Lossless and Lossy Broadcast System Source Codes: Theoretical Limits, Optimal Design, and Empirical Performance *

Qian Zhao   Michelle Effros

Abstract

Broadcast systems are a class of networks where one system node (transmitter) simultaneously sends both common and independent information to multiple nodes (receivers) in the system. Compressing the messages transmitted in such systems using traditional (single-transmitter, single-receiver) techniques requires use of a collection of independent source codes, one for each message sent through the system. The result of this approach is a system with multiple independent encoders at the transmitter and multiple independent decoders at each receiver. An alternative approach is to design a single joint encoder at the transmitter and a single joint decoder at each receiver. We call the resulting code a "broadcast system source code". This paper treats the theory and practice of optimal lossless and lossy (fixed- and variable-rate) broadcast system source codes. The results given include: theoretical limits for lossless source code performance on broadcast systems; an optimal lossless source code design algorithm; an optimal lossy source code design algorithm that generalizes the generalized Lloyd algorithm to broadcast systems; and experimental results for fixed- and variable-rate code performance in a two-receiver broadcast system.

I Introduction

A broadcast system is a network in which one "transmitter" simultaneously sends information to a collection of "receivers." The information sent by the transmitter may include both "common" information, intended for a collection of receivers, and "private" information, intended for only one receiver. Broadcast systems play a role in many network communication environments. For example, a base station sending information to a collection of hand-held units in a wireless communication network is a broadcast system; similarly, a video conferencing system may be modeled as a collection of broadcast networks, where each node broadcasts one user's voice and video messages to all other nodes in the network. Paging systems and computer

1

networks are also examples of network technologies where broadcast systems play a central role. In networks such as wireless communications systems, where resources such as power and bandwidth are critically limited, it is important to compress data maximally before transmission. Broadcast system source codes (BSSC), introduced in [1], are source codes explicitly designed for broadcast systems.

Consider a single-transmitter, M-receiver broadcast system. We label the receivers by $1, \ldots, M$ and denote the set $\{1, \ldots, M\}$ by $\mathcal{M}$. Let $\mathcal{S}$ be the set of all nonempty subsets of $\mathcal{M}$. Then each set $S \in \mathcal{S}$ describes a group of receivers to whom a particular message may be sent. For an M-receiver system, the cardinality of $\mathcal{S}$ is $2^M - 1$, and thus there can be up to $2^M - 1$ messages to be described. We assume an arbitrary but fixed ordering on $\mathcal{S}$; thus $\mathcal{S} = \{S(1), \ldots, S(2^M - 1)\}$. For any set $S \in \mathcal{S}$, $X_S[1], X_S[2], \ldots$ denotes the random source sequence to be described to all receivers in set S, and $\mathcal{X}_S$ denotes the corresponding source alphabet; thus for each $i \geq 1$, source symbol $X_S[i]$ may take on any value $x_S \in \mathcal{X}_S$. If $|S| > 1$, then X_S is the common information received by all of the receivers in S; if $|S| = 1$, then X_S is the private information received by the single receiver in S.

For any M-receiver broadcast system, a broadcast system source code contains one encoder and M decoders. In the most general case, the broadcast system transmitter uses the single encoder to jointly describe all sources X_S such that $S \in \mathcal{S}$. Each receiver $r \in \mathcal{M}$ uses a single decoder to jointly decode the descriptions of all sources X_S such that $r \in S$. The following notation is useful in making these ideas concrete.

This work treats block BSSCs. For each $S \in \mathcal{S}$, we block the source sequence $\{X_S[i]\}_{i=1}^{\infty}$ into vectors of length n and use $X_S^n \in \mathcal{X}_S^n$ to denote a single n-vector. The encoder jointly encodes all source vectors X_S^n with $S \in \mathcal{S}$. For any receiver $r \in \mathcal{M}$, the decoder at receiver r jointly decodes all messages X_S^n with $S \in \mathcal{S}_r = \{S \in \mathcal{S} : r \in S\}$. Now for any set $\mathcal{W} \subseteq \mathcal{S}$, let $\mathbf{X}_{\mathcal{W}}^n$ denote the vector $(X_S^n : S \in \mathcal{W})$ and $\mathcal{X}_{\mathcal{W}}^n = \prod_{S \in \mathcal{W}} \mathcal{X}_S^n$ denote the corresponding product alphabet; thus $\mathbf{X}_{\mathcal{W}}^n$ can take on any value $\mathbf{x}_{\mathcal{W}}^n \in \mathcal{X}_{\mathcal{W}}^n$, where the components of $\mathbf{X}_{\mathcal{W}}^n$ and its alphabet $\mathcal{X}_{\mathcal{W}}^n$ are ordered according to the fixed ordering on $\mathcal{S}$. Then $\mathbf{X}_{\mathcal{S}}^n \in \mathcal{X}_{\mathcal{S}}^n$ is the vector of sources described by the encoder and $\mathbf{X}_{\mathcal{S}_r}^n \in \mathcal{X}_{\mathcal{S}_r}^n$ is the vector of sources intended for receiver r. For notational simplicity, we abbreviate this notation as $\mathbf{X}^n = \mathbf{X}_{\mathcal{S}}^n$, $\mathcal{X}^n = \mathcal{X}_{\mathcal{S}}^n$, $\mathbf{X}_r^n = \mathbf{X}_{\mathcal{S}_r}^n$ and $\mathcal{X}_r^n = \mathcal{X}_{\mathcal{S}_r}^n$. Finally, for each $S \in \mathcal{S}$ and $r \in S$, let $\hat{\mathcal{X}}_{S,r}$ denote the reproduction alphabet used by receiver r in its reproduction of source X_S. Then for any $S \in \mathcal{S}$, we use $\hat{X}_{S,r}^n$ to denote receiver r's reproduction of source vector X_S^n and $\hat{\mathbf{X}}_r^n = (\hat{X}_{S,r}^n : S \in \mathcal{S}_r)$ to denote the full vector of reproductions at receiver r. Thus $\hat{X}_{S,r}^n$ and $\hat{\mathbf{X}}_r^n$ can respectively take on any value $\hat{x}_{S,r}^n \in \hat{\mathcal{X}}_{S,r}^n$ and $\hat{\mathbf{x}}_r^n \in \hat{\mathcal{X}}_r^n = \prod_{S \in \mathcal{S}_r} \hat{\mathcal{X}}_{S,r}^n$.

The system encoder $\mathbf{a} : \mathcal{X}^n \to \mathcal{C}$ maps the space of input source vectors $\mathcal{X}^n$ to the transmitter's product channel codebook $\mathcal{C} = \prod_{S \in \mathcal{S}} \mathcal{C}_S$, where for each $S \in \mathcal{S}$, the individual channel codebook $\mathcal{C}_S$ is a subset of the set $\{0, 1\}^*$ of finite binary strings. The encoder $\mathbf{a}$ comprises two parts, $\mathbf{a} = \gamma \circ \alpha$, where $\alpha : \mathcal{X}^n \to \mathcal{I}$, $\gamma : \mathcal{I} \to \mathcal{C}$, and $\circ$ denotes composition. The intermediate alphabet $\mathcal{I}$ is the index set of the channel codebook $\mathcal{C}$, where index set $\mathcal{I}$ is a product alphabet of form $\mathcal{I} = \prod_{S \in \mathcal{S}} \mathcal{I}_S$ and $\mathcal{I}_S$ is the index set associated with source X_S. The first component α maps each source vector $\mathbf{X}^n \in \mathcal{X}^n$ to a channel codeword index $i \in \mathcal{I}$. The function γ maps index i into

the corresponding channel codeword $c_i \in \mathcal{C}$. For each $r \in \mathcal{M}$, we use $\mathbf{t}_r : \mathcal{C} \to \mathcal{C}_r$ to denote the action of the channel seen by receiver r. Receiver r's decoder $\mathbf{b}_r : \mathcal{C}_r \to \hat{\mathcal{X}}_r^n$ maps that receiver's product channel codebook $\mathcal{C}_r = \prod_{S \in \mathcal{S}_r} \mathcal{C}_S$ to reproduction space $\hat{\mathcal{X}}_r^n$ by way of index set $\mathcal{I}_r = \prod_{S \in \mathcal{S}_r} \mathcal{I}_S$. Thus $\mathbf{b}_r = \beta_r \circ \delta_r$, where $\delta_r : \mathcal{C}_r \to \mathcal{I}_r$ and $\beta_r : \mathcal{I}_r \to \hat{\mathcal{X}}_r^n$. Here $\mathcal{I}_r$ may be interpreted as the index set of the channel codebook $\mathcal{C}_r$. The function δ_r converts each received channel codeword $c_r \in \mathcal{C}_r$ to an index $i_r \in \mathcal{I}_r$, and the function β_r maps this index i_r to a reproduction vector $\hat{\mathbf{X}}_r^n \in \hat{\mathcal{X}}_r^n$. Thus for any $\mathbf{x}^n \in \mathcal{X}^n$, $\mathbf{a}(\mathbf{x}^n) = \gamma(\alpha(\mathbf{x}^n)) = \mathbf{c}$ for some $\mathbf{c} = (c_S : S \in \mathcal{S}) \in \mathcal{C}$, and for each receiver r, $\mathbf{c}_r = (c_S : S \in \mathcal{S}_r) \in \mathcal{C}_r$ and $\mathbf{b}_r(\mathbf{c}_r) = \beta_r(\delta_r(\mathbf{c}_r)) = \hat{\mathbf{x}}_r^n$ for some $\hat{\mathbf{x}}_r^n \in \hat{\mathcal{X}}_r^n$, where $c_S = \gamma_S(\alpha(\mathbf{x}^n))$. We use $\beta_{r,S}(\delta_r(\mathbf{c}_r)) = \hat{x}_{S,r}^n$ to denote receiver r's reproduction of x_S^n. In order for each decoder to be able to successfully decode a sequence of source vector descriptions, the product channel codebook $\mathcal{C}_r$ must be uniquely decodable for each $r \in \mathcal{M}$. We consider both fixed-and variable-rate channel codebooks.

This paper treats the performance limits, optimal design, and empirical behavior of lossless and lossy BSSC. The work described here builds on work described in [1]. For the case of lossless coding, that paper demonstrates how lossless BSSCs can achieve optimal performance for 2-receiver systems. In this paper, we prove a source coding theorem for lossless BSSCs with more than 2 receivers and discuss optimal lossless BSSC design. For the case of lossy coding, [1] gives a general description of a practical BSSC algorithm. In this paper, we extend those results, giving a detailed description of the optimal design algorithm for fixed- and variable-rate broadcast channel vector quantization (BCVQ). The optimal design algorithm, which relies on an k-ary tree structure for some $k \geq 3$, generalizes the optimal multiple description vector quantizer (MDVQ) design algorithm of [2]. We demonstrate the empirical performance of BCVQ for a 2-receiver broadcast system and compare that performance with the performance achievable with independent single-transmitter, single-receiver codes in the same system, thereby demonstrating the benefits of BSSC.

This paper is organized as follows. In Section II, we derive the achievable rate region for lossless broadcast system source coding in an M-receiver system. In Section III we describe an optimal design algorithm for lossless BSSCs. Section IV contains a detailed description of the optimal design algorithm for fixed- and variable-rate BCVQs for general M-receiver broadcast systems. Section V contains experimental results. The key contributions of this paper are summarized in Section VI.

II   Theoretical Bounds on Lossless Codes

We derive the achievable rate region for lossless source coding in an M-receiver broadcast system. The derivation is similar to that of the Slepian-Wolf theorem [3], which describes the achievable rate region for multiple-access system source coding.

For our discussion of lossless BSSC performance bounds, we make the following assumptions. For each $S \in \mathcal{S}$, assume that $\mathcal{X}_S$ is a finite source alphabet. Further, assume that $\mathbf{X}[1], \mathbf{X}[2], \ldots$ is drawn i.i.d. with known probability mass function $p(\mathbf{x})$ on source alphabet $\mathcal{X}$. For any subset $\mathcal{W}$ of $\mathcal{S}$, we use $H(\mathbf{X}_{\mathcal{W}})$ to specify the entropy of source vector $\mathbf{X}_{\mathcal{W}} = (X_S : S \in \mathcal{W})$. For any subsets $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{S}$, $I(\mathbf{X}_{\mathcal{U}}; \mathbf{X}_{\mathcal{V}})$

denotes the mutual information of $\mathbf{X}_{\mathcal{U}}$ and $\mathbf{X}_{\mathcal{V}}$.

Definition 1 *The set $A_{\epsilon,r}^{(n)} \subset \mathcal{X}_r^n$ of ϵ-typical n-vectors is defined by*

$$A_{\epsilon,r}^{(n)} = A_{\epsilon,r}^{(n)}(\mathbf{X}_r) = \{\mathbf{x}_r^n : |-(1/n)\log p(\mathbf{x}_{\mathcal{W}}^n) - H(\mathbf{X}_{\mathcal{W}})| < \epsilon, \quad \forall\ \mathcal{W} \subseteq \mathcal{S}_r\}.$$

For any $\mathcal{U}, \mathcal{V} \subseteq \mathcal{S}_r$, $A_{\epsilon,r}^{(n)}(\mathbf{X}_{\mathcal{U}}|\mathbf{x}_{\mathcal{V}}^n)$ is then defined as the set of $\mathbf{x}_{\mathcal{U}}^n$ vectors that are jointly ϵ-typical with a particular $\mathbf{x}_{\mathcal{V}}^n$ vector, i.e.

$$A_{\epsilon,r}^{(n)}(\mathbf{X}_{\mathcal{U}}|\mathbf{x}_{\mathcal{V}}^n) = \{\mathbf{x}_{\mathcal{U}}^n : |-(1/n)\log p(\mathbf{x}_{\mathcal{U}}^n|\mathbf{x}_{\mathcal{V}}^n) - H(\mathbf{X}_{\mathcal{U}}|\mathbf{X}_{\mathcal{V}})| < \epsilon, \quad \forall\ \mathcal{U} \subseteq \mathcal{S}_r\}.$$

Definition 2 *A $((2^{nR_{S(1)}}, \dots, 2^{nR_{S(|\mathcal{S}|)}}), n)$ reduced broadcast system source code for an M-receiver broadcast system consists of $|\mathcal{S}|$ encoder maps $f_S : \mathcal{X}_S^n \to \{1, \dots, 2^{nR_S}\}$, $S \in \mathcal{S}$, and M decoder maps $g_r : \prod_{S \in \mathcal{S}_r}\{1, 2, \dots, 2^{nR_S}\} \to \prod_{S \in \mathcal{S}_r} \mathcal{X}_S^n, \quad r \in \mathcal{M}$.*

The idea of a reduced BSSC is prompted by the following observation. Consider a 3-receiver system, the sources are $X_{\{1,2,3\}}$, $X_{\{1,2\}}$, $X_{\{1,3\}}$, $X_{\{2,3\}}$, $X_{\{1\}}$, $X_{\{2\}}$, $X_{\{3\}}$, abbreviated as X_{123}, X_{12}, X_{13}, X_{23}, X_1, X_2, and X_3 for simplicity. Much of the performance benefits associated with broadcast system source coding can be achieved through sequential description of the sources $\{X_S : S \in \mathcal{S}\}$. In particular, we first losslessly describe X_{123}, then describe X_{12}, X_{13}, and X_{23} using conditional entropy codes conditioned on X_{123}. This sequential approach is decodable at all decoders since the description of X_{123} is received by all receivers. Similarly, the coding of each private message X_r can be based on knowledge of all the common information r receives. Note, however, that the coding of the common information which is shared by only 2 receivers, i.e., X_{12}, X_{13}, X_{23}, cannot be based on each other because, e.g., not all receivers that receive X_{12} also receive X_{13} or X_{23}. Given this observation, we seek codes in which the description of X_{12} is useful in the decoding of both X_{13} and X_{23} but not dependent on either. Only then can the description of X_{12} be successfully decoded at receivers 1 and 2. Towards this end, we replace the BSSC's single encoder with a collection of independent encoders and use an approach similar to the Slepian-Wolf coding strategy. The resulting code is the *reduced* BSSC. The rate region of the lossless source code achieved by this coding scheme is given by Theorem 1.

Theorem 1 *The set of achievable rate vectors for reduced broadcast system source coding in a single-transmitter, M-receiver broadcast system is given by*

$$R_{\mathcal{W}} \geq \max_{r \in \mathcal{M}: \mathcal{W} \subseteq \mathcal{S}_r} H(\mathbf{X}_{\mathcal{W}}|\mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}), \quad \forall\ \mathcal{W} \subseteq \mathcal{S}_r \text{ for some }\ r \in \mathcal{M}, \tag{1}$$

where $R_{\mathcal{W}} = \sum_{S \in \mathcal{W}} R_S$ and superscript c denotes the set complement operator.

Proof of Achievability: The proof follows an approach similar to the proof of the Slepian-Wolf theorem. We first randomly design a collection of encoders as follows. For each $S \in \mathcal{S}$, independently assign every $x_S^n \in \mathcal{X}_S^n$ to one of 2^{nR_S} bins according to a uniform distribution on $\{1, 2, \dots, 2^{nR_S}\}$. The resulting mappings are the encoders $\{f_S\}$. Reveal f_S to all receivers r, such that $S \in \mathcal{S}_r$. The decoders are then designed as follows. For any $r \in \mathcal{M}$, receiver r receives the index vector $(i_S, S \in \mathcal{S}_r)$ and

4

declares $\hat{\mathbf{x}}_r^n = \mathbf{x}_r^n$, if there is one and only one n-vector $\mathbf{x}_r^n \in \mathcal{X}_r^n$ such that for all $S \in \mathcal{S}_r$, $f_S(x_S^n) = i_S$ and $\mathbf{x}_r^n \in A_{\epsilon,r}^{(n)}$. Otherwise the decoder declares an error.

For any $r \in \mathcal{M}$, receiver r's probability of error is defined as $P_{e,r}^{(n)} = P(\hat{\mathbf{X}}_r \neq \mathbf{X}_r)$. We have an error if $\mathbf{X}_r^n \notin A_{\epsilon,r}^{(n)}$ or if there is another typical sequence in the same bin. Let $\mathbf{X}_r \sim p(\mathbf{x}_r)$, $\bar{x}_S^n \in \mathcal{X}_S^n$, $\bar{\mathbf{x}}_{\mathcal{W}}^n \in \mathcal{X}_{\mathcal{W}}^n$. Define the events

$$
\begin{aligned}
E_{0,r} &= \{ \mathbf{X}_r^n \notin A_{\epsilon,r}^{(n)} \}, \\
E_{\mathcal{W},r} &= \{ \exists \bar{\mathbf{x}}_{\mathcal{W}}^n : \forall S \in \mathcal{W}, \ \bar{x}_S^n \neq X_S^n, \ f_S(\bar{x}_S^n) = f_S(X_S^n) \\
&\qquad \text{and } (\bar{\mathbf{x}}_{\mathcal{W}}^n, \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}^n) \in A_{\epsilon,r}^{(n)} \} \quad \forall \mathcal{W} \subseteq \mathcal{S}_r.
\end{aligned}
$$

By the AEP, for any $\epsilon > 0$, $P(E_{0,r}) < \epsilon$ for n large enough. For any $\mathcal{W} \subseteq \mathcal{S}_r$,

$$
\begin{aligned}
P(E_{\mathcal{W},r}) &= \sum_{\mathbf{x}_r^n} p(\mathbf{x}_r^n) P\{ \exists \bar{\mathbf{x}}_{\mathcal{W}}^n : \forall S \in \mathcal{W}, \bar{x}_S^n \neq x_S^n, f_S(\bar{x}_S^n) = f_S(x_S^n), (\bar{\mathbf{x}}_{\mathcal{W}}^n, \mathbf{x}_{\mathcal{S}_r \cap \mathcal{W}^c}^n) \in A_{\epsilon,r}^{(n)} \} \\
&\leq \sum_{\mathbf{x}_r^n} p(\mathbf{x}_r^n) (\prod_{S \in \mathcal{W}} 2^{-nR_S}) |A_{\epsilon,r}^{(n)}(\mathbf{X}_{\mathcal{W}} | \mathbf{x}_{\mathcal{S}_r \cap \mathcal{W}^c}^n)| \leq (\prod_{S \in \mathcal{W}} 2^{-nR_S}) 2^{n(H(\mathbf{X}_{\mathcal{W}} | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}) + \epsilon)}.
\end{aligned}
$$

If $\sum_{S \in \mathcal{W}} R_S > H(\mathbf{X}_{\mathcal{W}} | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c})$, then $P(E_{\mathcal{W},r}) < \epsilon$ for n large enough. Combining the above results for all $r \in \mathcal{M}$ and $\mathcal{W} \subseteq \mathcal{S}_r$, we can bound the probability of error $P_e^{(n)}$ of the whole system for n large enough. In particular, if $\sum_{S \in \mathcal{W}} R_S > H(\mathbf{X}_{\mathcal{W}} | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c})$, $\forall r \in \mathcal{M}, \forall \mathcal{W} \subseteq \mathcal{S}_r$ (which defines the rate region in Theorem 1), then $P_e^{(n)} = \sum_{r \in \mathcal{M}} P_{e,r}^{(n)} \leq \sum_r (P(E_{0,r}) + \sum_{\mathcal{W} \subseteq \mathcal{S}_r} P(E_{\mathcal{W},r})) < K\epsilon$, where K is a constant. Thus there exists at least one code with probability of error $< K\epsilon$, and we can construct a sequence of codes with $P_e^{(n)} \to 0$. $\qquad\square$

Proof of Converse: Consider a reduced BSSC with fixed encoders $\{f_S\}$, decoders $\{g_r\}$ and probability of error $P_e^{(n)} \to 0$ as $n \to \infty$. Let $I_S = f_S(X_S^n)$. Since $P_e^{(n)} = \sum_{r \in \mathcal{M}} P_{e,r}^{(n)}$ goes to 0 when $n \to \infty$, we have for all r, $P_{e,r}^{(n)} \to 0$, as $n \to \infty$. For any $\mathcal{W} \subseteq \mathcal{S}$, let $\mathbf{I}_{\mathcal{W}}$ denote the index vector $(I_S : S \in \mathcal{W})$, and abbreviate the notation $\mathbf{I}_{\mathcal{S}_r}$ as $\mathbf{I}_r$. Then by Fano's inequality

$$
H(\mathbf{X}_r^n | \mathbf{I}_r) \leq P_{e,r}^{(n)} n \sum_{S \in \mathcal{S}_r} (\log |\mathcal{X}_S|) + 1 = n\epsilon_n \text{ and } H(\mathbf{X}_{\mathcal{W}}^n | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}^n, \mathbf{I}_r) \leq n\epsilon_n \quad \forall \mathcal{W} \subseteq \mathcal{S}_r,
$$

where $\epsilon_n \to 0$ as $n \to \infty$. Then for any $\mathcal{W} \subseteq \mathcal{S}_r$, any $r \in \mathcal{M}$,

$$
\begin{aligned}
n \sum_{S \in \mathcal{W}} R_S &\geq H(\mathbf{I}_{\mathcal{W}}) \geq H(\mathbf{I}_{\mathcal{W}} | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}^n) \geq I(\mathbf{X}_{\mathcal{W}}^n; \mathbf{I}_{\mathcal{W}} | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}^n) \\
&= H(\mathbf{X}_{\mathcal{W}}^n | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}^n) - H(\mathbf{X}_{\mathcal{W}}^n | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}^n, \mathbf{I}_r) \\
&\geq H(\mathbf{X}_{\mathcal{W}}^n | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}^n) - n\epsilon_n = nH(\mathbf{X}_{\mathcal{W}} | \mathbf{X}_{\mathcal{S}_r \cap \mathcal{W}^c}) - n\epsilon_n
\end{aligned}
$$

Dividing by n and taking the limit as $n \to \infty$, we have the desired converse. $\qquad\square$

III Lossless Code Design

In this section, we consider optimal lossless code design. While the theoretical lossless coding arguments allow for arbitrarily small but nonzero error probabilities, our

lossless code design requires a zero probability of error. For the sake of example, we focus on a 3-receiver broadcast system. Our method generalizes to broadcast systems with more than three receivers.

In a 3-receiver system, the sources that need to be described are X_{123}, X_{12}, X_{13}, X_{23}, X_1, X_2, and X_3. As discussed in Section II, we first losslessly describe common information X_{123} using minimum rate $H(X_{123})$. Then we losslessly describe common information X_{12}, X_{13} and X_{23} using some sort of conditional entropy code conditioned on X_{123}. Finally, for each receiver r, we losslessly describe private information X_r based on all X_S, such that $S \in \mathcal{S}_r$ and $|S| > 1$. The problem here is that the three receivers receive different subsets of the source description. As a result, there is no straightforward sequential description of X_{12}, X_{13}, and X_{23} that gives all receivers all of the conditioning information needed to decode. Given this difficulty, we first work on a system with only three sources X_{12}, X_{13}, and X_{23} and then use the results on this simplified problem to solve the more general case. Theorem 2 treats the simplified system while Corollary 1 treats its generalization.

Theorem 2 *Given a 3-receiver system with simplified message set $\mathcal{S} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$, the optimal truly lossless ($P_e = 0$) code for this broadcast system encodes and decodes all sources independently. Thus the optimal rates for i.i.d. sources are*

$$R_S = H(X_S) \quad \forall\ S \in \mathcal{S}.$$

Proof: For simplicity, the proof is given as a scalar coding argument. The same argument extends naturally to block codes. Let $c_S(x_S)$, $l_S(x_S)$ and $p_S(x_S)$, $S \in \mathcal{S}$, $x_S \in \mathcal{X}_S$ denote the codeword, codeword length, and probability respectively for symbol x_S of source X_S. Each receiver r receives a concatenated codeword $\mathbf{c}_r(\mathbf{x}_r) = (c_S(x_S) : S \in \mathcal{S}_r)$. Let $\mathbf{l}_r(\mathbf{x}_r) = \sum_{S \in \mathcal{S}_r} l_S(x_S)$ denote the length of codeword $\mathbf{c}_r(\mathbf{x}_r)$. Since the product channel codebook $\mathcal{C}_r$ is uniquely decodable for each $r \in \mathcal{M}$, the codeword length function $\mathbf{l}_r$ must satisfy the Kraft Inequality $\sum_{\mathbf{x}_r \in \mathcal{X}_r} 2^{-\mathbf{l}_r(\mathbf{x}_r)} \leq 1$ for each $r \in \mathcal{M}$.

The aim of the optimal lossless code design is to minimize the weighted sum of expected codeword lengths $\mathbf{l}_r$ subject to the constraint of unique decodability. So we minimize the Lagrangian

$$J = \sum_r \omega_r E_{\mathbf{X}_r}[\mathbf{l}_r(\mathbf{X}_r)] + \sum_r \lambda_r \sum_{\mathbf{x}_r} 2^{-\mathbf{l}_r(\mathbf{x}_r)}.$$

Consider $\sum_{x_S} p_S(x_S) = 1$ and taking partial derivatives, $\partial J / \partial l_S(x_S) = 0, \forall S \in \mathcal{S}, x_S \in \mathcal{X}_S$, we get $\lambda_r = \omega_r / \ln 2$, $r \in \{1, 2, 3\}$. Further, from the unique decodability constraints, $\sum_{\mathbf{x}_r} 2^{-\mathbf{l}_r(\mathbf{x}_r)} = \prod_{S \in \mathcal{S}_r} (\sum_{x_S \in \mathcal{X}_S} 2^{-l_S(x_S)}) = 1, \forall r \in \mathcal{M}$, we get $\sum_{x_S \in \mathcal{X}_S} 2^{-l_S(x_S)} = 1, \forall S \in \mathcal{S}$. The surprising implication of this result is that in a 3-receiver system with sources X_{12}, X_{13} and X_{23}, if the codebook $\mathcal{C}_r$ for each receiver r is uniquely decodable, then the codebook $\mathcal{C}_S$ for each source X_S must also be uniquely decodable. The optimal codeword lengths are then given by

$$l_S(x_S) = -\log p_S(x_S), \forall S \in \mathcal{S}, x_S \in \mathcal{X}_S;$$

thus independent encoding and decoding of X_{12}, X_{13} and X_{23} is optimal. The resulting expected codeword lengths are $\mathrm{E}[l_S(\mathrm{X}_S)] = \mathrm{H}(\mathrm{X}_S) \ \forall S \in \mathcal{S}$. These optimal results can be achieved to arbitrary accuracy using traditional block coding techniques such as Huffman codes. $\qquad\square$

Corollary 1 *Given a 3-receiver system with full message set $\mathcal{S} = \{\{1, 2, 3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1\}, \{2\}, \{3\}\}$, the optimal truly lossless ($P_e = 0$) code for this broadcast system achieves, to arbitrary accuracy, the following rates for i.i.d. sources*

$$
\begin{aligned}
R_{123} &= H(X_{123}), \\
R_S &= H(X_S | X_{123}), \ \forall S \in \{\{12\}, \{13\}, \{23\}\}, \\
R_{\{r\}} &= H(X_S | X_{123}, X_U, X_V), \ \forall r \in \{1, 2, 3\}, \ U, V \in \mathcal{S}_r \cap \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}
\end{aligned}
$$

Proof: We begin by coding X_{123} losslessly. Using Theorem 2, we then code X_{12}, X_{13} and X_{23} using independent conditional entropy codes matched to $p(x_S | x_{123})$ for each $S \in \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. Finally, for each $r \in \{1, 2, 3\}$, we code $X_{\{r\}}$ using a conditional entropy code matched to the distribution $p(x_r | \mathbf{x}_{S_r \cap \{r\}^c})$. $\qquad\square$

As a result, the optimal expected rates at receivers 1, 2 and 3 respectively are

$$
\sum_{S \in \mathcal{S}_r} R_S = H(\mathbf{X}_r) + I(X_U, X_V | X_{123}), \quad U, V \in \mathcal{S}_r \cap \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}. \tag{2}
$$

Notice that this result apparently contradicts the result given by Theorem 1, which states that lossless broadcast system source coding can achieve rates

$$
\sum_{S \in \mathcal{S}_r} R_S = H(\mathbf{X}_r), \tag{3}
$$

at the three receivers. The source of this discrepancy lies in the fact that Theorem 1 addresses the lossless source coding problem given an arbitrarily small but *nonzero* error probability; the lossless codes we design in Section III use lossless codes with probability of error zero. Theorem 2 and Corollary 1 generalize to zero-error lossless BSSCs on N-receiver broadcast channels. The optimal lossless source coding design problem for nonzero error probability codes (e.g., for use in lossy broadcast system source codes) will be addressed in future work.

IV   Lossy Code Design

Our BCVQ design algorithm is based on a tree structure similar to that used for the MDVQ algorithm [2]. To understand that tree structure, first consider the k-ary tree with $k = 5$ shown in Figure 1. The given tree is intended for use in an n-dimensional, 2-receiver BCVQ with index set $\mathcal{I}_{12} = \mathcal{I}_1 = \mathcal{I}_2 = \mathcal{I} = \{1, 2, 3, 4\}$. Each horizontal branch of the tree denotes an unreceived source description, while each downward path denotes a received index from $\mathcal{I}$. For the code given in Figure 1, each receiver receives 2 indices: receiver 1 receives $\mathbf{i}_1 = (i_{12}, i_1)$ and receiver 2 gets $\mathbf{i}_2 = (i_{12}, i_2)$. Thus each receiver can receive one of 16 possible messages. Associated with each
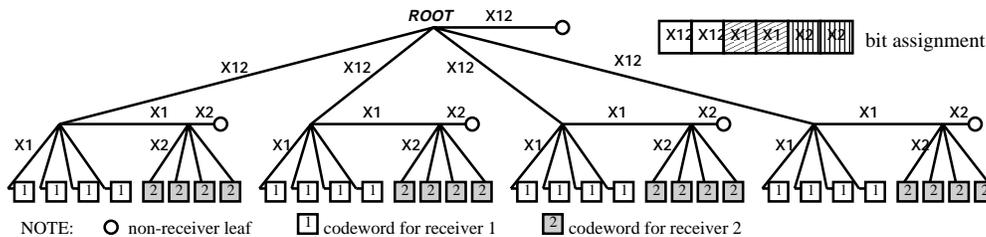
7

Figure 1: The k-ary tree structure for 2-receiver broadcast system lossy code.

possible message is a vector $(\hat{X}_S^n : S \in \mathcal{S})$ of reproduction vectors or *codewords*. The codewords reside at leaves of the tree. In particular, the codewords for receivers 1 and 2 are indicated by $\boxed{1}$ and $\boxed{2}$, respectively in Figure 1. The given figure shows the indices in order i_{12}, i_1, i_2 from top to bottom in the tree. (The order is arbitrary.) Since receiver 1 receives i_{12} and i_1 but not i_2, the codewords used by receiver 1 sit at the 16 nodes corresponding to two downward paths followed by a horizontal path through the tree. Receiver 2's codewords can be found in the same way. Nodes with no reproduction codewords correspond to combinations of received and unreceived messages not used in the current broadcast system configuration. The given structure generalizes to arbitrary M-receiver broadcast systems (by adding more levels to the tree) with arbitrary finite index sets (by allowing an arbitrary number of downward branches) and may be used for both fixed-rate coding (by describing each index using its natural binary description) and variable-rate coding (by describing each index using an appropriate entropy code, as discussed in Section III).

We now adapt the generalized Lloyd algorithm to BSSC design. For any receiver r and any set $S \in \mathcal{S}_r$, let $D_{S,r} = (1/n)\mathrm{E}_S[d(X_S^n, \hat{X}_{S,r}^n)]$ denote the average distortion of the reproduction of source X_S made by receiver r. For any set $S \in \mathcal{S}$, let $R_S = (1/n)\mathrm{E}_S[R_S(X_S^n)]$ denote the average rate used to transmit X_S, where $R_S(X_S^n)$ denotes the rate used to describe source vector X_S^n.

Let $\mathbf{D} = (D_{S,r} : S \in \mathcal{S}_r, r \in \mathcal{M})$ and $\mathbf{R} = (R_S : S \in \mathcal{S})$. The set of distortion-rate pairs $(\mathbf{D}, \mathbf{R})$ achievable through BCVQ (of arbitrary dimension) defines a convex region in the $(\sum_r |\mathcal{S}_r| + |\mathcal{S}|)$-dimensional distortion-rate space. This convex hull is entirely characterized by its support functional $J = \sum_r \sum_{S \in \mathcal{S}_r} \mu_{S,r}(D_{S,r} + \lambda_S R_S)$, where $\mu_{S,r}$ and λ_S are the associated Lagrangian constants. The functional J may be viewed as a Lagrangian for minimizing any combination of rates and distortions subject to constraints on the remaining quantities.

The goal of the algorithm is to minimize

$$
\begin{aligned}
&J(\alpha, \{\gamma_S\}, \{\delta_r\}, \{\beta_{r,S}\}) \\
=\ &\frac{1}{n}\mathrm{E}_{X_S^n}[\sum_{S \in \mathcal{S}_r, r} \mu_{S,r}[d(X_S^n, \beta_{r,S}(\delta_r(\mathbf{t}_r(\gamma(\alpha(\mathbf{X}^n)))))) + \lambda_S|\gamma_S(\alpha(X_S^n))|]],
\end{aligned}
$$

where α, $\{\gamma_S\}$, $\{\delta_r\}$ and $\{\beta_{r,S}\}$ are the encoders and decoders defined in Section I. The design of the BCVQ employs a convergent iterative descent approach reminiscent of the generalized Lloyd algorithm [4, 5]. Each iteration proceeds as follows.

1. **Optimize the encoder** α. For a given $\{\gamma_S\}$, $\{\delta_r\}$ and $\{\beta_{r,S}\}$, the optimal encoder $\alpha^\star$ is the one that maps each source vector $\mathbf{x}^n$ to the index i, such that

8

the reproductions $\hat{x}_{S,r}^n$ will result in the smallest J. Thus $\alpha^\star$ is given by

$$\alpha^\star(x^n) = \arg\min_{i\in\mathcal{I}} \sum_{S\in\mathcal{S}_{r,r}} \mu_{S,r}[d(X_S^n, \beta_{r,S}(\delta_r(\mathbf{t}_r(\gamma(i))))) + \lambda_S|\gamma_S(i)|].$$

2. **Optimize the decoder** β_r. For a given α, γ and $\{\delta_r\}_{r\in\mathcal{M}}$, the optimal decoder $\beta_r^\star(i_r)$, $r \in \mathcal{M}$ is the one that minimizes the expected reproduction distortion given i_r. The mathematical expression of $\beta_r^\star(i_r)$ is given by

$$\beta_r^\star(i_r) = \arg\min_{\hat{\mathbf{x}}_r^n\in\hat{\mathcal{X}}_r^n} \sum_S \mu_{S,r} \left[\mathrm{E}_{X_S^n} \left[d(X_S^n, \hat{x}_{S,r}^n)|\delta_r(\mathbf{t}_r(\gamma(\alpha(\mathbf{X}^n)))) = i_r \right] \right].$$

In particular, for the squared error distortion measure, $\beta_r^\star(i_r)$ is given by $\beta_r^\star(i_r)$: $\hat{x}_{S,r}^n = \mathrm{E}_{X_S^n}[X_S^n|\delta_r(\mathbf{t}_r(\gamma(\alpha(\mathbf{X}^n)))) = i_r]$.

3. **Optimize the lossless encoder** γ **and the lossless decoder** δ_r. For a given α and $\{\beta_r\}_{r\in\mathcal{M}}$, the optimal decoder γ and $\{\delta_r\}_{r\in\mathcal{M}}$ are the ones that minimize the expected description length. If we assume zero error probability, then the optimal lossless BSSC γ and $\{\beta_r\}_{r\in\mathcal{M}}$ is described by a generalization of Theorem 1. The optimal lossless code for nonzero error probabilities will be addressed in future work.

V   Experimental results

We include results for experiments performed for the two-receiver broadcast system using the algorithm described in Section IV. The results given here expand on those in [1] and rely on the same collection of assumptions and Lagrangian constants. In particular, $\mu_{\{2\},1} = \mu_{\{1\},2} = 0$, $\mu_{\{1\},1} = \sigma$, $\mu_{\{2\},2} = 1 - \sigma$ and $\mu_{\{1,2\},1} = \mu_{\{1,2\},2} = \theta/2$.

Further, each receiver is assigned a fixed rate of 4 bits per pixel (bpp), and the system uses a total rate of 6 bpp. To meet these rate constraints, in fixed-rate coding, we use a 5-ary tree and set $\lambda_S = 0$. In variable-rate coding, we use a 17-ary tree and for each (θ, σ) pair adjust λ_S to meet the desired rate constraint. By fixing the rates in this way, we can focus on the convex hull of the achieved distortions $\mathbf{D} = ((\mathbf{D}_{\{1,2\},1} + \mathbf{D}_{\{1,2\},2})/2, \mathbf{D}_{\{1\},1}, \mathbf{D}_{\{2\},2})$.

For comparison purposes, we use either fixed-rate VQ or entropy constrained VQ (ECVQ) to independently encode and decode the sources, and compare the resulting achievable distortions $\mathbf{D} = (\mathbf{D}_{\{1,2\}}, \mathbf{D}_{\{1\}}, \mathbf{D}_{\{2\}})$ with those of BCVQ. Since the rate assigned to each receiver is identical for the BCVQ and for the independent codes, the lower the convex hull of distortions in 3-dimension space, the better the performance.

Figure 2 compares the distortion triple of fixed-rate BCVQ with that of fixed-rate independent VQ, showing the 3-dimensional distortion space from three different angles. Figure 3 compares the distortion triple of variable-rate BCVQ with that of independent ECVQ. In both comparisons, the convex hull achieved by the BCVQ (solid line) is significantly lower than that achieved by the independent source code (dashed line) in the 3-dimensional space. Thus it demonstrates the performance benefits of BCVQs over independent source codes in a 2-receiver system.
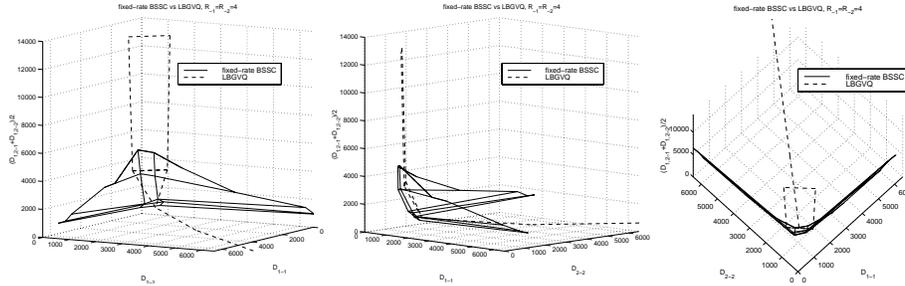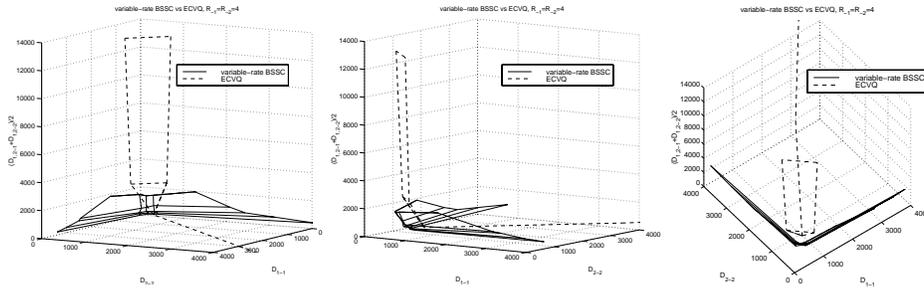
9

Figure 2: Fixed-rate BCVQ vs fixed-rate VQ.

Figure 3: Variable-rate BCVQ vs ECVQ.

VI Summary

In this paper, we prove the theoretical limits of lossless source coding performance in M-receiver broadcast systems. This proof also shows that only through joint coding can we hope to achieve optimal performance for broadcast systems. We then briefly describe a practical lossless coding scheme for a 3-receiver broadcast system. A detailed description of the practical algorithm for BCVQ is also given followed by experimental results, which show the benefits of BCVQ over independent coding.

References

[1] Q. Zhao and M. Effros. Broadcast system source codes: a new paradigm for data compression. In *Conference Record, Thirty-Third Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, October 1999. IEEE.

[2] M. Fleming and M. Effros. Generalized multiple description vector quantization. In *Proceedings of the Data Compression Conference*, pages 3–12, Snowbird, UT, March 1999. IEEE.

[3] D. Slepian and J. K. Wolf. A coding theorem for multiple access channels with correlated information sources. *Bell System Technical Journal*, 52:1037–1076, 1973.

[4] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28:84–95, January 1980.

[5] P. A. Chou, T. Lookabaugh, and R. M. Gray. Entropy-constrained vector quantization. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(1):31–42, January 1989.