

# A Strong Converse for a Collection of Network Source Coding Problems

WeiHsin Gu

Department of Electrical Engineering  
California Institute of Technology  
Pasadena, CA 91125, USA  
Email: wgu@caltech.edu

Michelle Effros

Department of Electrical Engineering  
California Institute of Technology  
Pasadena, CA 91125, USA  
Email: effros@caltech.edu

**Abstract**—We prove a strong converse for particular source coding problems: the Ahlswede-Körner (coded side information) problem, lossless source coding for multicast networks with side-information at the end nodes, and the Gray-Wyner problem. Source and side-information sequences are drawn i.i.d. according to a given distribution on a finite alphabet. The strong converse discussed here states that when a given rate vector  $\mathbf{R}$  is not  $\mathbf{D}$ -achievable, the probability of observing distortion  $\mathbf{D}$  for any sequence of block codes at rate  $\mathbf{R}$  must decrease exponentially to 0 as the block length grows without bound. This strong converse implies the prior strong converses for the point-to-point network, Slepian-Wolf problem, and Ahlswede-Körner (coded side information) problem.

## I. INTRODUCTION

In the traditional source coding scenario, here called the point-to-point network, comprised of one source node describing source  $X$  to one sink node across a single link, strong converses for both lossless source coding and lossy source coding have previously appeared in the literature. For example, [1] treats i.i.d. finite-alphabet source sequences, and [2], [3], and [4] treat more general source sequences. While the lossless source coding theorem describes the family of rates that can be achieved with arbitrarily small error probability, the strong converse states that for any rate outside the rate region, the probability of a correct reconstruction approaches 0 as the blocklength grows without bound. Similarly, the rate-distortion theorem describes the set of rates that can be achieved with expected distortion no greater than  $D$  while its strong converse demonstrates that the probability of observing distortion less than  $D$  at any rate outside this region approaches 0 as the blocklength grows without bound.

In this paper, we derive strong converses for three problems: the Ahlswede-Körner (coded side information) problem, lossless source coding for multicast networks with side-information at the end nodes, and the Gray-Wyner problem. The source sequences are drawn i.i.d. according to a finite-alphabet source distribution. Generalized from the strong converse for lossy source coding of the point-to-point network in [1], the strong converses of interest state that for any distortion vector  $\mathbf{D}$  when a rate vector  $\mathbf{R}$  is not in the  $\mathbf{D}$ -achievable rate region, the probability of observing distortion

at most  $\mathbf{D}$  with a rate- $\mathbf{R}$  code decreases exponentially to 0 as the blocklength  $n$  grows without bound. We call such a result an exponentially strong converse to emphasize the speed of convergence of the correct probabilities for the rate vectors outside the rate region. The exponentially strong converse for a network source coding problem is useful for a variety of applications beyond basic understanding of how achievable error probability varies with rate. For example, when the exponentially strong converse holds, we can show that any demands that can be achieved across a network with rate 0 across a given link can also be achieved when that link is absent [5]. This property is actually quite subtle since it requires demonstrating that asymptotically small rates across the given link are never critical to that network's operation. Notice that this property is not trivial since single-letter characterizations of the network with that additional link may not be available even when a single-letter characterization of the network without the 0-rate link is known. As mentioned above, the exponentially strong converse holds for the point-to-point network. For the point-to-point lossless case, the intuition is that the probability of the strongly typical set  $A_\epsilon^{*(n)}(X)$  for the finite-alphabet source  $X$  increases exponentially to 1 as the length  $n$  grows without bound. We denote exponent by  $\tau(\epsilon) > 0$ . Let  $B_\epsilon^{(n)}$  denote the intersection of  $A_\epsilon^{*(n)}(X)$  and the correct event for a given sequence of codes. When the code's probability of correctness equals  $2^{-nc(n)}$  for some  $c(n) \rightarrow 0$ ,

$$\frac{1}{n} \log \frac{|A_\epsilon^{*(n)}(X)|}{|B_\epsilon^{(n)}|}$$

can be made arbitrarily small when  $\epsilon > 0$  is sufficiently small and  $n$  is sufficiently large. This means that the rate  $R$  that is sufficient to describe the set  $B_\epsilon^{(n)}$  is asymptotically at least  $\frac{1}{n} \log |A_\epsilon^{*(n)}(X)|$ .

We prove that the exponentially strong converse holds for the lossless coded side information problem [6], lossless source coding for the family of multicast networks with side information at the end nodes, and the Gray-Wyner problem [7]. The cut-set bound is tight for multicast networks with side information at the end nodes by [8], and this family includes the family of multicast networks [9] as a subfamily, which includes the Slepian-Wolf problem [10] as a special case. The

<sup>0</sup>This material is based upon work partially supported by NSF Grant No. CCF-0325324 and Caltech's Lee Center for Advanced Networking.

strong converse has been proven for the coded side information problem in [11] and for the Slepian-Wolf problem in [12]. Neither of these results shows exponential decay.

The remainder of this paper is structured as follows. We define the exponentially strong converse in Section II. We show that the exponentially strong converse is true for the coded side information problem in Section III. In Section IV, we prove the exponentially strong converse for the family of multicast networks with side information at the end nodes. Finally, we prove the exponentially strong converse for the Gray-Wyner problem in Section V.

## II. DEFINITION OF THE EXPONENTIALLY STRONG CONVERSE

We define the exponentially strong converse here. Notice that for the network source coding problems discussed in this paper, lossless source coding can be treated as a special case of lossy source coding. Thus, it suffices to define the exponentially strong converse for lossy source coding. In the following definition, we use  $X_1, \dots, X_s$  to denote the source random variables and  $\mathcal{D} = \{(v, X_i) \mid v \text{ demands } X_i\}$  to denote the set of demands in the given network source coding problem [5].

*Definition 1:* We say that the exponentially strong converse holds for network source coding problem  $\mathcal{N}$  if and only if for any distortion vector  $\mathbf{D}$  and any rate vector  $\mathbf{R}$ ,  $\mathbf{R}$  is not in the achievable rate region if and only if for any sequence of rate- $\mathbf{R}$ , length- $n$  block codes  $\{\mathcal{C}_n\}$  the probability of observing distortion no greater than  $\mathbf{D}$  decreases to 0 exponentially, i.e.,

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log \Pr \left( \frac{1}{n} d(X_i^n, \hat{X}_i^n(v)) \leq D \forall (v, X_i) \in \mathcal{D} \right) > 0,$$

where for all  $(v, X_i) \in \mathcal{D}$ ,  $\hat{X}_i^n(v)$  is the reproduction of  $X_i^n$  at node  $v$  using  $\mathcal{C}_n$ .

Our approach relies on strong typicality. We briefly mention some properties that are useful here and fix the notation as follows. Let  $W$  be a finite-alphabet random variable. For any integer  $n$  and positive number  $\epsilon > 0$ , let  $A_\epsilon^{*(n)}(W)$  denote the strongly typical set. (For example, see [13].) Lemma 1 states that the probability of the atypical set  $(A_\epsilon^{*(n)}(W))^c$  decreases exponentially to 0 as  $n$  grows without bound with exponent greater than or equal to  $\tau(\epsilon)$  that depends only on the distribution of  $W$  and  $\epsilon$ . We state Lemma 1 without proof.

*Lemma 1:* For any  $\epsilon > 0$ , when  $n$  is sufficiently large,  $\Pr(A_\epsilon^{*(n)}(W)) \geq 1 - 2^{-n\tau(\epsilon)}$  for some  $\tau(\epsilon) > 0$ .

Let  $Z$  be another finite-alphabet random variable. For  $z^n \in \mathcal{Z}^n$ , let  $A_\epsilon^{*(n)}(W|Z = z^n)$  denote the set of sequences  $w^n \in \mathcal{W}^n$  of which  $(w^n, z^n)$  is strongly typical.

## III. THE LOSSLESS CODED SIDE INFORMATION PROBLEM

We here prove that the exponentially strong converse for the lossless coded side information problem [6]. (See Fig. 1.) Our proof follows the approach in [11], where the strong converse theorem for this network source coding problem without exponential decay has been proven. We start by redefining

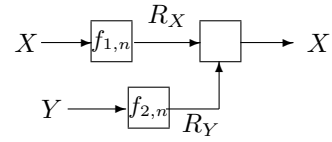


Fig. 1. The coded side information problem.

some terms introduced in [11] that are useful in this section. For simplicity, we use different notation and definitions from those in [11] that are sufficient for our problem. Please refer to [11] for more details.

*Definition 2:* For any positive integer  $n$ , positive number  $\delta > 0$ , and set  $\mathcal{B} \subseteq \mathcal{X}^n$ , define

$$\psi_\delta(\mathcal{B}) := \{y^n \mid \Pr(X^n \in \mathcal{B} \mid Y^n = y^n) \geq 2^{-n\delta}\}.$$

*Definition 3:* For any  $c > 0$ ,  $\epsilon > 0$ , and  $\delta > 0$ , define

$$\hat{S}_n(c, \epsilon, \delta) := \frac{1}{n} \log \min |\mathcal{B}|,$$

where the min is taken over all subsets  $\mathcal{B} \subseteq \mathcal{X}^n$  such that

$$\Pr(Y^n \in \psi_\delta(\mathcal{B}) \cap A_\epsilon^{*(n)}(Y)) \geq 2^{-cn}.$$

Define

$$\hat{T}(c) := \min H(X|U),$$

where the min is taken over all random variables  $U$  such that  $X \rightarrow Y \rightarrow U$  forms a Markov chain and  $I(Y; U) \leq c$ .

*Theorem 1:* [11, Theorem 1] For all  $\epsilon > 0$ ,  $\delta > 0$ , and  $c > 0$

$$\lim_{n \rightarrow \infty} \hat{S}_n(c, \epsilon, \delta) = \hat{T}(c).$$

*Theorem 2:* The exponentially strong converse holds for the coded side information problem.

*Proof.* Let  $(R_X, R_Y)$  be a rate pair such that there exists a sequence of length- $n$  rate- $(R_X, R_Y)$  block codes  $\{\mathcal{C}_n\}_{i=1}^\infty$  such that the correct probability  $\Pr(X^n = \hat{X}^n) = 2^{-nc(n)}$  for some sequence  $\{c(n)\}$  satisfying  $\lim_{n \rightarrow \infty} c(n) = 0$ . Here  $f_{1,n}$ ,  $f_{2,n}$ , and  $g_n$  are the encoding and decoding functions of  $\mathcal{C}_n$  (as shown in Fig. 1) and

$$\hat{X}^n = g_n(f_{1,n}(X^n), f_{2,n}(Y^n))$$

is the reproduction of  $X^n$  using code  $\mathcal{C}_n$ . We want to show that  $(R_X, R_Y)$  is achievable, i.e., that there exists a random variable  $U$  such that  $X \rightarrow Y \rightarrow U$  forms a Markov chain and that

$$R_X \geq H(X|U), \quad R_Y \geq I(Y; U).$$

For any particular value  $u \in \{1, 2, \dots, 2^{nR_Y}\}$  of the encoding function  $f_{2,n}$ , define

$$C^{(n)}(u) := \{x^n \mid x^n = g_n(f_{1,n}(x^n), u)\}.$$

By assumption,

$$2^{-nc(n)} \leq \sum_{y^n \in \mathcal{Y}^n} (\Pr(Y^n = y^n) \times \Pr(C^{(n)}(f_{2,n}(y^n)) \mid Y^n = y^n)).$$

Fix  $\epsilon > 0$ . By Lemma 1,

$$\Pr(A_\epsilon^{*(n)}(X, Y)) \geq 1 - 2^{-n\tau(\epsilon)}$$

for some  $\tau(\epsilon) > 0$ . Therefore when  $n$  is sufficiently large,

$$\sum_{y^n \in A_\epsilon^{*(n)}(Y)} (\Pr(Y^n = y^n) \times \quad (1)$$

$$\begin{aligned} & \Pr(C^{(n)}(f_{2,n}(y^n)) \mid Y^n = y^n) \\ & \geq 2^{-nc(n)} - 2^{-n\tau(\epsilon)} \geq 2^{-nb(n,\epsilon)} \quad (2) \end{aligned}$$

for some sequence of positive numbers  $\{b(n, \epsilon)\}$  such that  $\lim_{n \rightarrow \infty} b(n, \epsilon) = 0$  for all  $\epsilon > 0$ . Let  $S(n, \epsilon) \subseteq A_\epsilon^{*(n)}(Y)$  be the set of all  $y^n \in A_\epsilon^{*(n)}(Y)$  such that

$$\Pr(C^{(n)}(f_{2,n}(y^n)) \mid Y^n = y^n) \geq 2^{-nb(n,\epsilon)}.$$

Then (2) implies

$$\Pr(S(n, \epsilon)) + (1 - \Pr(S(n, \epsilon)))2^{-nb(n,\epsilon)} \geq 2^{-nb(n,\epsilon)},$$

which leads to  $\Pr(S(n, \epsilon)) > 2^{-n(b(n,\epsilon) + \frac{1}{n})}$  when  $n$  is sufficiently large. Now by definition

$$S(n, \epsilon) \subseteq \bigcup_u \left( \psi_{2b(n,\epsilon)}(C^{(n)}(u)) \cap A_\epsilon^{*(n)}(Y) \right).$$

Hence

$$\sum_u \Pr \left( \psi_{2b(n,\epsilon)}(C^{(n)}(u)) \cap A_\epsilon^{*(n)}(Y) \right) \geq 2^{-n(b(n,\epsilon) + \frac{1}{n})},$$

where the summation is taken over all  $u \in \{1, \dots, 2^{nR_Y}\}$ . Thus there exists an index  $u^*$  such that

$$\Pr \left( \psi_{2b(n,\epsilon)}(C^{(n)}(u^*)) \cap A_\epsilon^{*(n)}(Y) \right) \geq 2^{-n(b(n,\epsilon) + \frac{1}{n} + R_Y)}.$$

By the definition of  $\widehat{S}_n$ ,

$$\frac{1}{n} \log |C^{(n)}(u^*)| \geq \widehat{S}_n \left( b(n, \epsilon) + \frac{1}{n} + R_Y, \epsilon, 2b(n, \epsilon) \right).$$

Choose  $n$  sufficiently large so that  $\frac{1}{n} + b(n, \epsilon) < \epsilon$ . By Theorem 1,

$$\frac{1}{n} \log |C^{(n)}(u^*)| \geq \widehat{T}(\epsilon + R_Y) + \epsilon$$

when  $n$  is sufficiently large. Since  $C^{(n)}(u^*)$  is the set of  $x^n$  that can be correctly decoded when  $f_{2,n}(Y^n) = u^*$ ,  $|C^{(n)}(u^*)| \leq 2^{nR_X}$  and hence

$$R_X \geq \widehat{T}(\epsilon + R_Y) + \epsilon$$

when  $n$  is sufficiently large. By the definition of  $\widehat{T}$ , there exists an auxiliary random variable  $U$  such that  $X \rightarrow Y \rightarrow U$ ,  $R_X \geq H(X|U) + \epsilon$ , and  $R_Y \geq I(Y; U) + \epsilon$ . Letting  $\epsilon \rightarrow 0$  completes the proof.  $\square$

#### IV. LOSSLESS SOURCE CODING FOR THE MULTICAST NETWORKS WITH SIDE INFORMATION AT THE END NODES

In this section, we treat the family of multicast networks with side information at the end nodes. The simplest interesting example in this family is the problem of lossless

source coding with side information at the decoder. (See Fig. 2.) The infimum of the set of losslessly achievable rates is  $H(X|Y)$ , which corresponds to one of the two corner points in the rate region of the Slepian-Wolf problem. We prove the exponentially strong converse for this basic example in Lemma 2 and then use it to conclude in Theorem 3 that the exponentially strong converse holds for the family of multicast networks with side information at the end nodes.

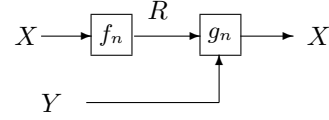


Fig. 2. The lossless source coding problem with side information at the decoder.

*Lemma 2:* The exponentially strong converse holds for lossless source coding problem with side information at the decoder.

*Proof.* Let  $R > 0$ . Suppose that there exists a sequence of length- $n$  rate- $R$  block codes  $\mathcal{C}_n$  with correct probability

$$\Pr(X^n = g_n(f_n(X^n), Y^n)) = 2^{-nc(n)}$$

for some sequence  $\{c(n)\}$  such that  $\lim_{n \rightarrow \infty} c(n) = 0$ , where  $g_n(f_n(X^n), Y^n)$  is the reproduction of  $X^n$  using code  $\mathcal{C}_n$ . We want to show that  $R$  is in the lossless rate region by showing that  $R \geq H(X|Y)$ .

For any positive integer  $n$  and positive real number  $\epsilon > 0$ , let

$$B_\epsilon^{(n)} := A_\epsilon^{*(n)}(X, Y) \cap \{(x^n, y^n) : x^n = g_n(f_n(x^n), y^n)\}.$$

Lemma 1 implies that

$$\Pr(A_\epsilon^{*(n)}(X, Y)) \geq 1 - 2^{-n\tau(\epsilon)}$$

for some  $\tau(\epsilon) > 0$ , so by an argument like the one used to get (2),

$$\Pr(B_\epsilon^{(n)}) \geq 2^{-nc(n)} - 2^{-n\tau(\epsilon)} = 2^{-nb(n,\epsilon)} \quad (3)$$

when  $n$  is sufficiently large for some sequence of positive numbers  $\{b(n, \epsilon)\}$  such that  $\lim_{n \rightarrow \infty} b(n, \epsilon) = 0$  for all  $\epsilon > 0$ . Hence there exists a  $y_0^n \in A_\epsilon^{*(n)}(Y)$  such that

$$\Pr(B_\epsilon^{(n)} | Y^n = y_0^n) \geq 2^{-nb(n,\epsilon)}. \quad (4)$$

Since for all  $x^n \in A_\epsilon^{*(n)}(X | Y^n = y_0^n)$

$$\Pr(X^n = x^n | Y^n = y_0^n) \leq 2^{-n(H(X|Y) - 2\epsilon)},$$

(4) implies that

$$|B_\epsilon^{(n)} \cap \{Y^n = y_0^n\}| \geq 2^{n(H(X|Y) - 2\epsilon - b(n,\epsilon))}. \quad (5)$$

Since  $\mathcal{C}_n$  has rate  $R$  and  $B_\epsilon^{(n)} \cap \{Y^n = y_0^n\}$  is by definition the set of pairs  $(x^n, y_0^n) \in A_\epsilon^{*(n)}(X, Y)$  such that  $x^n$  can be correctly decoded when  $Y^n = y_0^n$ ,

$$2^{nR} \geq |B_\epsilon^{(n)} \cap \{Y^n = y_0^n\}|.$$

Thus (5) implies that  $R \geq H(X|Y) - 2\epsilon - b(n, \epsilon)$  for all  $n$

and  $\epsilon$ . Since  $\epsilon > 0$  is arbitrary, letting  $n \rightarrow \infty$  gives

$$R \geq H(X|Y). \quad \square$$

Applying Lemma 2 and the tightness of the cut-set bound in this scenario, we conclude this section by Theorem 3.

*Theorem 3:* The exponentially strong converse holds for the multicast network with side information at the end nodes.

## V. THE GRAY-WYNER PROBLEM

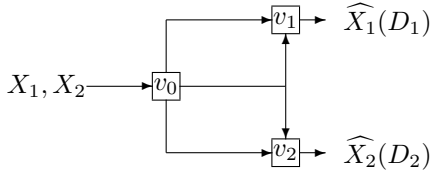


Fig. 3. The Gray-Wyner network.

Given  $\mathbf{D} = (D_1, D_2) \geq \mathbf{0}$ . The lossy rate region for the Gray-Wyner problem equals the closure of the set of all  $(R_0, R_1, R_2)$  for which there exists a random variable  $U$  with alphabet size  $|\mathcal{U}| \leq |\mathcal{X}_1| |\mathcal{X}_2| + 2$  such that

$$R_0 \geq I(\mathbf{X}; U), \quad R_1 \geq R_{X_1|U}(D_1), \quad R_2 \geq R_{X_2|U}(D_2), \quad (6)$$

where  $\mathbf{X} = (X_1, X_2)$  denotes the source vector and  $R_{X_i|U}(D_i)$  is the conditional rate-distortion function for source  $X_i$  and side information  $U_i$  at distortion  $D_i$  for  $i \in \{1, 2\}$ .

Theorem 4 proves the exponentially strong converse for the Gray-Wyner Problem. We sketch the proof in this paper and the details can be found in [14]. The main idea of the proof is that when the exponent of the probability of achieving distortion  $\mathbf{D}$  is asymptotically zero, the given rate vector is  $\mathbf{D}$ -achievable for another distribution  $\mathbf{X}$  that is close to  $P_{\mathbf{X}}$ . The approach then follows the method of proving the converse of the region (6) in [7] that turns the dimension- $n$  description of the rate vectors into a single-letter form. Lemma 3 shows the existence of another distribution which is close to  $P_{\mathbf{X}}$  such that the given rate vector  $\mathbf{R}$  is  $\mathbf{D}$ -achievable for this new distribution.

*Lemma 3:* Let  $W$  be a random variable with alphabet  $\mathcal{W}$  and distribution  $P_W$ . Let  $\{B^{(n)}\}$  be any sequence of sets  $B^{(n)} \subseteq \mathcal{W}^n$  such that  $P_{W^n}(B^{(n)}) = 2^{-nb(n)}$  for some sequence of non-negative numbers  $\{b(n)\}_{n=1}^{\infty}$  satisfying  $\lim_{n \rightarrow \infty} b(n) = 0$ . Then there exist a sequence  $\{a(n)\}_{n=1}^{\infty}$  of non-negative numbers and a sequence  $\{Q_{W^n}^{(n)}\}$  of distributions on  $\mathcal{W}^n$  such that

$$\begin{aligned} \lim_{n \rightarrow \infty} a(n) &= 0, \quad \lim_{n \rightarrow \infty} Q_{W^n}^{(n)}(B^{(n)}) = 1 \\ 2^{-na(n)} P_{W^n}(w^n) &\leq Q_{W^n}^{(n)}(w^n) \leq 2^{na(n)} P_{W^n}(w^n) \\ \forall w^n &\in \mathcal{W}^n. \end{aligned}$$

*Proof.* For all  $w^n \in B^{(n)}$ , define

$$Q_{W^n}^{(n)}(w^n) = \frac{2^{n(b(n) + \frac{1}{\sqrt{n}})} P_{W^n}(w^n)}{2^{n(b(n) + \frac{1}{\sqrt{n}})} P_{W^n}(B^{(n)}) + (1 - P_{W^n}(B^{(n)}))}.$$

For all  $w^n \notin B^{(n)}$ , define

$$Q_{W^n}^{(n)}(w^n) = \frac{P_{W^n}(w^n)}{2^{n(b(n) + \frac{1}{\sqrt{n}})} P_{W^n}(B^{(n)}) + (1 - P_{W^n}(B^{(n)}))}.$$

Then

$$Q_{W^n}^{(n)}(B^{(n)}) \geq \frac{2^{\sqrt{n}}}{2^{\sqrt{n}} + 1} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$$2^{-n(\frac{1}{\sqrt{n}} + \frac{1}{n})} P_{W^n}(w^n) \leq Q_{W^n}^{(n)}(w^n) \leq 2^{n(b(n) + \frac{1}{\sqrt{n}})} P_{W^n}(w^n) \\ \forall w^n \in \mathcal{W}^n$$

as desired.  $\square$

*Theorem 4:* The exponentially strong converse holds for the Gray-Wyner Problem.

*Proof.* Let  $\mathbf{R} = (R_0, R_1, R_2)$  be a rate vector and  $P_{\mathbf{X}}$  denote the source distribution. Suppose that there exists a sequence of length- $n$ , rate- $\mathbf{R}$  block codes  $\{\mathcal{C}_n\}$  such that

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \Pr \left( Ed(X_i^n, \hat{X}_i^n) \leq D_i \quad \forall i \in \{1, 2\} \right) = 0,$$

where  $\hat{X}_1^n$  and  $\hat{X}_2^n$  are reproductions of  $X_1^n$  and  $X_2^n$  at the nodes  $v_1$  and  $v_2$ , respectively. (See Fig. 3.) We want to show that  $\mathbf{R}$  is in the region described in (6).

For  $\epsilon > 0$ , let

$$B_\epsilon^{(n)} := A_\epsilon^{*(n)}(\mathbf{X}) \cap \{ \mathbf{x}^n \mid d(x_i^n, \hat{X}_i^n(\mathbf{x}^n)) \leq nD_i \quad \forall i \in \{1, 2\} \}.$$

Then following the argument used to prove (3) leads to

$$\Pr(B_\epsilon^{(n)}) = 2^{-nb(n, \epsilon)} \quad (7)$$

for some sequence of non-negative numbers  $\{b(n, \epsilon)\}$  such that  $\lim_{n \rightarrow \infty} b(n, \epsilon) = 0$  for all  $\epsilon > 0$ .

By Lemma 3, there exists a sequence of non-negative numbers  $\{a(n, \epsilon)\}_{n=1}^{\infty}$  and a sequence of distributions  $\{Q_{\mathbf{X}^n}^{(n, \epsilon)}\}$  such that for all  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} a(n, \epsilon) = 0, \quad \lim_{n \rightarrow \infty} Q_{\mathbf{X}^n}^{(n, \epsilon)}(B_\epsilon^{(n)}) = 1,$$

and for all  $\mathbf{x}^n \in \mathcal{X}_1^n \times \mathcal{X}_2^n$ ,

$$2^{-na(n, \epsilon)} P_{\mathbf{X}^n}(\mathbf{x}^n) \leq Q_{\mathbf{X}^n}^{(n, \epsilon)}(\mathbf{x}^n) \leq 2^{na(n, \epsilon)} P_{\mathbf{X}^n}(\mathbf{x}^n) \quad (8)$$

Let  $n(\epsilon)$  be a positive integer such that  $a(n(\epsilon), \epsilon) < \epsilon$  and  $Q_{\mathbf{X}^{n(\epsilon)}}^{(n(\epsilon), \epsilon)}(B_\epsilon^{(n(\epsilon))}) > 1 - \epsilon$ . Let  $Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}$  simply denote the distribution  $Q_{\mathbf{X}^{n(\epsilon)}}^{(n(\epsilon), \epsilon)}$ . By the continuity of the lossy rate region with respect to distortion vector, there exists  $\tau_1(\epsilon)$  with  $\lim_{\epsilon \rightarrow \infty} \tau_1(\epsilon) = 0$  such that the rate vector  $n(\mathbf{R} + \tau_1(\epsilon) \cdot \mathbf{1})$  is in the  $\mathbf{D}$ -achievable region for the Gray-Wyner problem with respect to distribution  $Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}$ .

Hence there exists a random variable  $U$  such that

$$n(\epsilon)(\mathbf{R} + \tau_1(\epsilon)) \cdot \mathbf{1} \geq (I_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)}; U), \\ R_{X_1^{n(\epsilon)}|U}(n(\epsilon)D_1, Q^{(\epsilon)}), R_{X_2^{n(\epsilon)}|U}(n(\epsilon)D_2, Q^{(\epsilon)}),$$

where  $I_{Q^{(\epsilon)}}$  and  $R_{X_i^{n(\epsilon)}|U}(D_i, Q^{(\epsilon)})$  (for  $i \in \{1, 2\}$ ) are the mutual information and conditional rate-distortion functions evaluated according to distribution  $Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}$ . Let  $J(\epsilon)$  be an independent random variable uniformly distributed over  $\{1, \dots, n(\epsilon)\}$ . Define  $U_j = (U, \mathbf{X}_1^{j-1})$  for  $j \in \{1, \dots, n(\epsilon)\}$ . Then

$$I_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)}; U) = \sum_{j=1}^{n(\epsilon)} I_{Q^{(\epsilon)}}(\mathbf{X}_j; U | \mathbf{X}_1^{j-1}) \\ = n(\epsilon)I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; U_{J(\epsilon)}, J(\epsilon)) + H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)}) \\ - n(\epsilon)H_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}).$$

Similarly, for  $i = 1, 2$ ,

$$R_{X_i^{n(\epsilon)}|U}(nD_i, Q^{(\epsilon)}) \geq n(\epsilon)R_{X_i, J(\epsilon)|U_{J(\epsilon)}, J(\epsilon)}(D_i, Q^{(\epsilon)}).$$

Since  $\mathbf{X}_{J(\epsilon)}$  has finite alphabet  $\mathcal{X}_1 \times \mathcal{X}_2$ , there exists a conditional distribution  $Q_{V_{n(\epsilon)}|\mathbf{X}_{J(\epsilon)}}$  for a random variable  $V_{n(\epsilon)}$  with alphabet size  $|\mathcal{X}_1||\mathcal{X}_2| + 2$  such that for  $i = 1, 2$ ,

$$I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; U_{J(\epsilon)}, J(\epsilon)) = I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; V_{n(\epsilon)}) \\ R_{X_i, J(\epsilon)|U_{J(\epsilon)}, J(\epsilon)}(D_i, Q^{(\epsilon)}) \geq R_{X_i, J(\epsilon)|V_{n(\epsilon)}}(D_i, Q^{(\epsilon)}).$$

We next show that

$$\lim_{\epsilon \rightarrow 0} Q_{\mathbf{X}_{J(\epsilon)}}^{(\epsilon)}(\mathbf{x}) = P_{\mathbf{X}}(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}_1 \times \mathcal{X}_2 \quad (9)$$

$$\lim_{\epsilon \rightarrow 0} \frac{H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)})}{n(\epsilon)} - H_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}) = 0. \quad (10)$$

First, for all  $\mathbf{x}^{n(\epsilon)} \in \mathcal{X}_1^{n(\epsilon)} \times \mathcal{X}_2^{n(\epsilon)}$  and for all  $\alpha \in \mathcal{X}_1 \times \mathcal{X}_2$ ,

$$Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}(\mathbf{X}_{J(\epsilon)} = \alpha | \mathbf{X}^{n(\epsilon)} = \mathbf{x}^{n(\epsilon)}) = \frac{|\{i \mid \mathbf{x}_i = \alpha\}|}{n(\epsilon)}.$$

Hence for all  $\mathbf{x}^{n(\epsilon)} \in B_\epsilon^{(n(\epsilon))} \subseteq A_\epsilon^{*(n(\epsilon))}(\mathbf{X})$  and for all  $\forall \alpha \in \mathcal{X}_1 \times \mathcal{X}_2$ ,

$$|Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}(\mathbf{X}_{J(\epsilon)} = \alpha | \mathbf{X}^{n(\epsilon)} = \mathbf{x}^{n(\epsilon)}) - P_{\mathbf{X}}(\alpha)| < \frac{\epsilon}{|\mathcal{X}_1||\mathcal{X}_2|}.$$

The fact that  $Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}(B_\epsilon^{(n(\epsilon))}) > 1 - \epsilon$  leads to

$$|Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}(\mathbf{X}_{J(\epsilon)} = \alpha) - P_{\mathbf{X}}(\alpha)| < \frac{\epsilon}{|\mathcal{X}_1||\mathcal{X}_2|} + \epsilon \quad \forall \alpha \in \mathcal{X}_1 \times \mathcal{X}_2,$$

which proves (9). By the uniform continuity of mutual information and entropy functions on finite-alphabet random variables, (9) implies that for  $i = 1, 2$ ,

$$|H_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}) - H_P(\mathbf{X})| < \tau_2(\epsilon) \\ |I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; V_{n(\epsilon)}) - I_P(\mathbf{X}; V_{n(\epsilon)})| < \tau_2(\epsilon) \\ |R_{X_i, J(\epsilon)|V_{n(\epsilon)}}(D_i, Q^{(\epsilon)}) - R_{X_i|V_{n(\epsilon)}}(D_i, P)| < \tau_2(\epsilon)$$

for some  $\tau_2(\epsilon)$  such that  $\lim_{\epsilon \rightarrow 0} \tau_2(\epsilon) = 0$ , and  $I_P$  and  $H_P$  are mutual information and entropy functions evaluated

according to the distribution  $P_{\mathbf{X}, V_{n(\epsilon)}} = P_{\mathbf{X}}Q_{V_{n(\epsilon)}|\mathbf{X}}$ . Hence for proving (10), it remains to show that

$$\lim_{\epsilon \rightarrow 0} \frac{H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)})}{n(\epsilon)} = H_P(\mathbf{X}).$$

By (8), for all  $\mathbf{x}^{n(\epsilon)} \in B_\epsilon^{(n(\epsilon))} \subseteq A_\epsilon^{*(n(\epsilon))}(\mathbf{X})$ ,

$$|-\frac{1}{n(\epsilon)} \log Q_{\mathbf{X}^{n(\epsilon)}}(\mathbf{x}^{n(\epsilon)}) + \frac{1}{n(\epsilon)} \log P_{\mathbf{X}^{n(\epsilon)}}(\mathbf{x}^{n(\epsilon)})| \leq a(n, \epsilon) < \epsilon$$

and

$$|\frac{1}{n(\epsilon)} \log P_{\mathbf{X}^{n(\epsilon)}}(\mathbf{x}^{n(\epsilon)}) - H_P(\mathbf{X})| < \tau_3(\epsilon)$$

for some  $\tau_3(\epsilon)$  such that  $\lim_{n \rightarrow \infty} \tau_3(\epsilon) = 0$ . Let

$$\tau_4(\epsilon) := \epsilon + \tau_3(\epsilon) + \epsilon \log |\mathcal{X}_1||\mathcal{X}_2|.$$

Since  $Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}(B_\epsilon^{(n(\epsilon))}) > 1 - \epsilon$ ,

$$|\frac{1}{n(\epsilon)} H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)}) - H_P(\mathbf{X})| < \tau_4(\epsilon),$$

which proves (10). Hence the rate vector

$$\mathbf{R} + (\tau_1(\epsilon) + \tau_2(\epsilon) + \tau_4(\epsilon)) \cdot \mathbf{1}$$

is in the achievable rate region of the Gray-Wyner problem for the distribution  $P_{\mathbf{X}}$ , which proves the desired result by letting  $\epsilon \rightarrow 0$ .  $\square$

## REFERENCES

- [1] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.:Academic, 1981.
- [2] J. C. Kieffer. Strong converses in source coding relative to a fidelity criterion. *IEEE Transactions on Information Theory*, IT-37(2):257–262, March 1991.
- [3] J. C. Kieffer. Sample converses in source coding theory. *IEEE Transactions on Information Theory*, IT-37(2):263–268, March 1991.
- [4] T. S. Han. *Information-Spectrum Method in Information Theory*. Springer-Verlag New York, LLC, 2003.
- [5] W. Gu, M. Effros, and M. Bakshi. A continuity theory for lossless source coding over networks. *Proceedings of the Allerton Conference on Communication, Control, and Computing*, September 2008.
- [6] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-21(6):629–637, November 1975.
- [7] R. M. Gray and A. D. Wyner. Source coding for a simple network. *Bell System Technical Journal*, 53(9):1681–1721, November 1974.
- [8] M. Bakshi and M. Effros. On achievable rates for multicast in the presence of side information. In *Proceedings of the IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.
- [9] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, IT-52(10):4413–4430, October 2006.
- [10] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.
- [11] R. Ahlswede, P. Gács, and J. Körner. Bounds on conditional probabilities with applications in multi-user communication. *Probability Theory and Related Fields*, 34(2):157–177, June 1976.
- [12] Y. Oohama and T. Han. Universal coding for the Slepian-wolf data compression system and the strong converse theorem. *IEEE Transactions on Information Theory*, IT-40(6):1908–1919, November 1994.
- [13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.
- [14] W. Gu. *On Achievable Rate Regions for Source Coding over Networks*. PhD thesis, California Institute of Technology, 2008.