{k+1} \oplus E'_k, 1 \leq k \leq K - 2 \), where \( E'_k \sim \text{Bern} (\tau'_k) \) is independent of all the other random variables. Define \( E_k = E'_{K-1} \oplus E'_{K-2} \oplus \cdots \oplus E'_k \sim \text{Bern} (\tau_k) \) with \( \tau_k = \tau'_{K-1} * \tau'_{K-2} * \cdots * \tau'_k \).
Then
\[ I (V_k; U_k | Z_k) = H (U_k | Z_k) - H (U_k | V_k, Z_k) \] (223)
\[ = H_2 (\beta_k \ast \tau_k) - \mu_k \sum_{v_k \in A_k} \lambda_{v_k} H (U_k | Z_k, V_k = v_k) \]
\[ - (1 - \mu_k) \sum_{v_k \in A_k} \frac{P (v_k)}{P (v_k \in A_k)} H (U_k | Z_k, V_k = v_k). \] (224)

For fixed \( v_k \), define a set of random variables \( (V'_k, S', U'_k, Z'_k) \sim 1 \{ v'_k = v_k \} p_{SU_k | Z_k} (s', t_k) \), then
\( H (U'_k | Z'_k | V'_k) = H (U_k | V_k) \) and \( H (Z'_k | V'_k) = H (Z_k | V_k) \). Since \( p_{SU_k | Z_k} \) satisfies
\[ p_{SU_k | Z_k} (s', t_k) = p_{S | V_k} (s' | t_k) p_{Z_k | S} (t_k | s') p_{U_k | S} (t_k | s'), \] (225)

it holds that \( Z'_k = S' \oplus B_k, U'_k = S' \oplus E_k \). Hence \( Z'_k \oplus U'_k = B_k \oplus E_k \).

For fixed \( v_k \), consider
\[ H (U_k | Z_k, V_k = v_k) = H (U_k | Z_k | V_k = v_k) - H (Z_k | V_k = v_k) \] (226)
\[ = H (U'_k | Z'_k | V'_k) - H (Z'_k | V'_k) \] (227)
\[ = H (U'_k | Z'_k | V'_k) \] (228)
\[ = H (U'_k \oplus Z'_k | Z'_k | V'_k) \] (229)
\[ = H (B_k \oplus E_k | Z'_k | V'_k) \] (230)
\[ \leq H (B_k \oplus E_k) \] (231)
\[ = H_2 (\beta_k \ast \tau_k). \] (232)

Combine (224) and (232), then it holds that
\[ I (V_k; U_k | Z_k) \geq H_2 (\beta_k \ast \tau_k) - \mu_k \sum_{v_k \in A_k} \lambda_{v_k} H (U_k | Z_k, V_k = v_k) - (1 - \mu_k) H_2 (\beta_k \ast \tau_k) \] (233)
\[ = \mu_k H_2 (\beta_k \ast \tau_k) - \mu_k \sum_{v_k \in A_k} \lambda_{v_k} H (U_k | Z_k, V_k = v_k). \] (234)

Now we consider the second term of (234).
\[ \sum_{v_k \in A_k} \lambda_{v_k} H (U_k | Z_k, V_k = v_k) = \sum_{v_k \in A_k} \lambda_{v_k} (H (U_k | Z_k | V_k = v_k) - H (Z_k | V_k = v_k)) \] (235)
\[ = \sum_{v_k \in A_k} \lambda_{v_k} (H_4 (d_{v_k}, \beta_k, \tau_k) - H_2 (d_{v_k} \ast \beta_k)). \] (236)
\[ = \sum_{v_k \in A_k} \lambda_{v_k} G_1 (d_{v_k}, \beta_k, \tau_k), \] (237)

where the function \( H_4 (x, y, z) \) is defined in (71) and
\[ G_1 (x, y, z) \triangleq H_4 (x, y, z) - H_2 (x \ast y). \] (238)

Equality (236) follows from calculating the entropies according to the definition.
Now we show that $G_1(x, y, z)$ is concave in $x$. To do this, we consider

$$\frac{\partial^2}{\partial x^2} G_1(x, y, z)$$

$$= - \frac{(yz - y\bar{z})^2}{xyz + xy\bar{z}} - \frac{(y\bar{z} - y\bar{z})^2}{xyz + xy\bar{z}} - \frac{(y\bar{z} - y\bar{z})^2}{xy\bar{z} + x\bar{y}z} + \frac{(y - \bar{y})^2}{xy\bar{z} + x\bar{y}z} + \frac{(y - \bar{y})^2}{xy + x\bar{y}}$$

$$= - \frac{(yz - y\bar{z})^2}{xyz + xy\bar{z}} + \frac{(y\bar{z} - y\bar{z})^2}{xy\bar{z} + x\bar{y}z} - \frac{(y - \bar{y})^2}{xy\bar{z} + x\bar{y}z} - \frac{(y - \bar{y})^2}{xy + x\bar{y}}$$

$$\leq 0,$$

where (241) follows from the following inequality

$$\frac{a_1^2}{b_1} + \frac{a_2^2}{b_2} = \frac{1}{b_1 + b_2} (b_1 + b_2) \left( \frac{a_1^2}{b_1} + \frac{a_2^2}{b_2} \right)$$

$$= \frac{1}{b_1 + b_2} \left( a_1^2 + a_2^2 + \frac{b_2 a_1^2}{b_1} + \frac{b_1 a_2^2}{b_2} \right)$$

$$\geq \frac{1}{b_1 + b_2} \left( a_1^2 + a_2^2 + 2a_1 a_2 \right)$$

$$= \frac{(a_1 + a_2)^2}{b_1 + b_2},$$

for $b_1, b_2 > 0$ and arbitrary real numbers $a_1, a_2$. (241) implies $G_1(x, y, z)$ is concave in $x$.

Then combining the concavity of $G_1(x, y, z)$ with (234) and (237), we have

$$I (V_k; U_k | Z_k) \geq \mu_k \left( H_2 (\beta_k * \tau_k) - G_1 \left( \sum_{v_k \in A_k} \lambda_{v_k} d_{v_k}, \beta_k, \tau_k \right) \right)$$

$$= \mu_k (H_2 (\beta_k * \tau_k) - G_1 (\alpha_k, \beta_k, \tau_k))$$

(246)

(247)

where

$$\alpha_k \triangleq \sum_{v_k \in A_k} \lambda_{v_k} d_{v_k}.$$  

(248)

From (221), $\alpha_k$ satisfies

$$\mu_k \alpha_k + (1 - \mu_k) \beta_k \leq D_k.$$  

(249)

Combine (249) with $D_k \leq \beta_k$ (i.e., (206)), then we have

$$0 \leq \alpha_k \leq D_k \leq \beta_k.$$  

(250)

Therefore,

$$I (V_k; U_k | Z_k) \geq \frac{\beta_k - D_k}{\beta_k - \alpha_k} \left( H_2 (\beta_k * \tau_k) - G_1 (\alpha_k, \beta_k, \tau_k) \right)$$

$$= \frac{\beta_k - D_k}{\beta_k - \alpha_k} \left( H_2 (\beta_k * \tau_k) - (H_k (\alpha_k, \beta_k, \tau_k) - H_2 (\alpha_k * \beta_k)) \right),$$

(251)

(252)

i.e., (222) holds.
Next we will show

\[
I (V_k; U_k | U_{k-1} Z_k) \geq \frac{\beta_k - D_k}{\beta_k - \alpha_k} (H_2 (\beta_k \ast \tau_k) - H_2 (\beta_k \ast \tau_{k-1}) - (H_4 (\alpha_k, \beta_k, \tau_k) - H_4 (\alpha_k, \beta_k, \tau_{k-1}))).
\]  

(253)

Consider

\[
I (V_k; U_k | U_{k-1} Z_k) = H (U_k | U_{k-1} Z_k) - H (U_k | U_{k-1} Z_k V_k)
\]

(254)

\[
= H (U_k | U_{k-1}) + H (U_k | Z_k) - H (U_k | U_{k-1} Z_k) - H (U_k | U_{k-1} Z_k V_k)
\]

(255)

\[
= H_2 (\tau'_{k-1}) + H_2 (\beta_k \ast \tau_k) - H_2 (\beta_k \ast \tau_{k-1}) - H (U_k | U_{k-1} Z_k V_k).
\]

(256)

Write the last term as

\[
H (U_k | U_{k-1} Z_k V_k)
\]

(257)

\[
= -(1 - \mu_k) \sum_{v_k \in A_k} \frac{\mathbb{P}(V_k = v_k)}{\mathbb{P}(V_k \in A_k)} H (U_k | U_{k-1}, Z_k, V_k = v_k) - \mu_k \sum_{u_k \in A_k} \lambda_{u_k} H (U_k | U_{k-1}, Z_k, V_k = v_k).
\]

For fixed \(v_k\), define \((V'_k, S', U'_k, U'_{k-1}, Z'_k) \sim 1 \{v'_k = v_k\} p_{SU_k U_{k-1} Z_k | V_k} (s', u'_k, u'_{k-1}, z'_k | v'_k)\). Since

\[
p_{SU_k U_{k-1} Z_k | V_k} (s', u'_k, u'_{k-1}, z'_k | v'_k) = p_{S | V_k} (s | v_k) p_{U_k | S} (u_k | s) p_{U_{k-1} | U_k} (u'_{k-1} | u'_k),
\]

(258)

we have \(Z'_k = S' \oplus B_k, U'_{k} = S' \oplus E_k, U'_{k-1} = U'_k \oplus E'_{k-1}\). Hence \(Z'_k \oplus U'_{k} = B_k \oplus E_k, Z'_k \oplus U'_{k-1} = B_k \oplus E_{k-1}\).

Similar to the derivation for \(H (U_k | U_{k-1}, V_k = v_k)\), we can write

\[
H (U_k | U_{k-1}, Z_k, V_k = v_k) = H (U'_k | U'_{k-1} Z'_k V'_k)
\]

(259)

\[
= H (U'_k \oplus Z'_k | U'_{k-1} \oplus Z'_k, Z'_k, V'_k)
\]

(260)

\[
\leq H (U'_k \oplus Z'_k | U'_{k-1} \oplus Z'_k)
\]

(261)

\[
= H (B_k \oplus E_k B_k \oplus E_{k-1})
\]

(262)

\[
= H (B_k \oplus E_k) + H (B_k \oplus E_{k-1} | B_k \oplus E_k) - H (B_k \oplus E_{k-1})
\]

(263)

\[
= H_2 (\tau'_{k-1}) + H_2 (\beta_k \ast \tau_k) - H_2 (\beta_k \ast \tau_{k-1}).
\]

(264)

Combine (256), (257) and (264), then we have

\[
I (V_k; U_k | U_{k-1} Z_k) \geq \mu_k (H_2 (\tau'_{k-1}) + H_2 (\beta_k \ast \tau_k) - H_2 (\beta_k \ast \tau_{k-1})) - \mu_k \sum_{v_k \in A_k} \lambda_{v_k} H (U_k | U_{k-1}, Z_k, V_k = v_k).
\]

(265)
Consider the last term of (265),

$$\sum_{v_k \in A_k} \lambda_{v_k} H (U_k | U_{k-1}, Z_k, V_k = v_k)$$

$$= \sum_{v_k \in A_k} \lambda_{v_k} (H (U_k | Z_k, V_k = v_k) + H (U_{k-1} | U_k, Z_k, V_k = v_k) - H (U_{k-1} | Z_k, V_k = v_k))$$

(266)

$$= \sum_{v_k \in A_k} \lambda_{v_k} (H (U_k | Z_k, V_k = v_k) + H_2 (\tau'_{k-1}) - H (U_{k-1}, Z_k | V_k = v_k))$$

(267)

$$= H_2 (\tau'_{k-1}) + \sum_{v_k \in A_k} \lambda_{v_k} (H_4 (d_{v_k}, \beta_k, \tau_k) - H_4 (d_{v_k}, \beta_k, \tau_k-1))$$

(268)

$$= H_2 (\tau'_{k-1}) + \sum_{v_k \in A_k} \lambda_{v_k} G_2 (d_{v_k}, \beta_k, \tau_k, \tau_k-1),$$

(269)

where (268) is by directly calculating the entropies according to the definition, and

$$G_2 (x, y, z) \triangleq H_4 (x, y, z) - H_4 (x, y, t).$$

(270)

Note that function $G_1 (x, y, z)$ is a special case of function $G_2 (x, y, z, t)$ given $t = \frac{1}{2}$, i.e.,

$$G_1 (x, y, z) = G_2 \left(x, y, z, \frac{1}{2}\right).$$

(271)

Now we show that $G_2 (x, y, z, t)$ is concave in $x$ when $0 \leq z \leq t \leq \frac{1}{2}$, which generalizes the concavity of $G_1 (x, y, z)$. To do this, we consider

$$\frac{\partial^2}{\partial x^2} G_2 (x, y, z, t)$$

$$= \frac{(y - y')^2}{xy + xy'} + \frac{(y - y')^2}{xy + xy'} - \frac{(y - y')^2}{xy + xy'}$$

(272)

and

$$\frac{\partial}{\partial t} \left( \frac{\partial^2}{\partial x^2} G_2 (x, y, z, t) \right) = \frac{\partial}{\partial t} \left( \frac{(y - y')^2}{xy + xy'} + \frac{(y - y')^2}{xy + xy'} \right) + \frac{\partial}{\partial t} \left( \frac{(y - y')^2}{xy + xy'} + \frac{(y - y')^2}{xy + xy'} \right)$$

(273)

$$= -y^2 \cdot y' \cdot (x + xy') \cdot (1 - 2t)$$

(274)

Hence for $0 \leq t \leq \frac{1}{2}$,

$$\frac{\partial}{\partial t} \left( \frac{\partial^2}{\partial x^2} G_2 (x, y, z, t) \right) \leq 0,$$

(275)

i.e., $\frac{\partial^2}{\partial x^2} G_2 (x, y, z, t)$ is decreasing in $t$. Then we have for $0 \leq z \leq t \leq \frac{1}{2}$,

$$\frac{\partial^2}{\partial x^2} G_2 (x, y, z, t) \leq \frac{\partial^2}{\partial x^2} G_2 (x, y, z, z) = 0,$$

(276)

It implies $G_2 (x, y, z, t)$ is concave in $x$ when $0 \leq z \leq t \leq \frac{1}{2}$.

Combining (265) and (269), and utilizing the concavity of $G_2 (x, y, z, t)$, we have

$$I (V_k; U_k | U_{k-1}, Z_k) \geq \mu_k \left( H_2 (\beta_k \ast \tau_k) - H_2 (\beta_k \ast \tau_k-1) - G_2 \left( \sum_{v_k \in A_k} \lambda_{v_k} d_{v_k}, \beta_k, \tau_k, \tau_k-1 \right) \right)$$

(277)

$$= \mu_k \left( H_2 (\beta_k \ast \tau_k) - H_2 (\beta_k \ast \tau_k-1) - G_2 (\alpha_k, \beta_k, \tau_k, \tau_k-1) \right)$$

(278)
where $\alpha_k$ is given by (248) and satisfies (249) and (250). Therefore,

$$ I (V_k; U_k | U_{k-1} Z_k) \geq \frac{\beta_k - D_k}{\beta_k - \alpha_k} \left( H_2 (\beta_k \star \tau_k) - H_2 (\beta_k \star \tau_{k-1}) - G_2 (\alpha_k, \beta_k, \tau_k, \tau_{k-1}) \right) $$

$$ = \frac{\beta_k - D_k}{\beta_k - \alpha_k} \left( H_2 (\beta_k \star \tau_k) - H_2 (\beta_k \star \tau_{k-1}) - (H_4 (\alpha_k, \beta_k, \tau_k) - H_4 (\alpha_k, \beta_k, \tau_{k-1})) \right), $$

i.e., (253) holds.

Combining (208), (222) and (253) gives Theorem 9.

**REFERENCES**


