Cascade, Triangular, and Two-Way Source Coding With Degraded Side Information at the Second User

Yeow-Khiang Chia, Haim H. Permuter, *Member, IEEE*, and Tsachy Weissman, *Senior Member, IEEE*

***Abstract*—In this paper, we consider the cascade and triangular rate-distortion problems where the same side information is available at the source node and user 1, and the side information available at user 2 is a degraded version of the side information at the source node and user 1. We characterize the rate-distortion region for these problems. For the cascade setup, we show that, at user 1, decoding and rebinning the codeword sent by the source node for user 2 is optimum. We then extend our results to the two-way cascade and triangular setting, where the source node is interested in lossy reconstruction of the side information at user 2 via a rate limited link from user 2 to the source node. We characterize the rate-distortion regions for these settings. Complete explicit characterizations for all settings are given in the quadratic Gaussian case. We conclude with two further extensions: a triangular source coding problem with a helper, and an extension of our two-way cascade setting in the quadratic Gaussian case.**

***Index Terms*—Cascade source coding, triangular source coding, two-way source coding, quadratic Gaussian, source coding with a helper.**

I. Introduction

The problem of lossy source coding through a cascade was first considered by Yamamoto [1], where a source node (node 0) sends a message to node 1, which then sends a message to node 2. Since Yamamoto's work, the cascade setting has been extended in recent years through incorporating side information at either nodes 1 or 2. This model of cascade source coding with side information has potential applications in peer-to-peer networking, such as video compression and transmission over a network, where each node may have side information, such as previous video frames, about a video to be sent from the source. In [2], Vasudevan *et al.* considered the cascade problem with side information Y at node 1 and Z at node 2, with the Markov chain $X - Z - Y$. They provided inner and outer bounds for this setup and showed that the bounds coincide for the Gaussian case. In [3], Cuff *et al.* considered the cascade problem where the side information is known only to the intermediate node and provided inner and outer bounds for this setup.

Of most relevance to this paper is the work in [4], where Permuter and Weissman considered the cascade source coding problem with side information available at both node 0 and node 1 and established the rate-distortion region for this setup. The cascade setting was then extended to the triangular source coding setting where an additional rate limited link is available from the source node to node 2.

Given the results in [4], a natural question is whether it can be extended to richer classes of cascade source coding problems. A related question is the following. The achievability scheme in the cascade result in [4] relies on node 1 decoding and retransmitting the codeword sent by node 0 to node 2. This is essentially a special case of the *decode and re-bin* scheme where node 1 decodes and re-bins the codeword sent by node 0 to node 2. When is this decode and re-bin scheme optimum and what is the statistical structure of the sources and network topology required for this scheme to be optimum? In this paper, we extend the cascade and triangular source coding setting in [4] to include additional side information Z at node 2, with the constraint that the Markov chain $X - Y - Z$ holds. Under the Markov constraint, we establish the rate-distortion regions for both the cascade and triangular settings. The cascade and triangular settings are shown in Figs. 1 and 2, respectively. In the cascade case, we show that, at node 1, the decode and re-bin scheme is optimum. To the best of our knowledge, this is the first *lossy* source coding setting where the decode and re-bin scheme at the cascade is shown to be optimum. (In [5], the decode and re-bin scheme was shown to be optimum for some classes of source statistics in a *lossless* setting.) The decode and re-bin appears to rely quite heavily on the fact that the side information at node 2 is degraded: Since node 1 can decode any codeword intended for node 2, there is no need for node 0 to send additional information for node 1 to relay to node 2 on the R_1 link. Node 0 can therefore tailor the transmission for node 1 and rely on node 1 to decode and minimize the rate required on the R_2 link. We also extend our results to two-way source coding through a cascade, where node 0 wishes to obtain a lossy version of Z through a rate limited link from node 2 to node 0. This setup generalizes the (two-rounds) two-way source coding result found in [6].[1] The two-way cascade source coding and two-way triangular source coding are given in Figs. 3 and 4, respectively.

The rest of the paper is as follows. In Section II, we provide the formal definitions and problem setup. In Section III,

Manuscript received October 18, 2010; revised March 31, 2011; accepted June 21, 2011. Date of publication September 12, 2011; date of current version January 06, 2012. This work was supported in part by the National Science Foundation (NSF) under Grants CCF-1049413 and 4101-38047, and the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant CCF-0939370. The material in this paper was presented in part at the Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2011 and the Allerton Conference on Communication, Control, and Computing, Monticello, IL, September–October, 2010.

Y.-K. Chia and T. Weissman are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: ykchia@stanford.edu; tsachy@stanford.edu).

H. H. Permuter is with the Department of Electrical and Computer Engineering, Ben Gurion University, Beer-Sheva 84105, Israel (e-mail: haimp@bgu.ac.il).

Communicated by S. Diggavi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2011.2167738

[1]Kaspi [6] considered multiple rounds. In this paper, we consider only two rounds and when we mention the results in [6], we mean the two-rounds version of the results

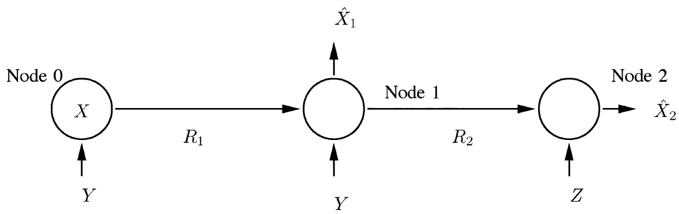

Fig. 1. Cascade source coding setting.

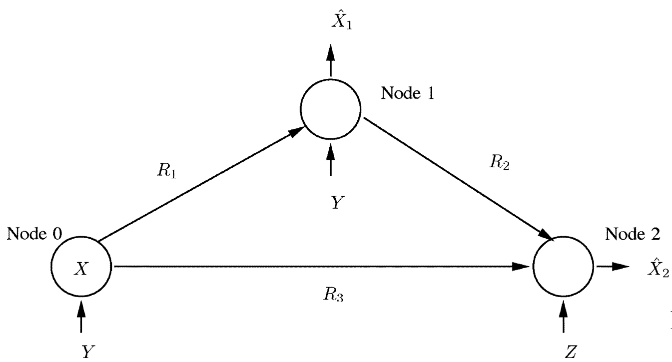

Fig. 2. Triangular source coding setting.

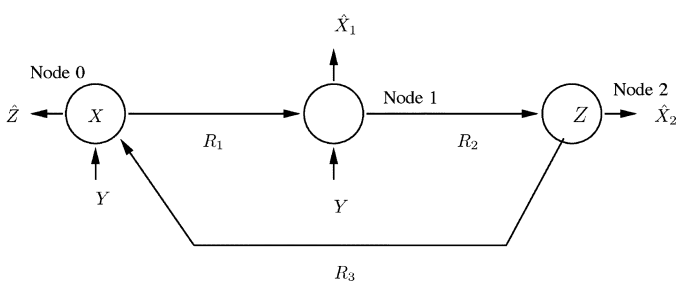

Fig. 3. Setup for two-way cascade source coding.

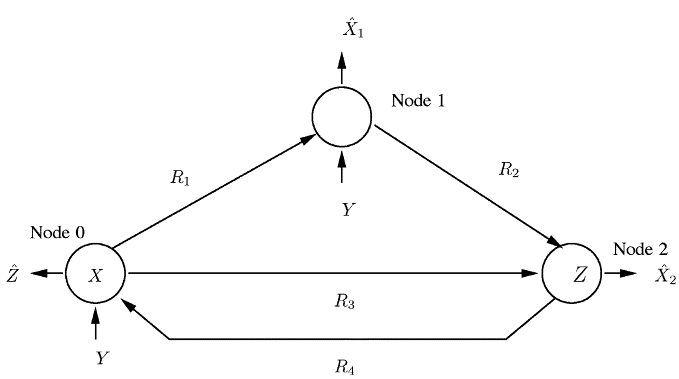

Fig. 4. Setup for two-way triangular source coding.

we present and prove our results for the aforementioned settings. In Section IV, we consider the quadratic Gaussian case. We show that Gaussian auxiliary random variables suffice to exhaust the rate-distortion regions, and their parameters may be found through solving a tractable low-dimensional optimization problem. We also showed that our quadratic Gaussian settings may be transformed into equivalent settings in [4] where explicit characterizations were given. In the quadratic Gaussian

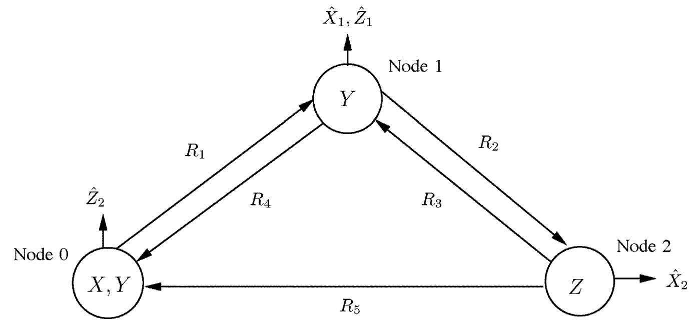

Fig. 5. Extended quadratic Gaussian two-way source coding.

case, we also extended our settings to solve a more general case of two-way cascade source coding. In Section V, we extend our triangular source coding setup to include a helper, which observes the side information Y, and has a rate limited link to node 2. Our two-way cascade quadratic Gaussian extension is shown in Fig. 5, while our helper extension is shown in Fig. 7. We conclude the paper in Section VI.

II. Problem Definition

In this section, we give formal definitions for the setups under consideration. We will follow the notation of [7, Lecture 1]. Unless otherwise stated, all logarithms in this paper are taken to base 2. The source sequences under consideration, $\{X_i \in \mathcal{X}, i = 1, 2, \dots\}$, $\{Y_i \in \mathcal{Y}, i = 1, 2, \dots\}$, and $\{Z_i \in \mathcal{Z}, i = 1, 2, \dots\}$, are drawn from finite alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. For any $i \geq 1$, the random variables (X_i, Y_i, Z_i) are independent and identically distributed according to $p(x, y, z) = p(x)p(y|x)p(z|y)$; i.e., $X - Y - Z$. The distortion measure between sequences is defined in the usual way. Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$. Then

$$d(x^n, \hat{x}^n) := \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i).$$

A. Cascade and Triangular Source Coding

We give formal definition for the triangular source coding setting (Fig. 2). The cascade setting follows from specializing the definitions for the triangular setting by setting $R_3 = 0$. A $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3})$ code for the triangular setting consists of three encoders

$$\begin{aligned} f_1 \text{ (at node 0) } &: \mathcal{X}^n \times \mathcal{Y}^n \to M_1 \in [1 : 2^{nR_1}] \\ f_2 \text{ (at node 1) } &: \mathcal{Y}^n \times [1 : 2^{nR_1}] \to M_2 \in [1 : 2^{nR_2}] \\ f_3 \text{ (at node 0) } &: \mathcal{X}^n \times \mathcal{Y}^n \to M_3 \in [1 : 2^{nR_3}] \end{aligned}$$

and two decoders

$$\begin{aligned} g_1 \text{ (at node 1) } &: \mathcal{Y}^n \times [1 : 2^{nR_1}] \to \hat{\mathcal{X}}_1^n \\ g_2 \text{ (at node 2) } &: \mathcal{Z}^n \times [1 : 2^{nR_2}] \times [1 : 2^{nR_3}] \to \hat{\mathcal{X}}_2^n. \end{aligned}$$

Given (D_1, D_2), a $(R_1, R_2, R_3, D_1, D_2)$ rate-distortion tuple for the triangular source coding setting is said to be *achievable* if, for any $\epsilon > 0$, and n sufficiently large, there exists a

$(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3})$ code for the triangular source coding setting such that

$$\mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} d_j(X_i, \hat{X}_{j,i})\right] \leq D_j + \epsilon, \qquad j = 1, 2$$

where $\hat{X}_1^n = g_1(Y^n, f_1(X^n, Y^n))$ and $\hat{X}_2^n = g_2(Z^n, f_2(Y^n, f_1(X^n, Y^n)), f_3(X^n, Y^n))$.

The *rate-distortion region* $\mathcal{R}(D_1, D_2)$ is defined as the closure of the set of all achievable rate-distortion tuples.

Cascade Source Coding: The cascade source coding setting corresponds to the case where $R_3 = 0$.

B. Two-Way Cascade and Triangular Source Coding

We give formal definitions for the more general two-way triangular source coding setting shown in Fig. 4. A $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3}, 2^{nR_4})$ code for the triangular setting consists of four encoders

$$\begin{aligned}
f_1 \text{ (at node 0) } &: \mathcal{X}^n \times \mathcal{Y}^n \to M_1 \in [1 : 2^{nR_1}] \\
f_2 \text{ (at node 1) } &: \mathcal{Y}^n \times [1 : 2^{nR_1}] \to M_2 \in [1 : 2^{nR_2}] \\
f_3 \text{ (at node 0) } &: \mathcal{X}^n \times \mathcal{Y}^n \to M_3 \in [1 : 2^{nR_3}] \\
f_4 \text{ (at node 2) } &: \mathcal{Z}^n \times [1 : 2^{nR_2}] \times [1 : 2^{nR_3}] \\
&\qquad \to M_4 \in [1 : 2^{nR_4}]
\end{aligned}$$

and three decoders

$$\begin{aligned}
g_1 \text{ (at node 1) } &: \mathcal{Y}^n \times [1 : 2^{nR_1}] \to \hat{\mathcal{X}}_1^n \\
g_2 \text{ (at node 2) } &: \mathcal{Z}^n \times [1 : 2^{nR_2}] \times [1 : 2^{nR_3}] \to \hat{\mathcal{X}}_2^n \\
g_3 \text{ (at node 0) } &: \mathcal{X}^n \times \mathcal{Y}^n \times [1 : 2^{nR_4}] \to \hat{\mathcal{Z}}^n.
\end{aligned}$$

Given (D_1, D_2, D_3), a $(R_1, R_2, R_3, R_4, D_1, D_2, D_3)$ rate-distortion tuple for the two-way triangular source coding setting is said to be *achievable* if, for any $\epsilon > 0$, and n sufficiently large, there exists a $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3}, 2^{nR_4})$ code for the two-way triangular source coding setting such that

$$\mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} d_j(X_i, \hat{X}_{j,i})\right] \leq D_j + \epsilon, \qquad j = 1, 2$$

and

$$\mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} d_3(Z_i, \hat{Z}_i)\right] \leq D_j + \epsilon$$

where $\hat{X}_1^n = g_1(Y^n, f_1(X^n, Y^n))$, $\hat{X}_2^n = g_2(Z^n, f_2(Y^n, f_1(X^n, Y^n)), f_3(X^n, Y^n))$, and $\hat{Z}^n = g_3(X^n, Y^n, f_4(Z^n, f_2(Y^n, f_1(X^n, Y^n)), f_3(X^n, Y^n)))$.

The *rate-distortion region* $\mathcal{R}(D_1, D_2, D_3)$ is defined as the closure of the set of all achievable rate-distortion tuples.

Two-Way Cascade Source Coding: The two-way cascade source coding setting corresponds to the case where $R_3 = 0$. In the special case of two-way cascade setting, we will use R_3, rather than R_4, to denote the rate from node 2 to node 0.

III. Main Results

In this section, we present our main results, which are single letter characterizations of the rate-distortion regions for the four settings introduced in Section II. The single-letter characterizations for the cascade source coding setting, triangular source coding setting, two-way cascade source coding setting, and two-way triangular source coding setting are given in Theorems 1, 2, 3, and 4, respectively. While Theorems 1–3 can be derived as special cases of Theorem 4, for clarity and to illustrate the development of the main ideas, we will present Theorems 1–4 separately. In each of the theorems, we will present a sketch of the achievability proof and proof of the converse. Details of the achievability proofs for Theorems 1–4 are given in Appendix A. Proofs of the cardinality bounds for the auxiliary random variables appearing in the theorems are given in Appendix B. In each of the theorems presented, the achievability scheme does not require the Markov structure $X - Y - Z$, and hence, they can be used even if the sources do not satisfy the Markov condition. The Markov condition is required for us to prove the converse.

A. Cascade Source Coding

Theorem 1 (Rate-Distortion Region for Cascade Source Coding): $\mathcal{R}(D_1, D_2)$ for the cascade source coding setting defined in Section II is given by the set of all rate tuples (R_1, R_2) satisfying

$$\begin{aligned}
R_2 &\geq I(U; X, Y|Z), \\
R_1 &\geq I(X; \hat{X}_1, U|Y)
\end{aligned}$$

for some $p(x, y, z, u, \hat{x}_1) = p(x)p(y|x)p(z|y)p(u|x, y) p(\hat{x}_1|x, y, u)$ and function $g_2 : \mathcal{U} \times \mathcal{Z} \to \hat{X}_2$ such that

$$\mathsf{E}d_j(X, \hat{X}_j) \leq D_j, \qquad j = 1, 2.$$

The cardinality of $\mathcal{U}$ is upper bounded by $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}| + 3$.

If $Z = \emptyset$, this region reduces to the cascade source coding region given in [4]. If $Y = X$, this setup reduces to the well-known Wyner–Ziv setup [8].

The coding scheme follows from a combination of techniques used in [4] and a new idea of decoding and re-binning at the cascade node (node 1). Node 0 generates a description U^n intended for nodes 1 and 2. Node 1 decodes U^n and then re-bins it to reduce the rate of communicating U^n to node 2 based on its side information. In addition, node 0 generates $\hat{X}_1^n$ to satisfy the distortion requirement at node 1. We now give a sketch of achievability and a proof of the converse.

Sketch of Achievability: We first generate $2^{n(I(X,Y;U)+\epsilon)}$ U^n sequences according to $\prod_{i=1}^n p(u_i)$. For each u^n and y^n sequences, we generate $2^{n(I(\hat{X}_1^n;X|U,Y)+\epsilon)}$ $\hat{X}_1^n$ sequences according to $\prod_{i=1}^n p(\hat{x}_i|u_i, y_i)$. Partition the set of U^n sequences into $2^{n(I(U;X|Y)+2\epsilon)}$ bins, $\mathcal{B}_1(m_{10})$. Separately and independently, partition the set of U^n sequences into $2^{n(I(U;X,Y|Z)+2\epsilon)}$ bins, $\mathcal{B}_2(m_2)$, $m_2 \in [1 : 2^{n(I(U;X,Y|Z)+2\epsilon)}]$.

Given x^n, y^n, node 0 looks for a jointly typical codeword u^n; that is, $(u^n, x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$. If there are more than one, it selects a codeword uniformly at random from the set of jointly typical codewords. This operation succeeds with high probability since there are $2^{n(I(X,Y;U)+\epsilon)}$ U^n sequences. Node 0 then looks for a $\hat{x}_1^n$ that is jointly typical with u^n, x^n, y^n. This operation succeeds with high probability since there are $2^{n(I(\hat{X}_1;X|U,Y)+\epsilon)}$

$\hat{x}_1^n$ sequences. Node 0 then sends out the bin index m_{10} such that $u^n \in \mathcal{B}_1(m_{10})$ and the index corresponding to $\hat{x}_1^n$. This requires a total rate of $R_1 = I(U;X|Y) + I(\hat{X}_1;X|U,Y) + 3\epsilon$.

At node 1, it recovers u^n by looking for the unique u^n sequence in $\mathcal{B}_1(m_{10})$ such that $(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$. Since there are only $2^{n(I(X,Y;U)-I(U;X|Y)-\epsilon)} = 2^{n(I(U;Y)-\epsilon)}$ sequences in the bin, this operation succeeds with high probability. Node 1 reconstructs x^n as $\hat{x}_1^n$. Node 1 then sends out m_2 such that $u^n \in \mathcal{B}_2(m_2)$. This requires a rate of $R_2 = I(U;X,Y|Z)+2\epsilon$.

At node 2, note that since $U - (X,Y) - Z$, the sequences (U^n, X^n, Y^n, Z^n) are jointly typical with high probability. Node 2 looks for the unique u^n in $\mathcal{B}_2(m_2)$ such that $(u^n, z^n) \in \mathcal{T}_\epsilon^{(n)}$. From the Markov chain $U-(X,Y)-Z$, $I(U;X,Y) - I(U;X,Y|Z) = I(U;Z)$. Hence, this operation succeeds with high probability since there are only $2^{n(I(U;Z)-\epsilon)}$ u^n sequences in the bin. It then reconstructs using $\hat{x}_{2i} = g_2(u_i, z_i)$ for $i \in [1:n]$.

Proof of Converse: Given a $(n, 2^{nR_1}, 2^{nR_2}, D_1, D_2)$ code, define $U_i = (X^{i-1}, Y^{i-1}, Z^{i-1}, Z_{i+1}^n, M_2)$. We have the following:

$$
\begin{aligned}
nR_2 &\geq H(M_2) \\
&\geq H(M_2|Z^n) \\
&= I(X^n, Y^n; M_2|Z^n) \\
&= \sum_{i=1}^n I(X_i, Y_i; M_2|Z^n, X^{i-1}, Y^{i-1}) \\
&= \sum_{i=1}^n \big(H(X_i, Y_i|Z^n, X^{i-1}, Y^{i-1}) \\
&\qquad - H(X_i, Y_i|Z^n, X^{i-1}, Y^{i-1}, M_2)\big) \\
&= \sum_{i=1}^n (H(X_i, Y_i|Z_i) - H(X_i, Y_i|Z_i, U_i)) \\
&= \sum_{i=1}^n I(X_i, Y_i; U_i|Z_i).
\end{aligned}
$$

Next

$$
\begin{aligned}
nR_1 &\geq H(M_1) \\
&\geq H(M_1|Y^n, Z^n) \\
&= H(M_1, M_2|Y^n, Z^n) \\
&= I(X^n; M_1, M_2|Y^n, Z^n) \\
&= \sum_{i=1}^n I(X_i; M_1, M_2|X^{i-1}, Y^n, Z^n) \\
&= \sum_{i=1}^n \big(H(X_i|X^{i-1}, Y^n, Z^n) \\
&\qquad - H(X_i|X^{i-1}, Y^n, Z^n, M_1, M_2)\big) \\
&= \sum_{i=1}^n (H(X_i|Y_i, Z_i) \\
&\qquad - H(X_i|X^{i-1}, Y^n, Z^n, M_1, M_2)) \\
&\overset{(a)}{=} \sum_{i=1}^n \Big(H(X_i|Y_i) - H(X_i|X^{i-1}, Y^n, \hat{X}_{1i}, Z^n, M_1, M_2)\Big) \\
&\geq \sum_{i=1}^n \Big(H(X_i|Y_i) - H(X_i|\hat{X}_{1i}, Y_i, U_i)\Big) \\
&= \sum_{i=1}^n I(X_i; \hat{X}_{1i}, U_i|Y_i).
\end{aligned}
$$

Step (a) follows from the Markov assumption $X-Y-Z$ and the fact that $\hat{X}_{1i}$ is a function of (Y^n, M_2). Next, let Q be a random variable uniformly distributed over $[1:n]$ and independent of (X^n, Y^n, Z^n). We note that $X_Q = X$, $Y_Q = Y$, $Z_Q = Z$, and

$$
\begin{aligned}
R_2 &\geq I(X_Q, Y_Q; U_Q|Q, Z_Q) \\
&= I(X_Q, Y_Q; U_Q, Q|Z_Q) \\
&= I(X, Y; U_Q, Q|Z), \\
R_1 &\geq I(X_Q; \hat{X}_{1Q}, U_Q|Y_Q, Q) \\
&= I(X; \hat{X}_{1Q}, U_Q, Q|Y).
\end{aligned}
$$

Defining $U = (U_Q, Q)$ and $\hat{X}_{1Q} = \hat{X}_1$ then completes the proof. The existence of the reconstruction function g_2 follows from the definition of U. The Markov chains $U - (X,Y) - Z$ and $Z - (U,X,Y) - \hat{X}_1$ required to factor the probability distribution stated in the theorem also follow from definitions of U and $\hat{X}_1$. □

We now extend Theorem 1 to the triangular source coding setting.

B. Triangular Source Coding

Theorem 2 (Rate-Distortion Region for Triangular Source Coding): $\mathcal{R}(D_1, D_2)$ for the triangular source coding setting defined in Section II is given by the set of all rate tuples (R_1, R_2, R_3) satisfying

$$
\begin{aligned}
R_1 &\geq I(X; \hat{X}_1, U|Y) \\
R_2 &\geq I(X, Y; U|Z) \\
R_3 &\geq I(X, Y; V|U, Z)
\end{aligned}
$$

for some $p(x,y,z,u,v,\hat{x}_1) = p(x)p(y|x)p(z|y)p(u|x,y)$ $p(\hat{x}_1|x,y,u)p(v|x,y,u)$ and function $g_2 : \mathcal{U} \times \mathcal{V} \times \mathcal{Z} \to \hat{\mathcal{X}}_2$ such that

$$
\mathsf{E}d_j(X, \hat{X}_j) \leq D_j, \qquad j = 1,2.
$$

The cardinalities for the auxiliary random variables can be upper bounded by $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}|+4$ and $|\mathcal{V}| \leq (|\mathcal{X}||\mathcal{Y}|+4)(|\mathcal{X}||\mathcal{Y}|+1)$.

If $Z = \emptyset$, this region reduces to the triangular source coding region given in [4].

The proof of the triangular case follows that of the cascade case, with the additional step of node 0 generating an additional description V^n that is intended for node 2. This description is then binned to reduce the rate, with the side information at node 2 being U^n and Z^n. Node 2 first decodes U^n and then V^n.

Sketch of Achievability: The Achievability proof is an extension of that in Theorem 1. The additional step we have here is that we generate $2^{n(I(V;X,Y|U)+\epsilon)}$ V^n sequences according to $\prod_{i=1}^n p(v_i|u_i)$ for each u^n sequence, and bin these sequences to $2^{n(I(V;X,Y|U,Z)+2\epsilon)}$ bins, $\mathcal{B}_3(m_3)$, $m_3 \in [1:2^{nR_3}]$. To send from node 0 to node 2, node 0 first finds a v^n sequence that is jointly typical with (u^n, x^n, y^n). This operation succeeds with high probability since we have $2^{n(I(V;X,Y|U)+\epsilon)}$ v^n sequences. We then send out m_3, the bin number for v^n. At node 2, from the probability distribution, we have the Markov chain $(V,U) - (X,Y) - Z$. Hence, the sequences are jointly typical with high probability. Node 2 reconstructs by looking for unique $v^n \in \mathcal{B}_3(m_3)$ such that (u^n, v^n, z^n) are jointly typical. This operation succeeds with high probability since the number of sequences in $\mathcal{B}_3(m_3)$ is $2^{n(I(V;Z|U)-\epsilon)}$. Node 2 then reconstructs using the function g_2.

Proof of Converse: The converse is proved in two parts. In the first part, we derive the required inequalities and in the second part, we show that the joint probability distribution can be restricted to the form stated in the theorem.

Given a $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3}, D_1, D_2)$ code, define $U_i = (X^{i-1}, Y^{i-1}, Z^{i-1}, Z_{i+1}^n, M_2)$ and $V_i = (U_i, M_3)$. We omit proof of the R_1 and R_2 inequalities since it follows the same steps as in Theorem 1. We have

$$nR_1 \geq \sum_{i=1}^{n} I(X_i; \hat{X}_{1i}, U_i|Y_i)$$
$$nR_2 \geq \sum_{i=1}^{n} I(X_i, Y_i; U_i|Z_i).$$

For R_3, we have

$$\begin{aligned} nR_3 &\geq H(M_3) \\ &\geq H(M_3|M_2, Z^n) \\ &= I(X^n, Y^n; M_3|M_2, Z^n) \\ &= \sum_{i=1}^{n} \big(H(X_i, Y_i|M_2, Z^n, X^{i-1}, Y^{i-1}) \\ &\qquad - H(X_i, Y_i|M_2, M_3, Z^n, X^{i-1}, Y^{i-1})\big) \\ &= \sum_{i=1}^{n} (H(X_i, Y_i|U_i, Z_i) - H(X_i, Y_i|U_i, V_i, Z_i)) \\ &= \sum_{i=1}^{n} I(X_i, Y_i; V_i|U_i, Z_i). \end{aligned}$$

Next, let Q be a random variable uniformly distributed over $[1 : n]$ and independent of (X^n, Y^n, Z^n). Defining $U = (U_Q, Q)$, $V = (V_Q, Q)$, and $\hat{X}_{1Q} = \hat{X}_1$ then gives us the bounds stated in Theorem 2. The existence of the reconstruction function g_2 follows from the definition of U and V. Next, from the definitions of U, V and $\hat{X}_1$, we note the following Markov relation: $(U, V, \hat{X}_1) - (X, Y) - Z$. The joint probability distribution can then be factored as $p(x, y, z, u, v, \hat{x}_1) = p(x, y, z)p(u|x, y)p(\hat{x}_1, v|x, y, u)$.

We now show that it suffices to restrict the joint probability distributions to the form $p(x, y, z)p(u|x, y)$ $p(\hat{x}_1|x, y, u)p(v|x, y, u)$ using a method in [4, Lemma 5]. The basic idea is that since the inequalities derived rely on $p(\hat{x}_1, v|x, y, u)$ only through the marginals $p(\hat{x}_1|x, y, u)$ and $p(v|x, y, u)$, we can obtain the same bounds even when the probability distribution is restricted to the form $p(x, y, z)p(u|x, y)p(\hat{x}_1|x, y, u)p(v|x, y, u)$.

Fix a joint distribution $p(x, y, z)p(u|x, y)p(\hat{x}_1, v|x, y, u)$ and let $\hat{p}(v|x, y, u)$ and $\hat{p}(\hat{x}_1|x, y, u)$ be the induced conditional distributions. Note that $p(x, y, z)p(u|x, y)p(\hat{x}_1, v|x, y, u)$ and $p(x, y, z)p(u|x, y)\hat{p}(\hat{x}_1|x, y, u)\hat{p}(v|x, y, u)$ have the same marginals $p(x, y, z, u, v)$ and $p(x, y, z, u, \hat{x}_1)$, and the Markov condition $(U, V, \hat{X}_1) - (X, Y) - Z$ continues to hold under $p(x, y, z)p(u|x, y)\hat{p}(\hat{x}_1|x, y, u)\hat{p}(v|x, y, u)$.

Finally, note that the rate and distortion constraints given in Theorem 2 depends on the joint distribution only through the marginals $p(x, y, z, u, v)$ and $p(x, y, z, u, \hat{x}_1)$. It therefore suffices to restrict the probability distributions to the form $p(x, y, z)p(u|x, y)\hat{p}(\hat{x}_1|x, y, u)\hat{p}(v|x, y, u)$. □

C. Two-Way Cascade Source Coding

We now extend the source coding settings to include the case where node 0 requires a lossy version of Z. We first consider the two-way cascade source coding setting defined in Section II (we will use R_3 to denote the rate on the link from node 2 to node 0). In the forward part, the achievable scheme consists of using the achievable scheme for the cascade source coding case. Node 2 then sends back a description of Z^n to node 0, with X^n, Y^n, U_1^n as side information at node 0. For the converse, we rely on the techniques introduced and also on a technique for establishing Markovity of random variables found in [6].

Theorem 3 (Rate-Distortion Region for Two-Way Cascade Source Coding): $\mathcal{R}(D_1, D_2, D_3)$ for two-way cascade source coding is given by the set of all rate tuples (R_1, R_2, R_3) satisfying

$$\begin{aligned} R_1 &\geq I(X; \hat{X}_1, U_1|Y) \\ R_2 &\geq I(U_1; X, Y|Z) \\ R_3 &\geq I(U_2; Z|U_1, X, Y) \end{aligned}$$

for some $p(x, y, z, u_1, u_2, \hat{x}_1) = p(x)p(y|x)p(z|y)p(u_1|x, y)$ $p(\hat{x}_1|u_1, x, y)p(u_2|z, u_1)$ and functions $g_2 : \mathcal{U}_1 \times \mathcal{Z} \to \hat{X}_2$ and $g_3 : \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X} \times \mathcal{Y} \to \hat{Z}$ such that

$$\begin{aligned} \mathsf{E}(d_j(X, \hat{X}_j)) &\leq D_j, \qquad j = 1, 2 \\ \mathsf{E}(d_3(Z, \hat{Z})) &\leq D_3. \end{aligned}$$

The cardinalities for the auxiliary random variables can be upper bounded by $|\mathcal{U}_1| \leq |\mathcal{X}||\mathcal{Y}| + 5$ and $|\mathcal{U}_2| \leq |\mathcal{U}_1|(|\mathcal{Z}| + 1)$.

If $Y = X$, this region reduces to the result for two-way (two rounds only) source coding found in [6].

Sketch of Achievability: The forward path (R_1 and R_2) follows from the cascade source coding case in Theorem 1. The reverse direction follows by the following. For each u_1^n, we generate $2^{n(I(U_2;Z|U_1)+\epsilon)}$ u_2^n sequences according to $\prod_{i=1}^{n} p(u_{2i}|u_{1i})$ and bin them to $2^{n(I(U_2;Z|U_1,X,Y)+2\epsilon)}$ bins, $\mathcal{B}_3(m_3)$, $m_3 \in [1 : 2^{nR_3}]$. Node 2 finds a u_2^n sequence that is jointly typical with (u_1^n, z^n). Since there are $2^{n(I(U_2;Z|U_1)+\epsilon)}$ sequences, this operation succeeds with high probability. It then sends out the bin index m_3, which the jointly typical v^n sequence is in. At node 0, it recovers u_2^n by looking for the unique sequence in $\mathcal{B}_3(m_3)$ such that (u_1^n, u_2^n, x^n, y^n) are jointly typical. From the Markov condition $U_2 - (U_1, Z) - (X, Y)$ and the Markov lemma [9], the sequences are jointly typical with high probability. Next, since there are only $2^{n(I(U_2;X,Y|U_1)-\epsilon)}$ sequences in the bin, the probability that we do not find the unique (correct) sequence goes to zero with n. Finally, node 0 reconstructs using the function g_3.

Proof of Converse: Given a $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3}, D_1, D_2, D_3)$ code, define $U_{1i} = (M_2, X^{i-1}, Y^{i-1}, Z_{i+1}^n)$ and $U_{2i} = M_3$. Note that unlike Theorems 1 and 2, U_{1i} does not contain Z^{i-1}. We have

$$\begin{aligned}
nR_1 &\geq H(M_1) \\
&\geq H(M_1|Y^n, Z^n) \\
&= H(M_1, M_2|Y^n, Z^n) \\
&= I(X^n; M_1, M_2|Y^n, Z^n) \\
&= \sum_{i=1}^n I(X_i; M_1, M_2|X^{i-1}, Y^n, Z^n) \\
&= \sum_{i=1}^n \big(H(X_i|X^{i-1}, Y^n, Z^n) \\
&\qquad - H(X_i|X^{i-1}, Y^n, Z^n, M_1, M_2)\big) \\
&= \sum_{i=1}^n \big(H(X_i|Y_i, Z_i) - H(X_i|X^{i-1}, Y^n, Z^n, M_1, M_2)\big) \\
&\overset{(a)}{=} \sum_{i=1}^n \big(H(X_i|Y_i) - H(X_i|X^{i-1}, Y^n, Z^n, M_1, M_2)\big) \\
&\overset{(b)}{=} \sum_{i=1}^n \Big(H(X_i|Y_i) - H(X_i|X^{i-1}, \hat{X}_{1i}, Y^n, Z^n, M_1, M_2)\Big) \\
&\geq \sum_{i=1}^n \Big(H(X_i|Y_i) - H(X_i|\hat{X}_{1i}, Y_i, U_{1i})\Big) \\
&= \sum_{i=1}^n I(X_i; \hat{X}_{1i}, U_{1i}|Y_i)
\end{aligned}$$

where step (a) follows from the Markov assumption $X_i - Y_i - Z_i$ and step (b) follows from $\hat{X}_{1i}$ being a function of (Y^n, M_1).

Consider now R_2

$$\begin{aligned}
nR_2 &= H(M_2) \\
&\geq H(M_2|Z^n) \\
&= I(M_2; X^n, Y^n|Z^n) \\
&= \sum_{i=1}^n \big(H(X_i, Y_i|Z^n, X^{i-1}, Y^{i-1}) \\
&\qquad - H(X_i, Y_i|Z^n, X^{i-1}, Y^{i-1}, M_2)\big) \\
&\geq \sum_{i=1}^n I(X_i, Y_i; U_{1i}|Z_i).
\end{aligned}$$

Next, consider R_3

$$\begin{aligned}
nR_3 &= H(M_3) \\
&\geq H(M_3|X^n, Y^n) \\
&\geq I(M_3; Z^n|X^n, Y^n) \\
&= H(Z^n|X^n, Y^n) - H(Z^n|X^n, Y^n, M_3) \\
&= H(Z^n|X^n, Y^n) - H(Z^n|X^n, Y^n, M_2, M_3) \\
&\geq \sum_{i=1}^n \big(H(Z_i|X_i, Y_i) - H(Z_i|Z_{i+1}^n, X^i, Y^i, M_2, M_3)\big) \\
&= \sum_{i=1}^n I(Z_i; U_{1i}, U_{2i}|X_i, Y_i) \\
&= \sum_{i=1}^n I(Z_i; U_{2i}|X_i, Y_i, U_{1i})
\end{aligned}$$

where the last step follows from the Markov relation $Z_i - (X_i, Y_i) - U_{1i}$ which we will now prove, together with other Markov relations between the random variables. The first two Markov relations below are used for factoring the joint probability distribution while Markov relations three and four are used for establishing the distortion constraints. We will use the following lemma from [6].

Lemma 1: Let A_1, A_2, B_1, B_2 be random variables with joint probability mass functions mf $p(a_1, a_2, b_1, b_2) = p(a_1, b_1)p(a_2, b_2)$. Let $\tilde{M}_1$ be a function of (A_1, A_2) and $\tilde{M}_2$ be a function of $(B_1, B_2, \tilde{M}_1)$. Then

$$I(A_2; B_1|\tilde{M}_1, \tilde{M}_2, A_1, B_2) = 0 \tag{1}$$

$$I(B_1; \tilde{M}_1|A_1, B_2) = 0 \tag{2}$$

$$I(A_2; \tilde{M}_2|\tilde{M}_1, A_1, B_2) = 0. \tag{3}$$

Now, let us show the following Markov relations.

1) $Z_i - (X_i, Y_i) - (U_{1i}, \hat{X}_{1i})$: To establish this relation, we show that $I(Z_i; U_{1i}, \hat{X}_{1i}|X_i, Y_i) = 0$

$$\begin{aligned}
&I(Z_i; \hat{X}_{1i}, U_{1i}|X_i, Y_i) \\
&= I(Z_i; \hat{X}_{1i}, M_2, X^{i-1}, Y^{i-1}, Z_{i+1}^n|X_i, Y_i) \\
&\leq I(Z_i; \hat{X}_{1i}, M_2, X^{i-1}, Y^{i-1}, X_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n|X_i, Y_i) \\
&= I(Z_i; X^{i-1}, Y^{i-1}, X_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n|X_i, Y_i) \\
&= 0.
\end{aligned}$$

2) $U_{2i} - (Z_i, U_{1i}) - (\hat{X}_{1i}, X_i, Y_i)$: Note that $U_{2i} = M_3$. Consider

$$\begin{aligned}
&I(\hat{X}_i, X_i, Y_i; U_{2i}|Z_i, U_{1i}) \\
&\leq I(\hat{X}_i, X_i^n, Y_i^n; M_3|Z_i^n, X^{i-1}, Y^{i-1}, M_2) \\
&= I(X_i^n, Y_i^n; M_3|Z_i^n, X^{i-1}, Y^{i-1}, M_2).
\end{aligned}$$

Now, using Lemma 1, set $A_1 = (X^{i-1}, Y^{i-1})$, $B_1 = Z^{i-1}$, $A_2 = (X_i^n, Y_i^n)$, $B_2 = (Z_i^n)$, $\tilde{M}_2 = M_3$, and $\tilde{M}_1 = M_2$. Then, using the third expression in the Lemma, we see that $I(X_i^n, Y_i^n; M_3|Z_i^n, X^{i-1}, Y^{i-1}, M_2) = 0$.

3) $Z^{i-1} - (U_{1i}, Z_i) - (X_i, Y_i)$: Consider

$$\begin{aligned}
&I(X_i, Y_i; Z^{i-1}|U_{1i}, Z_i) \\
&\leq I(X_i^n, Y_i^n; Z^{i-1}|X^{i-1}, Y^{i-1}, Z_i^n, M_2) \\
&= H(Z^{i-1}|X^{i-1}, Y^{i-1}, Z_i^n, M_2) \\
&\qquad - H(Z^{i-1}|X^n, Y^n, Z_i^n, M_2) \\
&\leq H(Z^{i-1}|X^{i-1}, Y^{i-1}, Z_i^n) \\
&\qquad - H(Z^{i-1}|X^n, Y^n, Z_i^n) \\
&= H(Z^{i-1}|X^{i-1}, Y^{i-1}) - H(Z^{i-1}|X^{i-1}, Y^{i-1}) \\
&= 0.
\end{aligned}$$

4) $(X_{i+1}^n, Y_{i+1}^n) - (U_{1i}, U_{2i}, X_i, Y_i) - Z_i$: Consider

$$\begin{aligned}
&I(X_{i+1}^n, Y_{i+1}^n; Z_i|U_{1i}, U_{2i}, X_i, Y_i) \\
&\leq I(X_{i+1}^n, Y_{i+1}^n; Z^i|M_2, M_3, Z_{i+1}^n, X^i, Y^i).
\end{aligned}$$

Applying the first expression in the lemma with $A_2 = (X_{i+1}^n, Y_{i+1}^n)$, $A_1 = (X^i, Y^i)$, $B_1 = Z^i$, and $B_2 = Z_{i+1}^n$ gives $I(X_{i+1}^n, Y_{i+1}^n; Z_i|U_{1i}, U_{2i}, X_i, Y_i) = 0$.

Distortion constraints: We show that the auxiliary definitions satisfy the distortion constraints by showing the existence of functions $\hat{x}_{2i}^*(U_{1i}, Z_i)$ and $\hat{z}_i^*(U_{1i}, U_{2i}, X_i, Y_i)$ such that

$$\begin{aligned}&\mathsf{E}(d_2(X_i, \hat{x}_{2i}^*(U_{1i}, Z_i)))\\ &\leq \mathsf{E}(d_2(X_i, \hat{x}_{2i}(M_2, Z^n))) \quad (4)\\ &\mathsf{E}(d_3(Z_i, \hat{z}_i^*(U_{1i}, U_{2i}, X_i, Y_i)))\\ &\leq \mathsf{E}(d_3(X_i, \hat{z}_{3i}(M_3, X^n, Y^n, Z^n))) \quad (5)\end{aligned}$$

where $\hat{x}_{2i}(M_2, Z^n)$ and $\hat{z}_i(M_3, X^n, Y^n)$ are the original reconstruction functions.

To prove the first expression (4), we have

$$\begin{aligned}&\mathsf{E}(d_2(X_i, \hat{x}_{2i}(M_2, Z^n)))\\ &= \sum p(x^i, y^i, z^n, m_2) d_2(x_i, \hat{x}_{2i}(m_2, z^n))\\ &\overset{(a)}{=} \sum \big(p(u_{1i}, z^i) p(x_i, y_i|u_{1i}, z^i).\\ &\qquad d_2(x_i, \hat{x}'_{2i}(u_{1i}, z_i, z^{i-1}))\big)\\ &= \sum \big(p(u_{1i}, z_i, z^{i-1}) p(x_i, y_i|u_{1i}, z_i).\\ &\qquad d_2(x_i, \hat{x}'_{2i}(u_{1i}, z_i, z^{i-1}))\big)\end{aligned}$$

where (a) follows from defining $\hat{x}'_{2i}(u_{1i}, z_i, z^{i-1}) = \hat{x}_{2i}(m_2, z^n)$ for all x^{i-1}, y^{i-1} and the last step follows from the Markov relation $Z^{i-1} - (U_{1i}, Z_i) - (X_i, Y_i)$. Finally, defining $(z^{i-1})^* = \arg\min_{z^{i-1}} \sum_{x_i, y_i} p(x_i, y_i|u_{1i}, z_i) d_2(x_i, \hat{x}'_{2i}(u_{1i}, z_i, z^{i-1}))$ and $\hat{x}_{2i}^*(u_{1i}, z_i) = \hat{x}'_{2i}(u_{1i}, z_i, (z^{i-1})^*)$ gives us

$$\begin{aligned}&\mathsf{E}(d_2(X_i, \hat{x}_{2i}(M_2, Z^n)))\\ &= \sum p(u_{1i}, z_i, z^{i-1}) \left(\sum_{x_i, y_i} (p(x_i, y_i|u_{1i}, z_i)\right.\\ &\qquad \left. d_2(x_i, \hat{x}'_{2i}(u_{1i}, z_i, z^{i-1}))) \right)\\ &\geq \sum p(u_{1i}, z_i, z^{i-1}) d_2(x_i, \hat{x}_{2i}^*(u_{1i}, z_i))\\ &= \mathsf{E}(d_2(X_i, \hat{x}_{2i}^*(U_{1i}, Z_i))).\end{aligned}$$

To prove the second expression (5), we follow similar steps. Considering the expected distortion, we have

$$\begin{aligned}&\mathsf{E}(d_3(Z_i, \hat{z}_i(M_3, X^n, Y^n)))\\ &= \sum p(z_i^n, x^n, y^n, m_3) d_3(z_i, \hat{z}_i(m_3, x^n, y^n))\\ &= \sum \big(p(u_{1i}, u_{2i}, x_i, y_i, x_{i+1}^n, y_{i+1}^n).\\ &\qquad p(z_i|u_{1i}, u_{2i}, x_i, y_i, x_{i+1}^n, y_{i+1}^n).\\ &\qquad d_3(z_i, \hat{z}'_{3i}(u_{1i}, u_{2i}, x_i, y_i, x_{i+1}^n, y_{i+1}^n))\big)\\ &= \sum \big(p(u_{1i}, u_{2i}, x_i, y_i, x_{i+1}^n, y_{i+1}^n).\\ &\qquad p(z_i|u_{1i}, u_{2i}, x_i, y_i).\\ &\qquad d_3(z_i, \hat{z}'_i(u_{1i}, u_{2i}, x_i, y_i, x_{i+1}^n, y_{i+1}^n))\big)\end{aligned}$$

where the last step uses Markov relation 4. The rest of the proof is omitted since it uses the same steps as the proof for the first distortion constraint.

Finally, using the standard time sharing random variable Q as before and defining $U_1 = (U_{1Q}, Q)$, $U_2 = U_{2Q}$,, and $\hat{X}_1 = \hat{X}_{1Q}$, we obtain the required outer bound for the rate-distortion region. The bound for the distortions follows from defining inequalities 4 and 5. We show the rest of the proof for D_2 and omit the proof for D_3 since it follows similar steps. Defining $\hat{x}_2^*(u_1, z_i) = \hat{x}_{2Q}^*(u_{1Q}, z_i)$, we have

$$\begin{aligned}\mathsf{E}(d_2(X, \hat{x}_2^*(U_1, Z))) &= \mathsf{E}_Q \mathsf{E}(d_2(X, \hat{x}_2^*(U_1, Z))|Q)\\ &= \frac{1}{n}\sum_{i=1}^n \mathsf{E}(d_2(X_i, \hat{x}_{2i}^*(U_{1i}, Z_i)))\\ &\leq \frac{1}{n}\sum_{i=1}^n \mathsf{E}(d_2(X_i, \hat{x}_{2i}(M_2, Z^n))\\ &\leq D_2. \qquad \square\end{aligned}$$

We now turn to the final case of two-way triangular source coding.

D. Two-Way Triangular Source Coding

Theorem 4 (Rate-Distortion Region for Two-Way Triangular Source Coding): $\mathcal{R}(D_1, D_2, D_3)$ for two-way triangular source coding is given by the set of all rate tuples (R_1, R_2, R_3, R_4) satisfying

$$\begin{aligned}R_1 &\geq I(X; \hat{X}_1, U_1|Y) \quad (6)\\ R_2 &\geq I(X, Y; U_1|Z) \quad (7)\\ R_3 &\geq I(X, Y; V|Z, U_1) \quad (8)\\ R_4 &\geq I(U_2; Z|U_1, V, X, Y) \quad (9)\end{aligned}$$

for some $p(x, y, z, u_1, u_2, v, \hat{x}_1) = p(x)p(y|x)p(z|y) p(u_1|x, y)p(\hat{x}_1|x, y, u_1).\ p(v|x, y, u_1)p(u_2|z, u_1, v)$ and functions $g_2 : \mathcal{U}_1 \times \mathcal{V} \times \mathcal{Z} \to \hat{X}_2$ and $g_3 : \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{V} \times \mathcal{X} \times \mathcal{Y} \to \hat{Z}$ such that

$$\begin{aligned}\mathsf{E}(d_1(X, \hat{X}_1)) &\leq D_1 \quad (10)\\ \mathsf{E}(d_2(X, \hat{X}_2)) &\leq D_2 \quad (11)\\ \mathsf{E}(d_3(Z, \hat{Z})) &\leq D_3. \quad (12)\end{aligned}$$

The cardinalities for the auxiliary random variables are upper bounded by $|\mathcal{U}_1| \leq |\mathcal{X}||\mathcal{Y}| + 6$, $|\mathcal{V}| \leq |\mathcal{U}_1|(|\mathcal{X}||\mathcal{Y}| + 3)$, and $|\mathcal{U}_2| \leq |\mathcal{U}_1||\mathcal{V}|(|\mathcal{Z}| + 1)$.

Sketch of Achievability: The forward direction (R_1, R_2, R_3) for two-way triangular source coding follows the procedure in Theorem 2. For the reverse direction (R_4), it follows Theorem 3 with (U_1, V) replacing the role of U_1 in Theorem 3.

Proof of Converse: Given a $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3}, 2^{nR_4}, D_1, D_2, D_3)$ code, define $U_{1i} = (M_2, X^{i-1}, Y^{i-1}, Z_{i+1}^n)$, $U_{2i} = M_4$, and $V_i = (M_3, U_{1i})$. The R_1 and R_2 bounds follow the same steps as in Theorem 3. For R_3, we have

$$\begin{aligned}nR_3 &\geq H(M_3)\\ &\geq H(M_3|M_2, Z^n)\end{aligned}$$

$$
\begin{aligned}
&= I(X^n, Y^n; M_3|M_2, Z^n) \\
&= \sum_{i=1}^{n} \big(H(X_i, Y_i|M_2, Z^n, X^{i-1}, Y^{i-1}) \\
&\qquad -H(X_i, Y_i|M_2, M_3, Z^n, X^{i-1}, Y^{i-1})\big) \\
&\geq \sum_{i=1}^{n} (H(X_i, Y_i|U_i, Z_i) - H(X_i, Y_i|U_{1i}, V_i, Z_i)) \\
&= \sum_{i=1}^{n} I(X_i, Y_i; V_i|U_{1i}, Z_i).
\end{aligned}
$$

Next, consider

$$
\begin{aligned}
nR_4 &= H(M_4) \\
&\geq H(M_4|X^n, Y^n) \\
&\geq I(M_4; Z^n|X^n, Y^n) \\
&= H(Z^n|X^n, Y^n) - H(Z^n|X^n, Y^n, M_4) \\
&= H(Z^n|X^n, Y^n) - H(Z^n|X^n, Y^n, M_2, M_3, M_4) \\
&\geq \sum_{i=1}^{n} H(Z_i|X_i, Y_i) - H(Z_i|Z_{i+1}^n, X^i, Y^i, M_2, M_3, M_4) \\
&= \sum_{i=1}^{n} I(Z_i; U_{1i}, V_i, U_{2i}|X_i, Y_i) \\
&= \sum_{i=1}^{n} I(Z_i; U_{2i}|X_i, Y_i, V_i, U_{1i})
\end{aligned}
$$

where the last step follows from the Markov relation $Z_i - (X_i, Y_i) - (V_i, U_{1i})$ which we will now prove together with other Markov relations between the random variables. The first two Markov relations are for factoring the probability distribution while Markov relations 3 and 4 are for establishing the distortion constraints.

Markov Relations

1) $Z_i - (X_i, Y_i) - (U_{1i}, V_i, \hat{X}_{1i})$: To establish this relation, we show that $I(Z_i; \hat{X}_{1i}, U_{1i}, V_i|X_i, Y_i) = 0$

$$
\begin{aligned}
&I(Z_i; \hat{X}_{1i}, U_{1i}, V_i|X_i, Y_i) \\
&= I(Z_i; \hat{X}_{1i}, M_3, M_2, X^{i-1}, Y^{i-1}, Z_{i+1}^n|X_i, Y_i) \\
&\leq I(Z_i; \hat{X}_{1i}, M_3, M_2, X^{i-1}, Y^{i-1}, X_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n|X_i, Y_i) \\
&= I(Z_i; X^{i-1}, Y^{i-1}, X_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n|X_i, Y_i) \\
&= 0.
\end{aligned}
$$

2) $U_{2i} - (Z_i, U_{1i}, V_i) - (\hat{X}_{1i}, X_i, Y_i)$: Consider

$$
\begin{aligned}
&I(\hat{X}_i, X_i, Y_i; U_{2i}|Z_i, U_{1i}, V_i) \\
&\leq I(\hat{X}_i, X_i^n, Y_i^n; M_4|Z_i^n, X^{i-1}, Y^{i-1}, M_2, M_3) \\
&= I(X_i^n, Y_i^n; M_4|Z_i^n, X^{i-1}, Y^{i-1}, M_2, M_3).
\end{aligned}
$$

Now, using Lemma 1, set $A_1 = (X^{i-1}, Y^{i-1})$, $B_1 = Z^{i-1}$, $A_2 = (X_i^n, Y_i^n)$, $B_2 = (Z_i^n)$, $\tilde{M}_2 = M_4$, and $\tilde{M}_1 = M_2$. Then, using the third expression in the lemma, we see that $I(X_i^n, Y_i^n; M_4|Z_i^n, X^{i-1}, Y^{i-1}, M_2) = 0$.

3) $Z^{i-1} - (U_{1i}, V_i, Z_i) - (X_i, Y_i)$: Consider

$$
\begin{aligned}
&I(X_i, Y_i; Z^{i-1}|U_{1i}, V_i, Z_i) \\
&\leq I(X_i^n, Y_i^n; Z^{i-1}|X^{i-1}, Y^{i-1}, Z_i^n, M_2, M_3) \\
&= \big(H(Z^{i-1}|X^{i-1}, Y^{i-1}, Z_i^n, M_2, M_3) \\
&\qquad -H(Z^{i-1}|X^n, Y^n, Z_i^n, M_2, M_3)\big) \\
&\leq H(Z^{i-1}|X^{i-1}, Y^{i-1}, Z_i^n) - H(Z^{i-1}|X^n, Y^n, Z_i^n) \\
&= H(Z^{i-1}|X^{i-1}, Y^{i-1}) - H(Z^{i-1}|X^{i-1}, Y^{i-1}) \\
&= 0.
\end{aligned}
$$

4) $(X_{i+1}^n, Y_{i+1}^n) - (U_{1i}, U_{2i}, V_i, X_i, Y_i) - Z_i$: Consider

$$
\begin{aligned}
&I(X_{i+1}^n, Y_{i+1}^n; Z_i|U_{1i}, U_{2i}, V_i, X_i, Y_i) \\
&\leq I(X_{i+1}^n, Y_{i+1}^n; Z^i|M_2, M_3, M_4, Z_{i+1}^n, X^i, Y^i).
\end{aligned}
$$

Applying the first expression in the lemma with $A_2 = (X_{i+1}^n, Y_{i+1}^n)$, $A_1 = (X^i, Y^i)$, $B_1 = Z^i$, and $B_2 = Z_{i+1}^n$ gives $I(X_{i+1}^n, Y_{i+1}^n; Z_i|U_{1i}, U_{2i}, X_i, Y_i) = 0$.

Distortion Constraints: The proof of the distortion constraints is omitted since it follows similar steps to the two-way cascade source coding case, with the new Markov relations 3 and 4, and (U_{1i}, V_i) replacing U_{1i} in the proof.

Using the standard time sharing random variable Q as before and defining $U_1 = (U_{1Q}, Q)$, $U_2 = U_{2Q}$, $\hat{X}_1 = \hat{X}_{1Q}$, and $V = V_Q$, we obtain an outer bound for the rate-distortion region for some probability distribution of the form $p(x, y, z, u_1, u_2, v, \hat{x}_1) = p(x, y, z) p(u_1|x, y)p(\hat{x}_1, v|x, y, u_1)p(u_2|z, u_1, v)$. It remains to show that it suffices to consider probability distributions of the form $p(x, y, z)p(u_1|x, y)p(\hat{x}_1|x, y, u_1)$. $p(v|x, y, u_1)p(u_2|z, u_1, v)$. This follows similar steps to proof of Theorem 2. Let

$$
\begin{aligned}
p_1 &= p(x, y, z)p(u_1|x, y)p(\hat{x}_1, v|x, y, u_1)p(u_2|z, u_1, v), \\
p_2 &= (p(x, y, z)p(u_1|x, y) \\
&\qquad \times \hat{p}(\hat{x}_1|x, y, u_1)\hat{p}(v|x, y, u_1)p(u_2|z, u_1, v))
\end{aligned}
$$

where $\hat{p}(\hat{x}_1|x, y, u_1)$ and $\hat{p}(v|x, y, u_1)$ are the marginals induced by p_1. Next, note that R_1, R_2, R_3, R_4, and the distortion constraints depend on p_1 only through the marginals $p(x, y, z, u_1, u_2, v)$ and $p(x, y, z, u_1, \hat{x}_1)$. Since these marginals are the same for p_1 and p_2, the rate and distortion constraints are unchanged. Finally, note that the Markov relations 1 and 2 implied by p_1 continue to hold under p_2. This completes the proof of the converse. □

IV. Quadratic Gaussian Distortion Case

In this section, we evaluate the rate-distortion regions when (X, Y, Z) are jointly Gaussian and the distortion is measured in terms of the mean square error. We will assume, without loss of generality, that $X = A + B + Z$, $Y = B + Z$, and $Z = Z$, where A, B, and Z are independent, zero mean Gaussian random variables with variances σ_A^2, σ_B^2, and σ_Z^2, respectively. While the results in Section III were proven only for discrete memoryless sources, the extension to the quadratic Gaussian case is standard and can be found in, for example, [10] and [7, Lecture 3].

A. Quadratic Gaussian Cascade Source Coding

Corollary 1 (Quadratic Gaussian Cascade Source Coding): First, we note that if $R_2 < \frac{1}{2}\log\frac{\sigma_A^2+\sigma_B^2}{D_2}$, then the distortion constraint D_2 cannot be met. Hence, given $D_1, D_2 > 0$ and $R_2 \geq \max\{\frac{1}{2}\log\frac{\sigma_A^2+\sigma_B^2}{D_2}, 0\}$, the rate-distortion region for quadratic Gaussian cascade source coding is characterized by the smallest rate R_1 such that (D_1, D_2, R_1, R_2) are achievable, which is

$$R_1 = \max\left\{\frac{1}{2}\log\frac{\sigma_A^2}{D_1}, \frac{1}{2}\log\frac{\sigma_A^2}{\sigma_{A|U,B}^2}\right\}$$

where $U = \alpha^* A + \beta^* B + Z^*$, $Z^* \sim N(0, \sigma_{Z^*}^2)$, with α^*, β^* and $\sigma_{Z^*}^2$ achieving the maximum in the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \sigma_{A|U,B}^2 \\ \text{subject to} \quad & R_2 \geq \frac{1}{2}\log\frac{\sigma_U^2}{\sigma_{Z^*}^2} \\ & D_2 \geq \sigma_{A+B|U}^2. \end{aligned}$$

The optimization problem given in the corollary can be solved following analysis in [4]. In our proof of the corollary, we will show that the rate-distortion region obtained is the same as the case when the degraded side information Z is available to all nodes.

Converse: Consider the case when the side information Z is available to all nodes. Without loss of generality, we can subtract the side information away from X and Y to obtain a rate-distortion problem involving only $A+B$ and B at node 0, B at node 1 and no side information at node 2. Characterization of this class of quadratic Gaussian cascade source coding problem has been carried out in [4] and following the analysis therein, we can show that the rate-distortion region is given by the region in Corollary 1. □

Achievability: We evaluate Theorem 1 using Gaussian auxiliaries random variables. Let $U' = \alpha^* X + (\beta^* - \alpha^*)Y + Z^* = \alpha^* A + \beta^*(B+Z) + Z^*$ and V be a Gaussian random variable that we will specify in the proof. We now rewrite $R_1 = I(X; U', \hat{X}_1|Y)$ as $R_1 = I(X; U', V|Y)$ with $\hat{X}_1 = V + \mathsf{E}(X|U', Y)$, V independent of U' and Y. Let $g_2(U', Z) = \mathsf{E}(X|U', Z)$. Evaluating R_1 and R_2 using this choice of auxiliaries, we have

$$\begin{aligned} R_1 &= I(X; U', V|Y) \\ &= h(A+B+Z|B+Z) - h(X|U', V, Y) \\ &= \frac{1}{2}\log\frac{\sigma_A^2}{\sigma_{X|U',V,Y}^2} \\ R_2 &= I(X, Y; U'|Z) \\ &= h(U'|Z) - h(U'|X, Y, Z) \\ &= \frac{1}{2}\log\frac{\sigma_{\alpha^* A+\beta^* B+Z^*}^2}{\sigma_{Z^*}^2} \\ &= \frac{1}{2}\log\frac{\sigma_U^2}{\sigma_{Z^*}^2}. \end{aligned}$$

Next, we have

$$\begin{aligned} \sigma_{X|U',Y}^2 &= \sigma_{A+B+Z|\alpha^* A+\beta^*(B+Z)+Z^*, B+Z}^2 \\ &= \sigma_{A|\alpha^* A+Z^*, B+Z}^2 \\ &= \sigma_{A|\alpha^* A+Z^*}^2 \\ &= \sigma_{A|U,B}^2. \end{aligned}$$

If $\sigma_{X|U',Y}^2 = \sigma_{A|U,B}^2 \leq D_1$, we set $V = 0$ to obtain $R_1 = \frac{1}{2}\log\frac{\sigma_A^2}{\sigma_{X|U',Y}^2}$. If $\sigma_{X|U',Y}^2 > D_1$, then we choose $V = X - \mathsf{E}(X|U', Y) + Z_2$ where $Z_2 \sim N(0, D_1\sigma_{X|U',Y}^2/(\sigma_{X|U',Y}^2 - D_1))$ so that $\sigma_{X|U',V,Y}^2 = D_1$ and obtain $R_1 = \frac{1}{2}\log\frac{\sigma_A^2}{D_1}$. Therefore, $R_1 = \max\{\frac{1}{2}\log\frac{\sigma_A^2}{D_1}, \frac{1}{2}\log\frac{\sigma_A^2}{\sigma_{A|U,B}^2}\}$.

Finally, we show that this choice of random variables satisfy the distortion constraints. For D_1, note that since $\mathsf{E}(X - \hat{X}_1)^2 = \sigma_{X|U',V,Y}^2$, the distortion constraint D_1 is always satisfied. For the second distortion constraint, we have

$$\begin{aligned} \mathsf{E}(X - \hat{X}_2)^2 &= \sigma_{X|U',Z}^2 \\ &= \sigma_{A+B|\alpha^* A+\beta^*(B+Z)+Z^*, Z}^2 \\ &= \sigma_{A+B|\alpha^* A+\beta^* B+Z^*, Z}^2 \\ &= \sigma_{A+B|\alpha^* A+\beta^* B+Z^*}^2 \\ &= \sigma_{A+B|U}^2 \\ &\leq D_2. \end{aligned}$$

Hence, our choice of auxiliary U' and V satisfies the rate-distortion region and distortion constraints given in the corollary, which completes our proof. □

B. Quadratic Gaussian Triangular Source Coding

Corollary 2 (Quadratic Gaussian Triangular Source Coding): Given $D_1, D_2 > 0$ and $R_2, R_3 \geq 0, R_2 + R_3 \geq \frac{1}{2}\log\frac{\sigma_A^2+\sigma_B^2}{D_2}$, the rate-distortion region for quadratic Gaussian triangular source coding is characterized by the smallest R_1 for which $(D_1, D_2, R_1, R_2, R_3)$ is achievable, which is

$$R_1 = \max\left\{\frac{1}{2}\log\frac{\sigma_A^2}{D_1}, \frac{1}{2}\log\frac{\sigma_A^2}{\sigma_{A|U,B}^2}\right\}$$

where $U = \alpha^* A + \beta^* B + Z^*$, $Z \sim N(0, \sigma_{Z^*}^2)$, with α^*, β^*, and $\sigma_{Z^*}^2$ satisfying the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \sigma_{A|U,B}^2 \\ \text{subject to} \quad & R_2 \geq \frac{1}{2}\log\frac{\sigma_U^2}{\sigma_{Z^*}^2} \\ & 2^{2R_3} D_2 \geq \sigma_{A+B|U}^2. \end{aligned}$$

As with Corollary 1, the optimization problem given this corollary can be solved following analysis in [4].

Converse: The converse uses the same approach as Corollary 1. Consider the case when the side information Z is available to all nodes. Without loss of generality, we can subtract the side information away from X and Y to obtain a rate-distortion problem involving only $A+B$ and B at node 0, B at node 1 and no side information at node 2. Characterization of this class of quadratic Gaussian triangular source coding problem has been carried out in [4] and following the analysis therein, we can show that the rate-distortion region is given by the region in Corollary 2. □

Achievability: We evaluate Theorem 2 using Gaussian auxiliary random variables. Let $U' = \alpha^* X + (\beta^* - \alpha^*)Y + Z^* = \alpha^* A + \beta^*(B+Z) + Z^*$ and $V' = X + \eta U' + Z_3$, $Z_3 \sim$

$N(0,\sigma^2_{Z_3})$. Following the analysis in Corollary 1, the inequalities for the rates are

$$R_1 = \max\left\{\frac{1}{2}\log\frac{\sigma^2_A}{D_1}, \frac{1}{2}\log\frac{\sigma^2_A}{\sigma^2_{A|U,B}}\right\}$$
$$R_2 \geq \frac{1}{2}\log\frac{\sigma^2_U}{\sigma^2_{Z^*}}$$
$$R_3 \geq I(X,Y;V|Z,U') = I(X;V'|Z,U') = \frac{1}{2}\log\frac{\sigma^2_{X|Z,U'}}{\sigma^2_{X|Z,U',V'}}.$$

As with Corollary 1, the distortion constraint D_1 is satisfied with an appropriate choice of $\hat{X}_1$. For the distortion constraint D_2, we have

$$D_2 \geq \sigma^2_{X|Z,U',V'}.$$

Next, note that we can assume equality for R_3, since we can adjust η and $\sigma^2_{Z_3}$ so that inequality is met. Since this operation can will only decrease $\sigma^2_{X|Z,U',V'}$, the distortion constraint D_2 will still be met. Therefore, setting $R_3 = \frac{1}{2}\log\frac{\sigma^2_{X|Z,U'}}{\sigma^2_{X|Z,U',V'}}$, we have

$$D_2 \geq \sigma^2_{X|Z,U',V'} = \frac{\sigma^2_{X|Z,U'}}{2^{2R_3}}.$$

Since $\sigma^2_{X|Z,U'} = \sigma^2_{A+B|U}$, this completes the proof of achievability. □

Remark: As alternative characterizations, we show in Appendix C that the cascade and triangular settings in Corollaries 1 and 2 can be transformed into equivalent problems in [4] where explicit characterizations of the rate-distortion regions were given.

C. Quadratic Gaussian Two-Way Source Coding

It is straightforward to extend Corollaries 1 and 2 to quadratic Gaussian two-way cascade and triangular source coding using the observation that in the quadratic Gaussian case, side information at the encoder does not reduce the required rate. Therefore, the backward rate from node 2 to node 0 is always lower bounded by $\frac{1}{2}\log\frac{\sigma^2_{Z|B+Z}}{D_3}$. This rate (and distortion constraint D_3) can be achieved by simply encoding Z. We therefore state the following corollary without proof.

Corollary 3 (Quadratic Gaussian Two-Way Triangular Source Coding): Given $D_1, D_2, D_3 > 0$, $R_2, R_3 \geq 0$, $R_2 + R_3 \geq \frac{1}{2}\log\frac{\sigma^2_A+\sigma^2_B}{D_2}$, and $R_4 \geq \max\{\frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_3}, 0\}$, the rate-distortion region for quadratic Gaussian two-way triangular source coding is characterized by the smallest R_1 for which $(R_1, R_2, R_3, R_4, D_1, D_2, D_3)$ is achievable, which is

$$R_1 = \max\left\{\frac{1}{2}\log\frac{\sigma^2_A}{D_1}, \frac{1}{2}\log\frac{\sigma^2_A}{\sigma^2_{A|U,B}}\right\}$$

where $U = \alpha^* A + \beta^* B + Z^*$, $Z \sim N(0, \sigma^2_{Z^*})$, with α^*, β^* and $\sigma^2_{Z^*}$ satisfying the following optimization problem:

$$\text{maximize} \quad \sigma^2_{A|U,B}$$
$$\text{subject to} \quad R_2 \geq \frac{1}{2}\log\frac{\sigma^2_U}{\sigma^2_{Z^*}}$$
$$2^{2R_3}D_2 \geq \sigma^2_{A+B|U}.$$

Remark: The special case of two-way cascade quadratic Gaussian source coding can be obtained as a special case by setting $R_3 = 0$.

Next, we present an extension to our settings for which we can characterize the rate-distortion region in the quadratic Gaussian case. In this extended setting, we have cascade setting from node 0 to node 2 and a triangular setting from node 2 to node 0, with the additional constraint that node 1 also reconstructs a lossy version of Z. As formal definitions are natural extensions of those presented in Section II, we will omit them here. The setting is shown in Fig. 5.

Theorem 5 (Extended Quadratic Gaussian Two-Way Cascade Source Coding): Given $D_1, D_2 > 0$, $0 < D_{Z_1}, D_{Z_2} \leq \sigma^2_{Z|Y}$ and $R_2 \geq \max\{\frac{1}{2}\log\frac{\sigma^2_A+\sigma^2_B}{D_2}, 0\}$, the rate-distortion region for the extended quadratic Gaussian two-way cascade source coding is given by the set of $R_1, R_3, R_4, R_5 \geq 0$ satisfying the following equalities and inequalities:

$$R_1 = \max\left\{\frac{1}{2}\log\frac{\sigma^2_A}{D_1}, \frac{1}{2}\log\frac{\sigma^2_A}{\sigma^2_{A|U,B}}\right\}$$

where $U = \alpha^* A + \beta^* B + Z^*$, $Z^* \sim N(0, \sigma^2_{Z^*})$, with α^*, β^* and $\sigma^2_{Z^*}$ satisfying the following optimization problem:

$$\text{maximize} \quad \sigma^2_{A|U,B}$$
$$\text{subject to} \quad R_2 \geq \frac{1}{2}\log\frac{\sigma^2_U}{\sigma^2_{Z^*}}$$
$$D_2 \geq \sigma^2_{A+B|U}$$

and

$$R_3 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_1}}$$
$$R_3 + R_5 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{\min\{D_{Z_1}, D_{Z_2}\}}$$
$$R_4 + R_5 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}.$$

Proof: Converse: For the forward direction (R_1, R_2), we note that node 2 can only send a function of (M_1, Y^n, Z^n) to nodes 0 and 1 using the R_4 and R_5 links. Since M_1 and Y^n are available at both nodes 0 and 1, the forward rates are lower bounded by the setting where Z^n is available to all nodes. Further, in this setting, the distortion constraints D_{Z_1} and D_{Z_2} are automatically satisfied since Z is available at nodes 0 and 1. Therefore, (R_3, R_4, R_5) do not affect the achievable (R_1, R_2) rates in this modified (lower bound) setting. (R_1, R_2) are then obtained by the observation in Corollary 1 that the rate-distortion region obtained for our quadratic Gaussian cascade setting in Corollary 1 is equivalent to the case where the side information Z is available at all nodes.

For the reverse direction, the lower bounds are derived by letting the side information (X, Y) to be available at node 2, and for side information X to be available at node 1. The D_1 and D_2

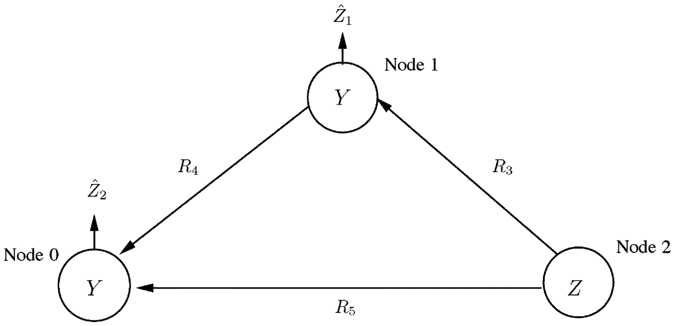

Fig. 6. Setup for analysis of achievability of backward rates.

distortion constraints are then automatically satisfied since X is available at all nodes. We then observed that (R_1, R_2) do not affect the achievable (R_3, R_4, R_5) rates in this modified (lower bound) setting. The stated inequalities for R_3, R_4, R_5 are then obtained from standard cutset bound arguments and the fact that $X - Y - Z$ form a Markov chain.

Achievability: We analyze only the backward rates R_3, R_4, and R_5 since the forward direction follows from Corollary 1. For the backward rates, we now show that the rates are achievable without the assumption of (X, Y) being available at node 2. We will rely on results on successive refinement of Gaussian sources with common side information given in [11]. A simplified figure of the setup for analyzing the backward rates is given in Fig. 6. We have three cases to consider.

Case 1: $D_{Z_1} \leq D_{Z_2}$. In this case, the inequalities in the lower bound reduce to

$$R_3 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_1}}$$
$$R_4 + R_5 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}.$$

From the successive refinement results in [11], we can show that the following rates are achievable:

$$\begin{aligned} R_3 &= I(U_1, U_2, U_3; Z|Y) \\ R_4 &= I(U_2; Z|Y) \\ R_5 &= I(U_3; Z|Y, U_2) \end{aligned}$$

for some conditional distribution $F(U_1, U_2, U_3|Z)$, $\hat{Z}_1(U_1, U_2, U_3, Y)$, and $\hat{Z}_2(U_1, U_2, Y)$ satisfying the distortion constraints. Now, for fixed $R_4 \leq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$, choose $D'(\geq D_{Z_2})$ such that $R_4 = \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D'}$. We now choose the auxiliary random variables and reconstruction functions in the following manner. Define $Q(x) := \frac{x\sigma^2_{Z|Y}}{\sigma^2_{Z|Y} - x}$

$$\begin{aligned} U_1 &= Z + W_1, \quad \text{where } W_1 \sim N(0, Q(D_{Z_1})) \\ U_3 &= U_1 + W_3, \quad \text{where } W_3 \sim N(0, Q(D_{Z_2}) - Q(D_{Z_1})) \\ U_2 &= U_3 + W_2, \quad \text{where } W_2 \sim N(0, Q(D') - Q(D_{Z_2})) \\ \hat{Z}_1 &= \mathsf{E}(Z|U_1, Y) \\ \hat{Z}_2 &= \mathsf{E}(Z|U_3, Y). \end{aligned}$$

From this choice of auxiliary random variables, it is easy to verify the following:

$$\begin{aligned} R_3 &= I(U_1, U_2, U_3; Z|Y) \\ &= I(U_1; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_1}} \\ R_4 &= I(U_2; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D'} \\ R_4 + R_5 &= I(U_2; Z|Y) + I(U_3; Z|Y, U_2) \\ &= I(U_3, U_2; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}} \\ \mathsf{E}(Z - \hat{Z}_1)^2 &= D_{Z_1} \\ \mathsf{E}(Z - \hat{Z}_2)^2 &= D_{Z_2}. \end{aligned}$$

Case 2: $D_{Z_1} > D_{Z_2}$, $R_3 \geq R_4$. In this case, the active inequalities are

$$R_3 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_1}}$$
$$R_4 + R_5 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}.$$

From [11], the following rates are achievable:

$$\begin{aligned} R_3 &= I(U_1, U_2; Z|Y) \\ R_4 &= I(U_2; Z|Y) \\ R_5 &= I(U_3, U_1; Z|Y, U_2). \end{aligned}$$

First, assume $R_3 \leq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$. Choose $D_{Z_2} \leq D' \leq D'' \leq D_{Z_1}$. We choose the auxiliary random variables and reconstruction functions as follows:

$$\begin{aligned} U_3 &= Z + W_3, \quad \text{where } W_3 \sim N(0, Q(D_{Z_2})) \\ U_1 &= U_3 + W_1, \quad \text{where } W_1 \sim N(0, Q(D') - Q(D_{Z_2})) \\ U_2 &= U_1 + W_2, \quad \text{where } W_2 \sim N(0, Q(D'') - Q(D')) \\ \hat{Z}_1 &= \mathsf{E}(Z|U_1, Y) \\ \hat{Z}_2 &= \mathsf{E}(Z|U_3, Y). \end{aligned}$$

From this choice of auxiliary random variables, it is easy to verify the following:

$$\begin{aligned} R_3 &= I(U_1, U_2; Z|Y) \\ &= I(U_1; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D'} \\ R_4 &= I(U_2; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D''} \\ R_4 + R_5 &= I(U_2; Z|Y) + I(U_3, U_1; Z|Y, U_2) \\ &= I(U_3, U_1, U_2; Z|Y) \\ &= I(U_3; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}} \\ \mathsf{E}(Z - \hat{Z}_1)^2 &= D' \leq D_{Z_1} \\ \mathsf{E}(Z - \hat{Z}_2)^2 &= D_{Z_2}. \end{aligned}$$

Next, consider $R_3 > \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$ and $R_4 > \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$. Then, it is easy to see from our achievability scheme that we can obtain $R_4' < R_4$, $R_3' < R_3$, and $R_5 = 0$ by setting $D' = D'' = D_{Z_2}$. Finally, consider the case where $R_3 > \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$ and $R_4 \leq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$. Then, we observe from our achievability scheme that we can achieve $R_3' = \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}} < R_3$ for any R_4 and R_5 satisfying the inequalities by setting $D' = D_{Z_2}$.

Case 3: $D_{Z_1} > D_{Z_2}$, $R_3 < R_4$. In this case, the active inequalities are

$$R_3 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_1}}$$
$$R_3 + R_5 \geq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}.$$

We first consider the case where $R_3 \leq \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$. We exhibit a scheme for which $R_4' = R_3(<R_4)$ and still satisfies the constraints. This procedure is done by letting U_2 in case 2 to be equal to U_1. For $D_{Z_2} \leq D' \leq D_{Z_1}$, define the auxiliary random variables and reconstruction functions as follows:

$$U_3 = Z + W_3, \quad \text{where } W_3 \sim N(0, Q(D_{Z_2}))$$
$$U_1 = U_3 + W_1, \quad \text{where } W_1 \sim N(0, Q(D') - Q(D_{Z_2}))$$
$$\hat{Z}_1 = \mathsf{E}(Z|U_1, Y)$$
$$\hat{Z}_2 = \mathsf{E}(Z|U_3, Y).$$

Then, we have the following:

$$\begin{aligned} R_3 &= I(U_1; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D'} \\ R_4' &= I(U_1; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D'} \\ R_3 + R_5 &= I(U_1; Z|Y) + I(U_3; Z|Y, U_1) \\ &= I(U_3, U_1; Z|Y) \\ &= I(U_3; Z|Y) \\ &= \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}} \\ \mathsf{E}(Z - \hat{Z}_1)^2 &= D' \leq D_{Z_1} \\ \mathsf{E}(Z - \hat{Z}_2)^2 &= D_{Z_2}. \end{aligned}$$

Finally, we note that in the case where $R_3 > \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$, we can always achieve $R_3' = \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$, $R_4' = \frac{1}{2}\log\frac{\sigma^2_{Z|Y}}{D_{Z_2}}$, and $R_5' = 0$ by letting $D' = D_{Z_2}$. □

Remark 1: The two-way cascade source coding setup given in Section II can be obtained as a special case by setting $R_3 = R_4 = 0$ and $D_{Z_1} \to \infty$.

Remark 2: The rate-distortion region is the same regardless of whether node 2 sends first, or node 0 sends first. This observation follows from i) our result in Corollary 1 where we showed that the rate-distortion region for the cascade setup is equivalent to the setup where all nodes have the degraded side information Z; and ii) our proof above where we showed that the backward rates are the same as in the case where the side information (X, Y) is available at all nodes.

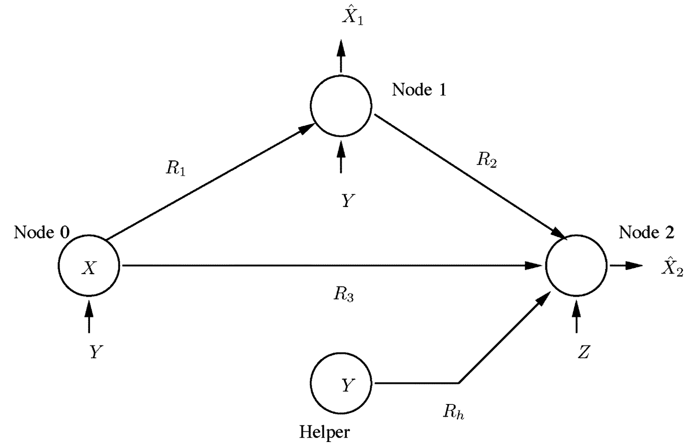

Fig. 7. Triangular source coding with a helper.

Remark 3: For arbitrary sources and distortions, the problem is open in general. Even in the Gaussian case, the problem is open without the Markov chain $X - Y - Z$. One may also consider the setting where there is a triangular source coding setup in the forward path from node 0 to node 2. This setting is still open, since the tradeoff in sending from node 0 to node 2 and then to node 1 versus sending directly to node 1 from node 0 is not clear.

V. Triangular Source Coding With a Helper

We present an extension to our triangular source coding setup by also allowing the side information Y to be observed at the second node through a rate limited link (or helper). The setup is shown in Fig. 7. As the formal definitions are natural extensions of those given in Section II, we will omit them here.

Theorem 6: The rate-distortion region for triangular source coding with a helper is given by the set of rate tuples

$$\begin{aligned} R_1 &\geq I(X; \hat{X}_1, U_1|Y, U_h) \\ R_2 &\geq I(U_1; X, Y|Z, U_h) \\ R_3 &\geq I(X, Y; U_2|U_1, U_h, Z) \\ R_h &\geq I(U_h; Y|Z) \end{aligned}$$

for some $p(x, y, z, u_1, u_2, u_h, \hat{x}_1) = p(x)p(y|x)p(z|y)$ $p(u_h|y)p(u|x, y, u_h)$. $p(\hat{x}_1|x, y, u_1, u_h)p(u_2|x, y, u_1, u_h)$ and function $g_2 : \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{U}_h \times \mathcal{Z} \to \hat{X}_2$ such that

$$\mathsf{E}d_j(X_j, \hat{X}_j) \leq D_j, \qquad j = 1, 2.$$

We give a proof of the converse in Appendix D. As the achievability techniques used form a straightforward extension of the techniques described in Appendix A, we give only a sketch of achievability.

Sketch of Achievability: The achievability follows that of triangular source coding, with an additional step of generating a lossy description of Y^n. The codebook generation consists of the following steps.

- Generate $2^{n(I(Y;U_h)+\epsilon)}$ U_h^n sequences according to $\prod_{i=1}^n p(u_{hi})$. Partition the set of U_h^n sequences into $2^{n(I(U_h;Y|Z)+2\epsilon)}$ bins, $\mathcal{B}_h(m_h)$, $m_h \in [1 : 2^{n(I(U_h;Y|Z)+2\epsilon)}]$.

- Generate $2^{n(I(X,Y,U_h;U_1)+\epsilon)}$ U_1^n sequences according to $\prod_{i=1}^n p(u_{1i})$. Partition the set of U_1^n sequences into $2^{n(I(U_1;X|Y,U_h)+2\epsilon)}$ bins, $\mathcal{B}_1(m_{10})$. Separately and independently, partition the set of U^n sequences into $2^{n(I(U_1;X,Y|Z,U_h)+2\epsilon)}$ bins, $\mathcal{B}_2(m_2)$, $m_2 \in [1 : 2^{n(I(U;X,Y|Z)+2\epsilon)}]$.
- For each (u_1^n, u_h^n, y^n) sequence, generate $2^{n(I(\hat{X}_1^n;X|U_1,Y,U_h)+\epsilon)}$ $\hat{X}_1^n$ sequences according to $\prod_{i=1}^n p(\hat{x}_i|u_{1i}, u_{hi}, y_i)$.
- Generate $2^{n(I(U_2;X,Y|U_h,U_1)+\epsilon)}$ U_2^n sequences according to $\prod_{i=1}^n p(u_{2i}|u_{1i}, u_{hi})$ for each (u_1^n, u_h^n) sequence, and partition these sequences to $2^{n(I(U_2;X,Y|U_1,U_h,Z)+2\epsilon)}$ bins, $\mathcal{B}_3(m_3)$.

Encoding consists of the following steps.

- Helper node: The helper node (and nodes 0 and 1) looks for a u_h^n sequence such that $(u_h^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$. This step succeeds with high probability since there are $2^{n(I(Y;U_h)+\epsilon)}$ U_h^n sequences. The helper then sends out the bin index m_h such that $u_h^n \in \mathcal{B}(m_h)$. The sequences (u_h^n, x^n, y^n, z^n) are jointly typical with high probability due to the Markov chain $(X, Z) - Y - U_h$.
- Node 0: Given $(x^n, y^n, u_h^n) \in \mathcal{T}_\epsilon^{(n)}$, node 0 looks for a jointly typical codeword u_1^n. This operation succeeds with high probability since there are $2^{n(I(X,Y,U_h;U_1)+\epsilon)}$ U_1^n sequences. Node 0 then looks for a $\hat{x}_1^n$ that is jointly typical with (u_1^n, x^n, y^n, u_h^n). This operation succeeds with high probability since there are $2^{n(I(\hat{X}_1;X|U_1,U_h,Y)+\epsilon)}$ $\hat{x}_1^n$ sequences.
- Node 0 also finds a u_2^n sequence that is jointly typical with (u_1^n, u_h^n, x^n, y^n). This operation succeeds with high probability since we have $2^{n(I(U_2;X,Y|U_1,U_h)+\epsilon)}$ v^n sequences.
- Node 0 then sends out the bin index m_{10} such that $u_1^n \in \mathcal{B}_1(m_{10})$ and the index corresponding to $\hat{x}_1^n$ to node 1. This requires a total rate of $R_1 = I(U;X|Y) + I(\hat{X}_1^n;X|U,Y) + 3\epsilon$ to node 1. Node 0 also sends out the bin index m_3 such that $u_2^n \in \mathcal{B}(m_3)$ to node 2. This requires a rate of $I(U_2;X,Y|U_1,U_h,Z)+2\epsilon$.
- Node 1 decodes the codeword u_1^n and forwards the index m_2 such that $u_1^n \in \mathcal{B}(m_2)$ to node 2. This requires a rate of $I(U_1;X,Y|Z,U_h) + 2\epsilon$.

Decoding consists of the following steps.

- Node 1: Node 1 reconstructs u_1^n by looking for the unique U_1^n sequence in $\mathcal{B}_1(m_{10})$ such that $(U_1^n, U_h^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}$. Since there are only $2^{n(I(X,Y,U_h;U_1)-I(U_1;X|Y,U_h)-\epsilon)} = 2^{n(I(U_1;U_h,Y)-\epsilon)}$ sequences in the bin, this operation succeeds with high probability. Node 1 reconstructs X^n as $\hat{X}_1^n(m_{10}, m_{11})$. Since the sequence $(\hat{X}_1^n, X^n)$ are jointly typical with high probability, the expected distortion constraint is satisfied.
- Node 2: We note that since $(U_1, U_2, U_h, X) - Y - Z$, the sequences $(U_h^n, U_1^n, U_2^n, X^n, Y^n, Z^n)$ are jointly typical with high probability. Decoding at node 2 consists of the following steps.
  1) Node 2 first looks for u_h^n in $\mathcal{B}_h(m_h)$ such that $(u_h^n, z^n) \in \mathcal{T}_\epsilon^{(n)}$. This operation succeeds with high probability since there are only $2^{n(I(U_h;Z)-\epsilon}$ u_h^n sequences in the bin.
  2) It then looks for u_1^n in $\mathcal{B}_2(m_2)$ such that $(u_h^n, u_1^n, z^n) \in \mathcal{T}_\epsilon^{(n)}$. Since $I(U_1;X,Y,U_h) - I(U_1;X,Y|Z,U_h) = I(U_1;Z,U_h)$ by the Markov chain $Z - (X,Y,U_h) - U_1$, this operation succeeds with high probability as there are only $2^{n(I(U_1;Z,U_h)-\epsilon}$ u_1^n sequences in the bin.
  3) Finally, it looks for u_2^n in $\mathcal{B}_3(m_3)$ such that $(u_h^n, u_1^n, u_2^n, z^n) \in \mathcal{T}_\epsilon^{(n)}$. Since $I(U_2;X,Y|U_h,U_1) - I(U_2;X,Y|Z,U_h,U_1) = I(U_2;Z|U_1,U_h)$ by the Markov chain $Z - (X,Y,U_h,U_1) - U_2$, this operation succeeds with high probability as there are only $2^{n(I(U_2;Z|U_1,U_h)-\epsilon}$ u_2^n sequences in the bin.
  4) Node 2 then reconstructs using the function $\hat{x}_{2i} = g_2(u_{1i}, u_{2i}, u_{hi}, z_i)$ for $i \in [1 : n]$. Since the sequences $(X^n, Z^n, U_1^n, U_2^n, U_h^n)$ are jointly typical with high probability, the expected distortion constraint is satisfied. □

VI. Conclusion

Rate-distortion regions for the cascade, triangular, two-way cascade, and two-way triangular source coding settings were established. Decoding part of the description intended for node 2 and then re-binning it was shown to be optimum for our cascade and triangular settings. We also extended our triangular setting to the case where there is an additional rate constrained helper, which observes Y, for node 2. In the quadratic Gaussian case, we showed that the auxiliary random variables can be taken to be jointly Gaussian and that the rate-distortion regions obtained for the cascade and triangular setup were equivalent to the setting where the degraded side information is available at all nodes. This observation allows us to transform our cascade and triangular settings into equivalent settings for which explicit characterizations are known. Characterizations of the rate-distortion regions for the quadratic Gaussian cases were also established in the form of tractable low-dimensional optimization programs. Our two-way cascade quadratic Gaussian setting was extended to solve a more general two-way cascade scenario. The case of generally distributed X, Y, Z, without the degradedness assumption, remains open.

Appendix A
Achievability Proofs

A. Achievability proof of Theorem 1

1) Codebook Generation:

- Fix the joint distribution $p(x,y,z,u,\hat{x}_1) = p(x)p(y|x)$ $p(z|y)p(u|x,y)p(\hat{x}_1|x,y,u)$. Let $R = R_{10} + R_{11}$, $R_l \geq R_{10}$ and $R_2 \geq R_{10}$.
- Generate $2^{nR_{10}}$ $U^n(l)$ sequences, $l \in [1 : 2^{nR_1}]$, each according to $\prod_{i=1}^n p(u_i)$.
- Partition the set of U^n sequences into $2^{nR_{10}}$ bins, $\mathcal{B}_1(m_{10})$, $m_{10} \in [1 : 2^{nR_{10}}]$. Separately and independently, partition the set of U^n sequences into 2^{nR_2} bins, $\mathcal{B}_2(m_2)$, $m_2 \in [1 : 2^{nR_2}]$.
- For each $u^n(l)$ and y^n sequences, generate $2^{nR_{11}}$ $\hat{X}_1^n(l, m_{11})$ sequences according to $\prod_{i=1}^n p(\hat{x}_{1i}|u_i, y_i)$.

2) Encoding at the Encoder: Given a (x^n, y^n) pair, the encoder first looks for an index $l \in [1 : 2^{nR_l}]$ such that $(u^n(l), x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$, where $\mathcal{T}_\epsilon^{(n)}$ stands for the set of jointly typical sequences. If there are more than one such l, it selects one uniformly at random from the set of admissible indices. If there is none, it sends an index uniformly at random from $[1 : 2^{nR_l}]$.[2] Next, it finds the index m_{11} such that $(\hat{x}_1(l, m_{11}), u^n(m_{10}), x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$. As before, if there are more than one, it selects one uniformly at random from the set of admissible indices. If there is none, it sends an index uniformly at random from $[1 : 2^{nR_{11}}]$. Finally, it sends out (m_{10}, m_{11}), where m_{10} is the bin index such that $u^n(l) \in \mathcal{B}_1(m_{10})$. The total rate required is R.

3) Decoding and Reconstruction at Node 1: Given (m_{10}, m_{11}), node 1 looks for the unique $\hat{l}$ such that $(u^n(\hat{l}), y^n) \in \mathcal{T}_\epsilon^{(n)}$ and $u^n(\hat{l}) \in \mathcal{B}_1(l)$. It reconstructs x^n as $\hat{x}^n(\hat{l}, m_{11})$. If it failed to find a unique one, or if there are more than one, it outputs $\hat{l} = 1$ and performs the reconstruction as before.

4) Encoding at Node 1: Node 1 sends an index $\hat{m}_2$ such that $u^n(\hat{l}) \in \mathcal{B}_2(\hat{m}_2)$. This requires a rate of R_2.

5) Decoding and Reconstruction at Node 2: Node 2 looks for the index $\tilde{l}$ such that $(u^n(\tilde{l}), y^n) \in \mathcal{T}_\epsilon^{(n)}$ and $\tilde{l} \in \mathcal{B}_2(\hat{m}_2)$. It then reconstructs x^n according to $\hat{x}_{2i} = g_2(u^n(\tilde{l})_i, z_i)$ for $i \in [1 : n]$. If there is no such index, it reconstructs using $\tilde{l} = 1$.

6) Analysis of Expected Distortion: Using the typical average lemma in [7, Lecture 2] and following the analysis in [7, Lecture 3], it suffices to analyze the probability of "error"; i.e., the probability that the chosen sequences will not be jointly typical with the source sequences. Let L and M_{11} be the chosen indices at the encoder. Note that these define the bin indices M_{10} and M_2. Let $\hat{M}_2$ be the chosen index at node 1. Define the following error events.

1) $\mathcal{E}_0 := \{(X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$.
2) $\mathcal{E}_1 := \{(U^n(l), X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$ for all $l \in [1 : 2^{nR_l}]$.
3) $\mathcal{E}_2 := \{(U^n(l), X^n, Y^n, Z^n) \notin \mathcal{T}_\epsilon^{(n)}\}$ for all $l \in [1 : 2^{nR_l}]$.
4) $\mathcal{E}_3 := \{(U^n(L), \hat{X}^n(L, m_{11}), X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$ for all $m_{11} \in [1 : 2^{nR_{11}}]$.
5) $\mathcal{E}_4 := \{(U^n(\hat{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)}\}$ for some $\hat{l} \neq L$ and $U^n(\hat{l}) \in \mathcal{B}_1(M_{10})$.
6) $\mathcal{E}_5(\hat{M}_2) := \{(U^n(\tilde{l}), Z^n) \in \mathcal{T}_\epsilon^{(n)}\}$ for some $\tilde{l} \neq L$ and $U^n(\tilde{l}) \in \mathcal{B}_2(\hat{M}_2)$.

We can then bound the probability of error as

$$\mathsf{P}_e \leq \mathsf{P}\{\bigcup_{i=0}^{5} \mathcal{E}_i\} = \sum \mathsf{P}\{\mathcal{E}_i \cap (\bigcap_{j=0}^{i-1} \mathcal{E}_j^c)\}.$$

- $\mathsf{P}\{\mathcal{E}_0\} \to 0$ as $n \to \infty$ by law of large numbers (LLN).
- By the covering lemma in [7, Lecture 3], $\mathsf{P}\{\mathcal{E}_1 \cap \mathcal{E}_0^c\} \to 0$ as $n \to \infty$ if

$$R_l > I(U; X, Y) + \delta(\epsilon).$$

[2]For simplicity, we assume randomized encoding, but it is easy to see that the randomized encoding employed our proofs can be incorporated as part of the (random) codebook generation stage.

- $\mathsf{P}\{\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_0^c\} \to 0$ as $n \to \infty$ by the Markov relation $U - (X, Y) - Z$ and the conditional joint typicality lemma [7, Lecture 2].
- By the covering lemma in [7, Lecture 3], $\mathsf{P}\{\mathcal{E}_3 \cap (\bigcap_{j=0}^{2} \mathcal{E}_j^c\} \to 0$ as $n \to \infty$ if

$$R_{11} > I(\hat{X}_1; X | U, Y) + \delta(\epsilon).$$

- From the analysis of the Wyner–Ziv coding scheme (see [8] or [7, Lecture 12]), $\mathsf{P}\{\mathcal{E}_4 \cap (\bigcap_{j=0}^{3} \mathcal{E}_j^c\} \to 0$ as $n \to \infty$ if

$$R_l - R_{10} < I(U; Y) - \delta(\epsilon).$$

- For the last term, we have

$$\begin{aligned}
&\mathsf{P}\{\mathcal{E}_5(\hat{M}_2) \cap (\bigcap_{j=0}^{4} \mathcal{E}_j^c)\} \\
&= \mathsf{P}\{\mathcal{E}_5(\hat{M}_2) \cap (\bigcap_{j=0}^{4} \mathcal{E}_j^c) \cap \{\hat{M}_2 \neq M_2\}\} \\
&\qquad + \mathsf{P}\{\mathcal{E}_5(\hat{M}_2) \cap (\bigcap_{j=0}^{4} \mathcal{E}_j^c) \cap \{\hat{M}_2 = M_2\}\} \\
&\overset{(a)}{=} \mathsf{P}\{\mathcal{E}_5(\hat{M}_2) \cap (\bigcap_{j=0}^{4} \mathcal{E}_j^c) \cap \{\hat{M}_2 = M_2\}\} \\
&= \mathsf{P}\{\mathcal{E}_5(M_2) \cap (\bigcap_{j=0}^{4} \mathcal{E}_j^c) \cap \{\hat{M}_2 = M_2\}\} \\
&\leq \mathsf{P}\{\mathcal{E}_5(M_2) \cap \mathcal{E}_2^c\}.
\end{aligned}$$

Step (a) follows from the observation that $(\bigcap_{j=0}^{4} \mathcal{E}_j^c) \cap \{\hat{M}_2 \neq M_2\} = \emptyset$. The analysis of the probability of error therefore reduces to the analysis for the equivalent Wyner–Ziv setup with Z as the side information at node 2. Hence, $\mathsf{P}\{\mathcal{E}_5(\hat{M}_2) \cap (\bigcap_{j=0}^{4} \mathcal{E}_j^c)\} \to 0$ as $n \to \infty$ if

$$R_l - R_2 < I(U; Z) - \delta(\epsilon).$$

Eliminating R_l in the aforementioned inequalities then gives us the required rate region.

B. Achievability Proof of Theorem 2

As the achievability proof for the triangular source coding case follows that of the cascade source coding case closely, we will only include the additional steps required for generating R_3 and analysis of probability of error at node 2. The steps for generating R_1 and R_2, and for reconstruction at node 1 are the same as the cascade setup.

1) Codebook Generation:

- Fix $p(x, y, z, u, v, \hat{x}_1) = p(x)p(y|x)p(z|y)p(u|x, y) p(\hat{x}_1|x, y, u)p(v|x, y, u)$.
- For each $u^n(l)$, generate $V^n(l_3)$, $l_3 \in [1 : 2^{n\tilde{R}_3}]$, according to $\prod_{i=1}^{n} p(v_i|u_i)$. Partition the set of v^n sequences into 2^{nR_3} bins, $\mathcal{B}_3(m_3)$.

2) Encoding:

- Given a sequence (x^n, y^n) and $u^n(l)$ found through the steps in the cascade source coding setup, the encoder looks for an index l_3 such that $(u^n, v^n(l, l_3), x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$. If it finds more than one, it selects one uniformly at random from the set of admissible indices. If it finds none, it outputs an index uniformly at random from $[1:2^{n\tilde{R}_3}]$. The encoder then sends out m_3 such that $L_3 \in \mathcal{B}_3(m_3)$.

3) Decoding: The additional decoding step is in decoding L_3. Node 2 looks for the unique $\hat{l}_3$ such that $(u^n(\tilde{l}), v^n(\tilde{l}, \hat{l}_3), z^n) \in \mathcal{T}_\epsilon^{(n)}$ and $v^n(\hat{l}_3) \in \mathcal{B}_3(M_3)$. If there is none or more than one, it outputs $\hat{m}_3 = 1$.

4) Analysis of Distortion: Let L, M_{11} and M_3 be the indices chosen by the encoder. Note that these fix the indices M_{10} and M_2. We follow similar analysis as in the cascade case, with the same definitions for error events $\mathcal{E}_0$ to $\mathcal{E}_5$. We also require the following additional error events.

1) $\mathcal{E}_6 := \{(U^n(L), V^n(L, L_3), X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$.
2) $\mathcal{E}_7 := \{(U^n(L), V^n(L, L_3), X^n, Y^n, Z^n) \notin \mathcal{T}_\epsilon^{(n)}\}$.
3) $\mathcal{E}_8(\tilde{L}) := \{(U^n(\tilde{L}), V^n(\tilde{L}, \hat{l}_3), Z^n) \in \mathcal{T}_\epsilon^{(n)}\}$ for some $\hat{l}_3 \neq L_3$ and $\hat{l}_3 \in \mathcal{B}_3(M_3)$.

To bound the probability of error, we have the following additional terms.

- By the covering lemma, $\mathsf{P}(\mathcal{E}_6 \cap \mathcal{E}_2^c) \to 0$ as $n \to \infty$ if

$$\tilde{R}_3 > I(V; X, Y|U) + \delta(\epsilon).$$

- $\mathsf{P}(\mathcal{E}_7 \cap \mathcal{E}_6^c) \to 0$ as $n\infty$ from the Markov condition $(V, U) - (X, Y) - Z$ and the conditional joint typicality lemma.
- $\mathsf{P}\{\mathcal{E}_8(\tilde{L}) \cap \mathcal{E}_5^c(\hat{M}_2) \cap \mathcal{E}_7^c \cap (\bigcap_{j=0}^4 \mathcal{E}_j^c))\}$. We have

$$\begin{aligned}
&\mathsf{P}\left\{\mathcal{E}_8(\tilde{L}) \cap \mathcal{E}_5^c(\hat{M}_2) \cap \mathcal{E}_7^c \cap \left(\bigcap_{j=0}^4 \mathcal{E}_j^c\right)\right\} \\
&= \mathsf{P}\left\{\mathcal{E}_8(\tilde{L}) \cap \mathcal{E}_5^c(\hat{M}_2) \cap \mathcal{E}_7^c \right. \\
&\qquad \left. \cap \left(\bigcap_{j=0}^4 \mathcal{E}_j^c\right) \cap \{\hat{M}_2 = M_2\}\right\} \\
&\quad + \mathsf{P}\left\{\mathcal{E}_8(\tilde{L}) \cap \mathcal{E}_5^c(\hat{M}_2) \cap \mathcal{E}_7^c \right. \\
&\qquad \left. \cap \left(\bigcap_{j=0}^4 \mathcal{E}_j^c\right) \cap \{\hat{M}_2 \neq M_2\}\right\} \\
&= \mathsf{P}\left\{\mathcal{E}_8(\tilde{L}) \cap \mathcal{E}_5^c(\hat{M}_2) \cap \mathcal{E}_7^c \right. \\
&\qquad \left. \cap \left(\bigcap_{j=0}^4 \mathcal{E}_j^c\right) \cap \{\hat{M}_2 = M_2\}\right\} \\
&= \mathsf{P}\left\{\mathcal{E}_8(\tilde{L}) \cap \mathcal{E}_5^c(M_2) \cap \mathcal{E}_7^c \right. \\
&\qquad \left. \cap \left(\bigcap_{j=0}^4 \mathcal{E}_j^c\right) \cap \{\hat{M}_2 = M_2\}\right\} \\
&\leq \mathsf{P}\{\mathcal{E}_8(\tilde{L}) \cap \mathcal{E}_5^c(M_2) \cap \mathcal{E}_7^c\} \\
&\stackrel{(a)}{=} \mathsf{P}\{\mathcal{E}_8(\tilde{L}) \cap \mathcal{E}_5^c(M_2) \cap \mathcal{E}_7^c \cap \{\tilde{L} = L\}\} \\
&= \mathsf{P}\{\mathcal{E}_8(L) \cap \mathcal{E}_5^c(M_2) \cap \mathcal{E}_7^c \cap \{\tilde{L} = L\}\} \\
&\leq \mathsf{P}\{\mathcal{E}_8(L) \cap \mathcal{E}_7^c\}.
\end{aligned}$$

(a) follows from the observation that $\mathcal{E}_5^c(M_2) \cap \mathcal{E}_7^c \cap \{\tilde{L} \neq L\} = \emptyset$. It remains to bound $\mathsf{P}\{\mathcal{E}_8(L) \cap \mathcal{E}_7^c\}$. Note that the analysis of this term is equivalent to analyzing the setup where U^n is the side information at node 0 and (U^n, Z^n) is the side information at node 2. Hence, $\mathsf{P}\{\mathcal{E}_8(L) \cap \mathcal{E}_7^c\} \to 0$ as $n \to \infty$ if

$$\tilde{R}_3 - R_3 < I(V; Z|U) - \delta(\epsilon).$$

We then obtain the rate region by eliminating $\tilde{R}_3$ and R_l.

C. Achievability proof of Theorem 3

As with the case for the triangular setting, the proof for this case follows the cascade setting closely. We will therefore include only the additional steps. We have a change of notation from the cascade setting. We will use U_1 instead of U

1) Codebook Generation:

- Fix $p(x, y, z, u_1, u_2, \hat{x}_1) = p(x, y, z)p(u_1|x, y) p(\hat{x}_1|u_1, x, y)p(u_2|z, u_1)$.
- For each $u_1^n(l)$, generate 2^{nR_3} $U_2^n(l_3)$ sequences, $l \in [1:2^{n\tilde{R}_3}]$, each according to $\prod_{i=1}^n p(u_{2i}|u_{1i})$. Partition the set of U_2^n into 2^{nR_3} bins, $\mathcal{B}_3(m_3)$.

2) Encoding: The additional encoding step is at node 2. Node 2 looks for an index L_3 such that $(u_1^n(L), u_2^n(L, L_3), Z^n) \in \mathcal{T}_\epsilon^{(n)}$. As before, if it finds more than one, it selects an index uniformly at random from the set of admissible indices. If it finds none, it outputs an index uniformly at random from $[1:2^{n\tilde{R}_3}]$. It then outputs the bin index m_3 such that $L_3 \in \mathcal{B}_3(m_3)$.

3) Decoding: Additional decoding is required at node 0. Node 0 looks the index $\hat{l}_3$ such that $(u_1^n(l), u_2^n(l, \hat{l}_3), x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$ and $\hat{l}_3 \in \mathcal{B}_3(m_3)$.

4) Analysis of Distortion: Let $\mathcal{E}_{\text{cascade}}$ denote the event that an error occurs in the forward cascade path. In addition, we define the following error events.

- $\mathcal{E}_{TW-1}(\hat{L}) := \{(U_1^n(\hat{L}), U_2^n(\hat{L}, l_3), Z^n) \notin \mathcal{T}_\epsilon^{(n)}$ for all $l_3 \in [1:2^{n\tilde{R}_3}]\}$.
- $\mathcal{E}_{TW-2}(\hat{L}) := \{(U_1^n(\hat{L}), U_2^n(\hat{L}, L_3), Z^n, X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$.
- $\mathcal{E}_{TW-3}(\hat{L}) := \{(U_1^n(\hat{L}), U_2^n(\hat{L}, \hat{l}_3), X^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}$ for some $\hat{l}_3 \in \mathcal{B}_3(M_3), \hat{l}_3 \neq L_3\}$.
- $\mathsf{P}(\mathcal{E}_{TW-1}(\hat{L}) \cap \mathcal{E}_{\text{cascade}}^c) = \mathsf{P}(\mathcal{E}_{TW-1}(L) \cap \mathcal{E}_{\text{cascade}}^c) \to 0$ as $n \to \infty$ if

$$\tilde{R}_3 > I(U_2; Z|U_1) + \delta(\epsilon).$$

- $\mathsf{P}(\mathcal{E}_{TW-2}(\hat{L}) \cap \mathcal{E}_{\text{cascade}}^c) = \mathsf{P}(\mathcal{E}_{TW-2}(L) \cap \mathcal{E}_{\text{cascade}}^c) \to 0$ as $n \to \infty$ by the strong Markov lemma [9].
- $\mathsf{P}(\mathcal{E}_{TW-3}(\hat{L}) \cap \mathcal{E}_{\text{cascade}}^c) = \mathsf{P}(\mathcal{E}_{TW-3}(L) \cap \mathcal{E}_{\text{cascade}}^c) \to 0$ as $n \to \infty$ if

$$\tilde{R}_3 - R_3 < I(U_2; X, Y|U_1) - \delta(\epsilon).$$

Finally, eliminating $\tilde{R}_3$ and R_l gives us the required rate region.

D. Achievability Proof of Theorem 4

The achievability proof for two-way triangular source coding combines the proofs of the triangular source coding case and the two-way cascade case. As it is largely similar to these proofs, we will not repeat it here. We will just mention that the codebook generation, encoding, decoding, and analysis of distortion for the forward path from node 0 to node 2 follows that of the triangular source coding case, while codebook generation, encoding, decoding, and analysis of distortion for the reverse path from node 2 to node 0 follows that of the two-way cascade source coding case, with (U_2, V) taking the role of U_2.

Appendix B
Cardinality Bounds

We provide cardinality bounds for Theorems 1–4 stated in the paper. The main tool we will use is the Fenchel–Eggleston–Caratheodory theorem [12].

A. Proof of Cardinality Bound for Theorem 1

For each x, y, we have

$$f_j(p_{X,Y|U}(x,y|u)) = \sum_u p(u)p(x,y|u) = p(x,y).$$

We therefore have $|\mathcal{X}||\mathcal{Y}|-1$ continuous functions of $p(x,y|u)$. These set of equations preserves the distribution $p(x,y)$ and hence, by Markovity, $p(x,y,z)$. Next, observe that the following are similarly continuous functions of $p(x,y|u)$:

$$\begin{aligned} I(U;X,Y|Z) &= H(X,Y|Z) - H(X,Y,Z|U) + H(Z|U) \\ I(X;\hat{X}_1,U|Y) &= H(X|Y) - H(X|U) + H(X,\hat{X}_1,Y|U) \\ \mathsf{E}d_1(X,\hat{X}_1) &= \sum_{x,\hat{x}} p(x,\hat{x}_1)d(x,\hat{x}_1) \\ \mathsf{E}d_2(X,\hat{X}_2) &= \sum_{x,y,u} p(x,y,u)d(x,g_2(x,u)). \end{aligned}$$

These equations give us four additional continuous functions and hence, by Fenchel–Eggleston–Caratheodory theorem, there exists a U' with cardinality of $|\mathcal{X}||\mathcal{Y}|+3$ such that all the constraints are satisfied. Note that this construction does not preserve $p(\hat{x}_1)$, but this does not change the rate-distortion region since the associated rate and distortion are preserved.

B. Proof of Cardinality Bound for Theorem 2

We will first give a bound for the cardinality of U. We look at the following continuous functions of $p(x,y|u)$:

$$\begin{aligned} f_j(p_{X,Y|U}(x,y|u)) &= \sum_u p(u)p(x,y|u) = p(x,y) \qquad \forall x,y \\ I(U;X,Y|Z) &= H(X,Y|Z) \\ &\quad - H(X,Y,Z|U) + H(Z|U) \\ I(X;\hat{X}_1,U|Y) &= H(X|Y) - H(X|U) \\ &\quad + H(X,\hat{X}_1,Y|U) \\ I(X,Y;V|U,Z) &= H(X,Y,Z|U) - H(Z|U) \\ &\quad - H(X,Y,V,Z|U) + H(V,Z|U) \\ \mathsf{E}d_1(X,\hat{X}_1) &= \sum_{x,\hat{x}} p(x,\hat{x}_1)d(x,\hat{x}_1) \\ \mathsf{E}d_2(X,\hat{X}_2) &= \sum_{x,y,u,v} p(x,y,u,v)d(x,g_2(x,u)). \end{aligned}$$

From these equations, there exists a U' with $|\mathcal{U}'| \le |\mathcal{X}||\mathcal{Y}|+4$ such that the equations are satisfied. Note that the new U' induces a new V'. For each $U'=u$, consider the following continuous functions of $p(x,y|u,v)$:

$$\begin{aligned} p(x,y|u) &= \sum_v p(v|u)p(x,y|v,u) \\ I(X,Y;V|U=u,Z) &= H(X,Y|U=u,Z) \\ &\quad - H(X,Y|V,U=u,Z) \\ \mathsf{E}(d_2(X,\hat{X}_2)|U=u) &= \sum_{x,y,v} p(x,y,v|u)d(x,g_2(x,u)). \end{aligned}$$

From this set of equations, we see that for each $U'=u$, it suffices to consider V' such that $|\mathcal{V}'| \le |\mathcal{X}||\mathcal{Y}|+1$. Hence, the overall cardinality bound on V is $|\mathcal{V}| \le (|\mathcal{X}||\mathcal{Y}|+4)(|\mathcal{X}||\mathcal{Y}|+1)$. The joint $p(x,y,z)$ is preserved due to the Markov chain $(V,U)-(X,Y)-Z$.

C. Proof of Cardinality Bound for Theorem 3

The cardinality bounds on U_1 follow similar analysis as in the cascade source coding case. The proof is therefore omitted. For each $U_1=u_1$, the following are continuous functions of $p(z|u_2,u_1)$:

$$\begin{aligned} p(z|u_1) &= \sum_{u_2} p(u_2|u_1)p(z|u_2,u_1) \\ I(U_2;Z|U_1=u_1,X,Y) &= H(Z|U_1=u_1,X,Y) \\ &\quad - H(Z|U_1=u_1,U_2,X,Y) \\ \mathsf{E}(d_3(Z,\hat{Z})|U_1=u_1) &= \sum_{x,y,z,u_2} (p(x,y,z,u_2|u_1). \\ &\qquad d(z,g_3(x,y,u_1,u_2)))\,. \end{aligned}$$

From this set of equations, we see that for each $U_1=u_1$, it suffices to consider U_2' such that $|\mathcal{U}_2'| \le |\mathcal{Z}|+1$. Hence, the overall cardinality bound on U_2 is $|\mathcal{U}_2| \le |\mathcal{U}_1|(|\mathcal{Z}|+1)$. The joint $p(x,y,z)$ is preserved due to the Markov chains $U_1-(X,Y)-Z$ and $U_2-(Z,U_1)-(X,Y)$.

D. Proof of Cardinality Bound for Theorem 4

The cardinality bounds follow similar steps to those for the first three theorems. For the cardinality bound for $|\mathcal{U}_2|$, we find a cardinality bound for each $U_1=u_1$ and $V=v$. Details of the proof are omitted.

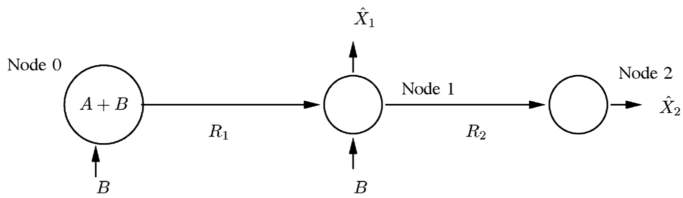

Fig. 8. Cascade source coding setting for the optimization problem in Corollary 1. $\hat{X}_1$ and $\hat{X}_2$ are lossy reconstructions of $A+B$.

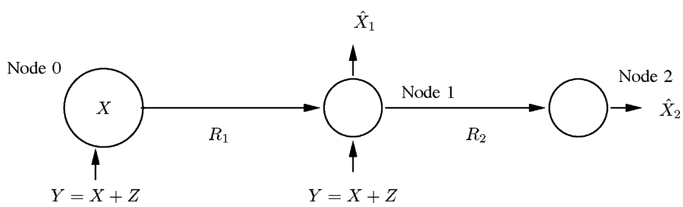

Fig. 9. Cascade source coding setting for the optimization problem in Corollary 1. $\hat{X}_1$ and $\hat{X}_2$ are lossy reconstructions of X and Z is independent X.

Appendix C
Alternative Characterizations of Rate-Distortion Regions in Corollaries 1 and 2

In this Appendix, we show that the rate-distortion regions in Corollaries 1 and 2 can alternatively be characterized by transforming them into equivalent problems found in [4], where explicit characterizations were given. We focus on the cascade case (Corollary 1), since the triangular case follows by the same analysis.

Fig. 8 shows the cascade source coding setting which the optimization problem in Corollary 1 solves.

In [4], explicit characterization of the cascade source coding setting in Fig. 9 was given.

We now show that the setting in Fig. 8 can be transformed into the setting in Fig. 9. First, we note that for the setting in Fig. 9, the rate-distortion regions are the same regardless of whether the sources are (X, Y) or $(X, \alpha Y)$ where $\alpha \neq 0$ since the nodes can simply scale Y by an appropriate constant.

Next, for Gaussian sources, the two settings are equivalent if we can show that the covariance matrix of $(X, \alpha Y)$ can be made equal to the covariance matrix of $(A+B, B)$. Equating coefficients in the covariance matrix, we require the following:

$$\begin{aligned} \sigma_X^2 &= \sigma_A^2 + \sigma_B^2 \\ \alpha\sigma_X^2 &= \sigma_B^2 \\ \alpha^2(\sigma_X^2 + \sigma_Z^2) &= \sigma_B^2. \end{aligned}$$

Solving these equations, we see that $\alpha = \sigma_B^2/(\sigma_A^2 + \sigma_B^2)$ and $\sigma_Z^2 = (\sigma_B^2 - \alpha^2\sigma_X^2)/\alpha^2$. Since $(\sigma_B^2 - \alpha^2\sigma_X^2) \geq 0$, this choice of σ_Z^2 is valid, which completes the proof.

Appendix D
Proof of Converse for Triangular Source Coding With Helper

Given a $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3}, 2^{nR_h}, D_1, D_2)$ code, define $U_{hi} = (Y^{i-1}, Z^{i-1}, Z_{i+1}^n, M_h)$, $U_{1i} = (X^{i-1}, M_2)$, and $U_{2i} = (U_{hi}, U_{1i}, M_3)$. Observe that we have the required Markov conditions $(X_i, Z_i) - Y_i - U_{hi}$ and $Z_i - (X_i, Y_i, U_{hi}) - (U_{1i}, U_{2i})$. For the helper condition, we have

$$\begin{aligned} nR_h &\geq I(M_h; Y^n|Z^n) \\ &= \sum_{i=1}^n \big(H(Y_i|Z_i) - H(Y_i|Y^{i-1}, M_h, Z^n)\big) \\ &= \sum_{i=1}^n I(U_{hi}; Y_i|Z_i). \end{aligned}$$

For the other rates, we have

$$\begin{aligned} nR_1 &\geq H(M_1) \\ &\geq H(M_1|Y^n, Z^n) \\ &= H(M_1, M_2|Y^n, Z^n) \\ &= I(X^n; M_1, M_2|Y^n, Z^n) \\ &= \sum_{i=1}^n I(X_i; M_1, M_2|X^{i-1}, Y^n, Z^n) \\ &= \sum_{i=1}^n \big(H(X_i|X^{i-1}, Y^n, Z^n) \\ &\qquad - H(X_i, Y_i|X^{i-1}, Y^n, Z^n, M_1, M_2)\big) \\ &= \sum_{i=1}^n (H(X_i|Y_i, Z_i) \\ &\qquad - H(X_i, Y_i|X^{i-1}, Y^n, Z^n, M_1, M_2)) \\ &\overset{(a)}{=} \sum_{i=1}^n (H(X_i|Y_i) \\ &\qquad - H(X_i, Y_i|X^{i-1}, Y^n, \hat{X}_{1i}, Z^n, M_1, M_2, M_h)\Big) \\ &\geq \sum_{i=1}^n \Big(H(X_i|Y_i, U_{hi}) - H(X_i|\hat{X}_{1i}, Y_i, U_{1i}, U_{hi})\Big) \\ &= \sum_{i=1}^n I(X_i; \hat{X}_{1i}, U_{1i}|Y_i, U_{hi}). \end{aligned}$$

(a) follows from the Markov chain condition. Next

$$\begin{aligned} nR_2 &\geq H(M_2|M_h) \\ &\geq H(M_2|Z^n, M_h) \\ &= I(X^n, Y^n; M_2|Z^n, M_h) \\ &= \sum_{i=1}^n I(X_i, Y_i; M_2|Z^n, X^{i-1}, Y^{i-1}, M_h) \\ &= \sum_{i=1}^n \big(H(X_i, Y_i|Z^n, X^{i-1}, Y^{i-1}, M_h) \\ &\qquad - H(X_i, Y_i|Z^n, X^{i-1}, Y^{i-1}, M_2, M_h)\big) \\ &= \sum_{i=1}^n (H(X_i, Y_i|Z_i, U_{hi}) - H(X_i, Y_i|Z_i, U_{1i}, U_{hi})) \\ &= \sum_{i=1}^n I(X_i, Y_i; U_{1i}|Z_i, U_{hi}). \end{aligned}$$

Next

$$\begin{aligned} nR_3 &\geq H(M_3) \\ &\geq H(M_3|M_2, M_h, Z^n) \\ &= I(X^n, Y^n; M_3|M_2, M_h Z^n) \\ &= \sum_{i=1}^{n} \big(H(X_i, Y_i|M_2, M_h, Z^n, X^{i-1}, Y^{i-1}) \\ &\quad - H(X_i, Y_i|M_2, M_3, M_h, Z^n, X^{i-1}, Y^{i-1})\big) \\ &= \sum_{i=1}^{n} I(X_i, Y_i; U_{2i}|U_{1i}, U_{hi}, Z_i). \end{aligned}$$

Finally, it remains to show that the joint probability distribution induced by our choice of auxiliary random variables $p(x)p(y|x)p(z|y)p(u_h|y)p(u|x, y, u_h)p(\hat{x}_1, u_2|x, y, u_1, u_h)$ can be decomposed into the required form. This step follows closely the similar step in the proof of Theorem 2, which we therefore omit.

Acknowledgment

The authors would like to thank the anonymous reviewers for helpful comments.

References

[1] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 3, pp. 299–308, May 1981.

[2] D. Vasudevan, C. Tian, and S. N. Diggavi, "Lossy source coding for a cascade communication system with side informations," in *Proc. Allerton Conf. Commun. Control Comput.*, 2006, DOI: 10.1.1.94.1174.

[3] P. Cuff, H.-I. Su, and A. El Gamal, "Cascade multiterminal source coding," in *Proc. IEEE Int. Conf. Symp. Inf. Theory*, Piscataway, NJ, 2009, pp. 1199–1203.

[4] H. Permuter and T. Weissman, "Cascade and triangular source coding with side information at the first two nodes," [Online]. Available: arXiv:1001.1679v

[5] P. Ishwar and S. S. Pradhan, "A relay-assisted distributed source coding problem," in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, 2008, DOI: 10.1109/ITA.2008.4601039.

[6] A. H. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 735–740, Nov. 1985.

[7] A. El Gamal and Y. H. Kim, "Lectures on network information theory," 2010 [Online]. Available: http://arxiv.org/abs/1001.3404

[8] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[9] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, Schl. Electr. Comput. Eng., Cornell Univ., Itahca, NY, 1978.

[10] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—Part II: General sources," *Inf. Control*, no. 38, pp. 60–80, 1978.

[11] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," Technion Dept. Electr. Eng., Haifa, Israel, CCIT Rep. 419, EE Pub. 1358, 2003.

[12] H. G. Eggleston, *Convexity*. Cambridge, U.K.: Cambridge Univ. Press, 1958.

Yeow-Khiang Chia received the M.Eng. degree in electrical and electronics engineering from Imperial College, London, U.K. Currently, he is working towards the Ph.D. degree at the Department of Electrical Engineering, Stanford University, Stanford, CA.

From 2005 to 2006, he was a Research Engineer with the Institute for Infocomm Research, Singapore.

Mr. Chia is a recipient of the Stanford Graduate Fellowship (SGF) and the National Science Scholarship (NSS) from the Agency for Science, Technology and Research, Singapore (A*STAR).

Haim H. Permuter (M'08) received the B.Sc. (*summa cum laude*) and M.Sc. (*summa cum laude*) degrees in electrical and computer engineering from Ben-Gurion University, Beer-Sheva, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2008.

Between 1997 and 2004, he was an Officer at a Research and Development unit of the Israeli Defense Forces. He is currently a Senior Lecturer at Ben-Gurion University.

Dr. Permuter is a recipient of the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and the 2009 U.S.–Israel Binational Science Foundation Bergmann Memorial Award.

Tsachy Weissman (S'99–M'02–SM'07) received the B.Sc. (*summa cum laude*) and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 1997 and 2001, respectively.

He then worked at Hewlett-Packard Laboratories with the information theory group before joining Stanford University, Stanford, CA, where he has been on the faculty of the Electrical Engineering Department since 2003, spending two academic years 2007–2009 on leave at the Technion. His research is focused on information theory, statistical signal processing, the interplay between them, and their applications.

Dr. Weissman received a National Science Foundation (NSF) CAREER award, a joint IT/COM societies best paper award, a Horev fellowship for Leaders in Science and Technology, and a Henry Taub prize for excellence in research. He is on the editorial board of the IEEE TRANSACTIONS ON INFORMATION THEORY, serving as an Associate Editor for Shannon Theory.