Cascade and Triangular Source Coding with Side Information at the First Two Nodes

Haim Permuter and Tsachy Weissman

Abstract

We consider the cascade and triangular rate-distortion problem where side information is known to the source encoder and to the first user but not to the second user. We characterize the rate-distortion region for these problems. For the quadratic Gaussian case, we show that it is sufficient to consider jointly Gaussian distributions, a fact that leads to an explicit solution.

Index Terms

Cascade source coding, empirical coordination, quadratic Gaussian, Pareto frontier, source coding, side information, rate distortion, triangular source coding

I. Introduction

Yamamoto [1] considered the cascade source coding problem, where a source sends a message to User 1, and then User 1 sends a message to User 2. In this paper, we extend Yamamoto's cascade source coding problem to the case where side information is known to the source and to User 1, but not to User 2. The problem is depicted in Fig. 1.

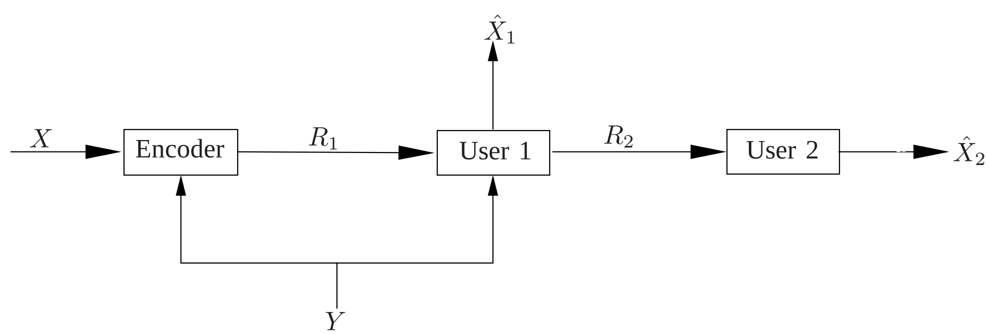

Fig. 1. A cascade rate distortion problem with three nodes (encoder, User 1, User 2), where the first two nodes have side information Y. User 1 and User 2 need to reconstruct the sourse X, within distortion criteria.

More recently, Vasudevan, Tian and Diggavi [2] considered the cascade source coding problem, where side information, Y, is known to the source encoder and to User 1, additional side information Z is known to User 2, and the Markov chain $X-Z-Y$ holds. Vasudevan et al. [2] provided an inner and an outer bound and showed that

Author's emails: haimp@bgu.ac.il, tsachy@stanford.edu

TABLE I

LITERATURE OVERVIEW OF CASCADE SOURCE CODING WITH SIDE INFORMATION AS SHOWN IN FIG. 2

Switch a	Switch b	Switch c	Gaussian quadratic case	General case
open	open	open	Solved [1]	Solved [1]
open	open	closed	Solved [2]	Upper and lower bounds [2]
open	closed	open	Upper and lower bounds [3]	Upper and lower bounds [3]
open	closed	closed	Solved [2]	Upper and lower bounds [2]
closed	open	open	Solved [1]	Solved [1]
closed	open	closed	Solved [2]	Upper and lower bounds [2]
closed	closed	open	Section IV	Section II

the bounds coincide for the Gaussian case. Cuff, Su and El-Gammal [3] considered the cascade problem where the side information is known only to the intermediate node and provided an inner and an outer bound. An additional related problem, which was considered and solved in [4], is that of cascade source coding when side information is known to all nodes with a limited rate. Table I summarizes the literature on cascade source coding with side information.

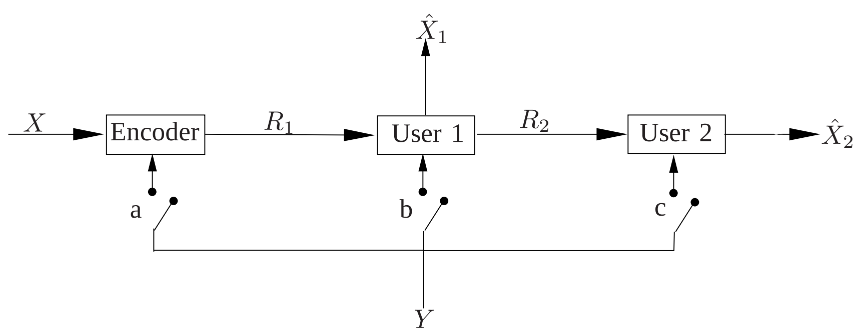

Fig. 2. A cascade rate distortion problem with several options of side information. Table I summarizes the lietrtaure on this problem.

Of special interest in lossy source coding is the Gaussian case with quadratic distortion, which in many source coding problems results in an analytical solution such as in the Wyner-Ziv problem [5] where side information is available to the decoder, the Heegard-Berger problem [6] where side information at the decoder may be absent, Kaspi's problem [7], [8] where side information is known to the encoder and may or may not be known to the decoder, the multiple description problem [9], [10], the two-way source coding problem [11], the multi-terminal problem [12] [13], the CEO problem [14]–[16], rate distortion with a helper [17], [18], and successive refinement [19] and its extension to successive refinement for the Wyner-Ziv problem [20].

Our main result in this paper is that the achievable region for the problem depicted in Fig. 1 is given by $\mathcal{R}(D_1, D_2)$, which is defined as the set of all rate-pairs (R_1, R_2) that satisfy

$$R_2 \quad \geq \quad I(Y, X; \hat{X}_2), \tag{1}$$

$$R_1 \geq I(X;\hat{X}_1,\hat{X}_2|Y), \tag{2}$$

for some joint distribution $P(x,y)P(\hat{x}_1,\hat{x}_2|x,y)$ for which

$$\mathbb{E}d_i(X,\hat{X}_i) \leq D_i, \quad i=1,2. \tag{3}$$

An extension of the cascade source coding problem is the triangular setting [21], where there is an additional direct link from the source encoder to User 2. We solve this problem where side information exists at the source encoder and User 1, but not at User 2.

The remainder of the paper is organized as follows. In Section II, we formally define the cascade problem and present the theorem establishing the achievable region. In Section III, we provide a converse and achievability proofs of the theorem, and in Section IV we explicitly compute the rate region for the Gaussian case. In Section V we extend our result to the triangular case (cf. Fig. 5), and in Section VI we further extend the results to multiple users and discuss the corresponding empirical coordination problem.

II. Cascade rate distortion: Problem definitions and main results

Here we formally define the cascade rate-distortion problem where side information is known to the source encoder and to User 1. We present a single-letter characterization of the achievable region. We use the regular definitions of rate distortion, and we follow the notation of [22]. The source sequences $\{X_i \in \mathcal{X},\ i=1,2,\cdots\}$, and the side information sequence $\{Y_i \in \mathcal{Y},\ i=1,2,\cdots\}$ are discrete random variables drawn from finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. The random variables (X_i,Y_i) are i.i.d. $\sim P(x,y)$. Let $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{X}}_2$ be the reconstruction alphabets, and $d_i:\ \mathcal{X}\times\hat{\mathcal{X}}_i \to [0,\infty)$, $i=1,2$, are single letter distortion measures. Distortion between sequences is defined in the usual way

$$d_i(x^n,\hat{x}_i^n) = \frac{1}{n}\sum_{j=1}^{n} d_i(x_j,\hat{x}_{i,j}), \quad i=1,2. \tag{4}$$

Let $\mathcal{M}_i$ denote a set of positive integers $\{1,2,..,M_i\}$ for $i=1,2$.

Definition 1 (Cascade rate distortion code with side information at the first two nodes): An (n,M_1,M_2,D_1,D_2) code for source X and side information Y consists of two encoders

$$\begin{aligned} f_1 &: \mathcal{X}^n\times\mathcal{Y}^n \to \mathcal{M}_1 \\ f_2 &: \mathcal{Y}^n\times\mathcal{M}_1 \to \mathcal{M}_2 \end{aligned} \tag{5}$$

and two decoders

$$\begin{aligned} g_1 &: \mathcal{Y}^n\times\mathcal{M}_1 \to \hat{\mathcal{X}}_1^n \\ g_2 &: \mathcal{M}_2 \to \hat{\mathcal{X}}_2^n \end{aligned} \tag{6}$$

such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} d_j(X_i,\hat{X}_{j,i})\right] \leq D_j, \quad j=1,2 \tag{7}$$

The rate pair (R_1, R_2) of the (n, M_1, M_2, D_1, D_2) code is defined by

$$R_i = \frac{1}{n} \log M_i; \quad i = 1, 2. \tag{8}$$

Definition 2: Given a distortion pair (D_1, D_2), a rate pair (R_1, R_2) is said to be *achievable* if, for any $\epsilon > 0$, and sufficiently large n, there exists an $(n, 2^{nR_1}, 2^{nR_2}, D_1 + \epsilon, D_2 + \epsilon)$ code for the source X with side information Y.

Definition 3: The (operational) achievable region $\mathcal{R}^O(D_1, D_2)$ of cascade rate distortion is the closure of the set of all achievable rate pairs.

Theorem 1 is the main result of this work.

Theorem 1: For the cascade rate distortion problem with side information at the source and User 1, as depicted in Fig. 1, the achievable region is given by

$$\mathcal{R}^O(D_1, D_2) = \mathcal{R}(D_1, D_2), \tag{9}$$

where the region $\mathcal{R}(D_1, D_2)$ is defined in (1)-(3).

III. Proof of Theorem 1

Achievability: The proof follows classical arguments, and therefore the technical details will be omitted. We describe only the coding structure and justify why the indicated region is achievable. We fix a joint distribution $P_{X,Y,\hat{X}_1,\hat{X}_2}$ for which (3) holds, and an $\epsilon > 0$, and we show that there exists a code with rates

$$R_2 = I(Y, X; \hat{X}_2) + \epsilon, \tag{10}$$

$$R_1 = I(X; \hat{X}_1, \hat{X}_2 | Y) + 3\epsilon, \tag{11}$$

complying with the distortion constraints.

Generate randomly $2^{n(I(X,Y;\hat{X}_2)+\epsilon)}$ codewords using an i.i.d. $\sim P_{\hat{X}_2}$. Then bin the codewords into $2^{n(I(X;\hat{X}_2|Y)+2\epsilon)}$ bins. In each bin, there are $2^{n(I(X,Y;\hat{X}_2)-I(X;\hat{X}_2|Y)-\epsilon)} = 2^{n(I(Y;\hat{X}_2)-\epsilon)}$ codewords. In addition, for any typical sequences $y^n, \hat{x}_2^n$ generate $2^{n(I(X;\hat{X}_1|Y,\hat{X}_2)+\epsilon)}$ codewords using the pmf $P(\hat{x}_1^n | y^n, \hat{x}_2^n) = \prod_{i=1}^n P_{\hat{X}_1|Y,\hat{X}_2}(\hat{x}_{1,i} | y_i, \hat{x}_{2,i})$.

The source-encoder receives the sequences x^n, y^n and first looks for a codeword $\hat{x}_2^n$ that is jointly typical with x^n, y^n. If there is such a codeword, the source encoder sends the index of the bin that includes this codeword to User 1. User 1 looks which codeword in the received bin is jointly typical with the side information y^n. Since there are less than $2^{n(I(Y;\hat{X}_2)}$ in the bin, with high probability only one codeword will be jointly typical with y^n and it would be the codeword sent by the encoder. User 1 then forwards the codeword to User 2.

Now we can think of a new problem where the source-encoder and User 1 have side information $Y^n, \hat{X}_2^n$ and hence a rate $I(X; \hat{X}_1 | Y, \hat{X}_2) + \epsilon$ is needed to generate $\hat{X}_1^n$ that is jointly typical with $(X^n, Y^n, \hat{X}_2)$. Therefore, a total rate to User 1 of $R_1 = I(X; \hat{X}_2 | Y) + 2\epsilon + I(X; \hat{X}_1 | Y, \hat{X}_2) + \epsilon = I(X; \hat{X}_1, \hat{X}_2 | Y) + 3\epsilon$ is needed, and an additional rate $R_2 = I(Y, X; \hat{X}_2) + \epsilon$ is needed from User 1 to User 2.

Converse: Assume that we have an $(n, M_1 = 2^{nR_1}, M_2 = 2^{nR_2}, D_1, D_2)$ code as in Definition 1. We will show the existence of a joint distribution $P_{X,Y,\hat{X}_1,\hat{X}_2}$ that satisfies (1)-(3). Denote $T_1 = f_1(X^n, Y^n) \in \{1, ..., 2^{nR_1}\}$, and $T_2 = f_2(T_1, Y^n) \in \{1, ..., 2^{nR_2}\}$. Then,

$$\begin{aligned} nR_2 &\geq H(T_2) \\ &\geq I(X^n, Y^n; T_2) \\ &= \sum_{i=1}^{n} H(X_i, Y_i) - H(X_i, Y_i | T_2, X^{i-1}, Y^{i-1}) \\ &\overset{(a)}{=} \sum_{i=1}^{n} H(X_i, Y_i) - H(X_i, Y_i | \hat{X}_{2,i}, T_2, X^{i-1}, Y^{i-1}) \\ &\geq \sum_{i=1}^{n} I(X, Y; \hat{X}_{2,i}), \end{aligned} \tag{12}$$

where equality (a) follows from the fact that the reconstruction at time i, $\hat{X}_{2,i}$, is a deterministic function of T_2. Now consider

$$\begin{aligned} nR_1 &\geq H(T_1) \\ &\geq H(T_1|Y^n) \\ &\overset{(a)}{=} H(T_1, T_2|Y^n) \\ &\geq I(X^n; T_1, T_2|Y^n) \\ &= \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|Y^n, T_1, T_2, X^{i-1}) \\ &\overset{(b)}{=} \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|Y^n, T_1, T_2, X^{i-1}, \hat{X}_{1,i}, \hat{X}_{2,i}) \\ &\geq \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|Y_i, \hat{X}_{1,i}, \hat{X}_{2,i}) \\ &= \sum_{i=1}^{n} I(X_i; \hat{X}_{1,i}, \hat{X}_{2,i}|Y_i), \end{aligned} \tag{13}$$

where equality (a) follows from the fact that T_2 is a deterministic function of T_1 and Y^n, and, similarly, equality (b) follows from the fact that $\hat{X}_{1,i}$ and $\hat{X}_{2,i}$ are deterministic functions of (T_1, Y^n) and T_2, respectively.

The proof is concluded in the standard way by letting Q be a random variable independent of X^n, Y^n, uniformly distributed over the set $\{1, 2, 3, .., n\}$, and considering the joint distribution of $X_Q, Y_Q, \hat{X}_{1,Q}, \hat{X}_{2,Q}$. For this joint distribution, inequalities (12) and (13) imply that (1) and (2) hold, respectively, and (7) implies that (3) holds. ■

IV. Cascade rate distortion: the Gaussian case

In this section we explicitly calculate the rate region $\mathcal{R}(D_1, D_2)$ for the cases where X and Y are jointly Gaussian and the distortion is the square-error distortion. The converse and the achievability in the previous sections are proved for the finite alphabet case, but it can be extended to the Gaussian case [5].

Our first step in finding the achievable region for the quadratic Gaussian case is to show that it suffices to consider only jointly Gaussian distributions $P_{X,Y,\hat{X}_1,\hat{X}_2}$ in order to exhaust the rate region. Then we solve an optimization problem to find the achievable rate-region explicitly.

Lemma 2 (Optimality of jointly Gaussian distributions): For the quadratic Gaussian cascade rate-distortion problem with side information known to the source-encoder and to User 1, i.e., X, Y are jointly Gaussian and $d_1(x,\hat{x}_1) = (x-\hat{x}_1)^2$, $d_2(x,\hat{x}_2) = (x-\hat{x}_2)^2$, it suffices to consider only jointly Gaussian distributions $P_{X,Y,\hat{X}_1,\hat{X}_2}$ in order to exhaust the rate region $\mathcal{R}(D_1,D_2)$ given in (1)-(3).

Proof: Let us fix a point (R_1,R_2,D_1,D_2) in the rate region and let $P_{X,Y,\hat{X}_1,\hat{X}_2}$ be a joint distribution that satisfies (1)-(3). Such a distribution must exist since Inequalities (1)-(3) define the rate region (Theorem 1). Let K denote the covariance matrix induced by $P_{X,Y,\hat{X}_1,\hat{X}_2}$ and let $\tilde{P}_{X,Y,\hat{X}_1,\hat{X}_2}$ denote a normal joint distribution with mean zero and covariance matrix K. Now let us show that (1)-(3) also hold where the joint distribution is $\tilde{P}_{X,Y,\hat{X}_1,\hat{X}_2}$. Inequality (3) is automatically satisfied, since it depends on the distribution of $(X,Y,\hat{X}_1,\hat{X}_2)$ only through the covariance matrix K. Consider,

$$\begin{aligned} R_1 &\geq I(X;\hat{X}_1,\hat{X}_2|Y), \\ &= h(X|Y) - h(X|\hat{X}_1,\hat{X}_2,Y), \\ &\overset{(a)}{=} h(X|Y) - h(X-(\alpha_1\hat{X}_1+\alpha_2\hat{X}_2+\alpha_3 Y)|\hat{X}_1,\hat{X}_2,Y), \\ &\overset{(b)}{\geq} h(X|Y) - h(X-(\alpha_1\hat{X}_1+\alpha_2\hat{X}_2+\alpha_3 Y)) \\ &\overset{(c)}{\geq} h(X|Y) - h_{\tilde{P}}(X-(\alpha_1\hat{X}_1+\alpha_2\hat{X}_2+\alpha_3 Y)) \\ &\overset{(d)}{=} I_{\tilde{P}}(X;\hat{X}_1,\hat{X}_2|Y), \end{aligned} \tag{14}$$

equality (a) is true for any set of scalars $(\alpha_1,\alpha_2,\alpha_3)$ and in particular if we choose those that are the linear estimator of X given $\hat{X}_1,\hat{X}_2,Y$. Note that the coefficients $(\alpha_1,\alpha_2,\alpha_3)$ and the variance $E(X-(\alpha_1\hat{X}_1+\alpha_2\hat{X}_2+\alpha_3 Y))^2$ are a function only of the covariance matrix K. Inequality (b) follows from the fact that conditioning reduces entropy, and (c) follows from the fact that, given a variance, the Gaussian distribution maximizes the differential entropy. The term $I_{\tilde{P}}(X;\hat{X}_1,\hat{X}_2|Y)$ denotes the mutual information induced by the Gaussian distribution $\tilde{P}_{X,Y,\hat{X}_1,\hat{X}_2}$, and equality (d) follows from the fact that for the Gaussian distribution the error, i.e., $X-(\alpha_1\hat{X}_1+\alpha_2\hat{X}_2+\alpha_3 Y)$, is independent of the observations $\hat{X}_1,\hat{X}_2,Y$.

Similarly, we have

$$\begin{aligned} R_2 &\geq I(Y,X;\hat{X}_2) \\ &= I(Y;\hat{X}_2) + I(X;\hat{X}_2|Y) \\ &\geq I_{\tilde{P}}(Y;\hat{X}_2) + I_{\tilde{P}}(X;\hat{X}_2|Y), \end{aligned} \tag{15}$$

where the last inequality follows from the same steps as (14). ■

The next theorem provides an explicit expression for the Gaussian case. The proof is provided in Appendix A and is based on Lemma 2 and on solving an optimization problem with quadratic constraints and a linear objective.

Theorem 3 (Cascade Gaussian case): The rate region of the cascade source coding with side information at the first two nodes, where the source X and the side information $Y = X + Z$ are jointly Gaussian distributed, where X and Z are mutually independent, and the distortion is quadratic, is given by

$$R_1(D_1, D_2, R_2) = \frac{1}{2}\max\left(\log\frac{\sigma^2_{X|Y}}{\sigma^2_{X|W,Y}}, \log\frac{\sigma^2_{X|Y}}{D_1}, 0\right), \tag{16}$$

where $\sigma^2_{X|W,Y}$ is given by the following four cases

$$\sigma^2_{X|W,Y}(D_1, D_2, R_2) = \begin{cases} \left(\frac{2^{2R_2}D_2 - \sigma^2_X}{\sigma^2_Z\sigma^2_X\alpha^2} + \sigma^{-2}_{X|Y}\right)^{-1}, & \text{if } D_2 \le \sigma^2_{X|Y} \text{ and } \frac{\sigma^2_X}{D_2} \le 2^{2R_2} \le \frac{\sigma^2_Z(\sigma^2_X - D_2)}{\sigma^2_Z\sigma^2_X - D_2\sigma^2_Z - D_2\sigma^2_X}\frac{\sigma^2_X}{D_2} \\ D_2, & \text{if } D_2 \le \sigma^2_{X|Y} \text{ and } 2^{2R_2} \ge \frac{\sigma^2_Z(\sigma^2_X - D_2)}{\sigma^2_Z\sigma^2_X - D_2\sigma^2_Z - D_2\sigma^2_X}\frac{\sigma^2_X}{D_2} \\ \left(\frac{2^{2R_2}D_2 - \sigma^2_X}{\sigma^2_Z\sigma^2_X\alpha^2} + \sigma^{-2}_{X|Y}\right)^{-1}, & \text{if } D_2 \ge \sigma^2_{X|Y} \text{ and } \frac{\sigma^2_X}{D_2} \le 2^{2R_2} \le \frac{\sigma^4_X}{\sigma^2_X D_2 + \sigma^2_Z D_2 - \sigma^2_X\sigma^2_Z} \\ \sigma^2_{X|Y}, & \text{if } D_2 \ge \sigma^2_{X|Y}, \text{ and } 2^{2R_2} \ge \frac{\sigma^4_X}{\sigma^2_X D_2 + \sigma^2_Z D_2 - \sigma^2_X\sigma^2_Z} \end{cases} \tag{17}$$

and $\alpha = \left(\frac{\sigma_Z}{\sigma_X}\sqrt{\frac{\sigma^2_X - D_2}{D_2 - \sigma^2_X 2^{-2R_2}}} - 1\right)^{-1}$.

Fig. 3 depicts the regions for two specific values of D_1 and D_2 such that it captures all four cases of Eq. (17).

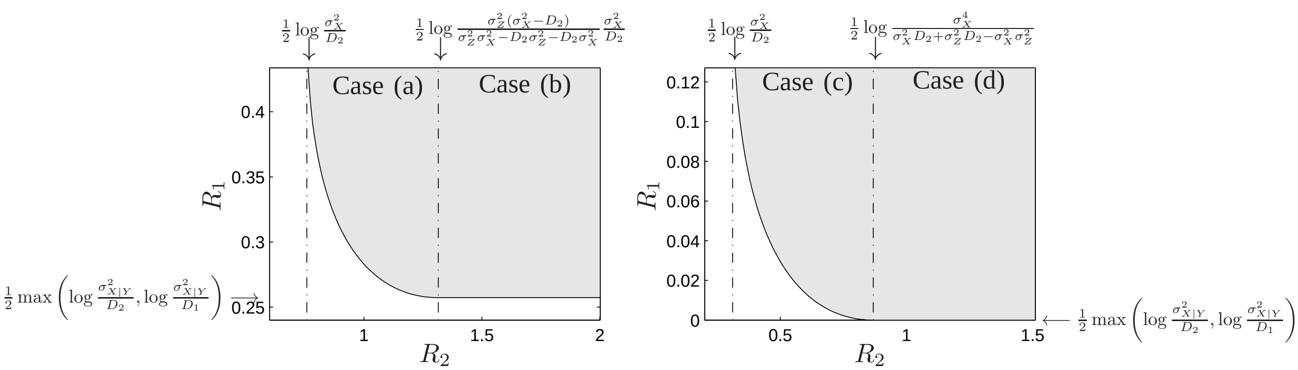

Fig. 3. The Gaussian quadratic rate region. The graph on the left hand side shows the rate region for the case where $\sigma^2_X = \sigma^2_Z = 1$, $D_2 = 0.35$ and $D_1 = 0.4$. Since $D_2 < \sigma^2_{X|Y}$, the rate region is given by Cases (a) and (b) in Eq. (17). The right hand side graph shows the rate region for the case where $\sigma^2_X = \sigma^2_Z = 1$, $D_2 = 0.65$ and $D_1 = 0.5$. Since $D_2 > \sigma^2_{X|Y}$, the rate region is given by Cases (c) and (d) in Eq. (17)

Now, let us consider several extreme cases that can be easily solved using Theorem 3.

1) Side information is independent of the source $X \perp Y$: This means that $\sigma^2_{X|Y} = \sigma^2_X$ and $\sigma^2_Z = \infty$. For such a case (17) becomes

$$\sigma^2_{X|W,Y}(D_1, D_2, R_2) = \begin{cases} \sigma^2_X, & \text{if } D_2 \le \sigma^2_X \text{ and } \frac{\sigma^2_X}{D_2} \le 2^{2R_2} \le \frac{\sigma^2_X}{D_2} \\ D_2, & \text{if } D_2 \le \sigma^2_X \text{ and } 2^{2R_2} \ge \frac{\sigma^2_X}{D_2} \\ \infty, & \text{if } D_2 \ge \sigma^2_X, \text{ and } 2^{2R_2} \ge 0 \end{cases} \tag{18}$$

and this implies that

$$R_1(D_1, D_2, R_2) = \frac{1}{2} \max \left(\log \frac{\sigma_X^2}{D_2}, \log \frac{\sigma_{X|Y}^2}{D_1}, 0 \right), \tag{19}$$

recovering a result that appears in the successive refinement source coding paper [19].

2) Side information equals the source, i.e., $X = Y$: For this case, $\sigma_{X|Y}^2 = 0$; hence $R_1 = 0$ and $2^{2R_2} \geq \frac{\sigma_X^2}{D_2}$, consistent with the well known rate distortion function of the Gaussian source.

3) $R_2 \to \infty$: If $D_2 \leq \sigma_{X|Y}^2$ then

$$R_1(D_1, D_2, R_2) = \frac{1}{2} \max \left(\log \frac{\sigma_{X|Y}^2}{D_2}, \log \frac{\sigma_{X|Y}^2}{D_1}, 0 \right), \tag{20}$$

and if $D_2 \geq \sigma_{X|Y}^2$

$$R_1(D_1, D_2, R_2) = \frac{1}{2} \max \left(\log \frac{\sigma_{X|Y}^2}{D_1}, 0 \right). \tag{21}$$

Note that for this case we can assume that the side information Y is known to all three nodes; hence only $\sigma_{X|Y}^2$ is manifested in the expression.

4) The message that User 2 receives depends only on the side information: In this extreme case, the rate R_2 and the distortion D_2 are large enough so that the message that User 2 receives depends only on the side information. This case is depicted in Fig. 4.

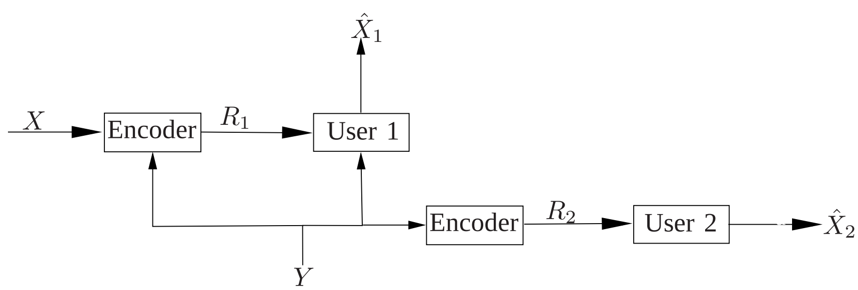

Fig. 4. An extreme case where the rate R_2 and the distortions D_2 are large enough so that the message that User 2 receives depends only on the side information.

For this extreme, the rate region is simply

$$\begin{aligned} R_1 &\geq I(X; \hat{X}_1 | Y), \\ R_2 &\geq I(Y; \hat{X}_2), \end{aligned} \tag{22}$$

for all joint Gaussian distributions that satisfy $\sigma_{X|Y,\hat{X}_1}^2 \leq D_1$ and $\sigma_{X|\hat{X}_2}^2 \leq D_2$.

More explicitly, this region is given by

$$D_2 \geq \frac{\sigma_X^2 (\sigma_X^2 2^{-2R_2} + \sigma_Z^2)}{\sigma_X^2 + \sigma_Z^2} \tag{23}$$

$$R_1 \geq \frac{1}{2} \max \left(\log \frac{\sigma_{X|Y}^2}{D_1}, 0 \right). \tag{24}$$

Indeed, if (23) holds, then according to Theorem 3, $R_1(D_1, D_2, R_2) = \frac{1}{2}\max\left(\log\frac{\sigma^2_{X|Y}}{D_1}, 0\right)$.

V. Triangular source coding with side information

In this section, we extend the cascade source coding discussed in previous sections by adding a direct link from the encoder to the second user, as depicted in Fig. 5. The definition of the code $(n, M_1, M_2, M_3, D_1, D_2)$ is similar to the one given in Def. 1 for the cascade case, with an additional message M_3 at rate R_3 sent from the source to User 2.

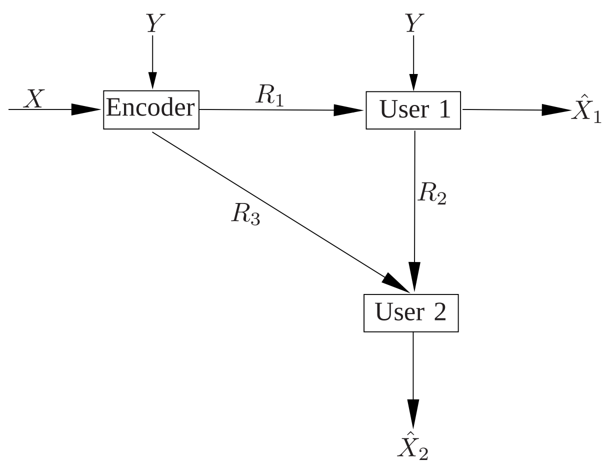

Fig. 5. A triangular rate distortion problem with three nodes (encoder, User 1, User 2), where side information Y is known to the encoder and User 1, but not to User 2. User 1 and User 2 need to reconstruct the sourse X to within distortion criteria.

A. Main theorem and its proof

Theorem 4 (The achievable rate region for the triangular case): The achievable region for the problem depicted in Fig. 5 is given by $\mathcal{R}_\Delta(D_1, D_2)$, which is defined as the set of all rate-triples (R_1, R_2, R_3) that satisfy

$$R_1 \geq I(X; \hat{X}_1, U|Y), \tag{25}$$

$$R_2 \geq I(Y, X; U), \tag{26}$$

$$R_3 \geq I(X; \hat{X}_2|U), \tag{27}$$

for some joint distribution $P(x, y)P(\hat{x}_1, \hat{x}_2, u|x, y)$ satisfying

$$\mathbb{E}d_i(X, \hat{X}_i) \leq D_i, \quad i = 1, 2, \tag{28}$$

where the cardinality of the auxiliary variable U may be bounded by $|U| \leq |\mathcal{X}||\mathcal{Y}||\hat{\mathcal{X}}_1||\hat{\mathcal{X}}_2| + 2$.

Lemma 5 below shows that one can restrict the joint distribution $P(x, y)P(\hat{x}_1, \hat{x}_2, u|x, y)$ to $P(x, y)P(\hat{x}_1, u|x, y)P(\hat{x}_2|x, u)$ without affecting the region.

Proof of Converse Part of Theorem 4: Assume that we have an $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_3}, D_1, D_2)$ code. We will show the existence of a joint distribution $P_{X,Y,U,\hat{X}_1,\hat{X}_2}$ that satisfies (25)-(28). Denote $T_1 = f_1(X^n, Y^n) \in \{1, ..., 2^{nR_1}\}$, and $T_2 = f_2(T_1, Y^n) \in \{1, ..., 2^{nR_2}\}$, and $T_3 = f_3(X^n, Y^n) \in \{1, ..., 2^{nR_3}\}$. Then,

$$\begin{aligned}
nR_1 &\geq H(T_1) \\
&\geq H(T_1|Y^n) \\
&\overset{(a)}{=} H(T_1, T_2|Y^n) \\
&\geq I(X^n; T_1, T_2|Y^n) \\
&= \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|Y^n, T_1, T_2, X^{i-1}) \\
&\overset{(b)}{=} \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|Y^n, T_1, T_2, X^{i-1}, \hat{X}_{1,i}, U_i) \\
&\geq \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|Y_i, \hat{X}_{1,i}, U_i) \\
&= \sum_{i=1}^{n} I(X_i; \hat{X}_{1,i}, U_i|Y_i), \qquad (29)
\end{aligned}$$

where equality (a) follows from the fact that T_2 is a deterministic function of T_1 and Y^n, and, similarly, equality (b) follows from the fact that $\hat{X}_{1,i}$ is a deterministic function of (T_1, Y^n) and from defining $\hat{U}_i \triangleq (T_2, X^{i-1}, Y^{i-1})$. Now, consider

$$\begin{aligned}
nR_2 &\geq H(T_2) \\
&\geq I(X^n, Y^n; T_2) \\
&= \sum_{i=1}^{n} H(X_i, Y_i) - H(X_i, Y_i|T_2, X^{i-1}, Y^{i-1}) \\
&\overset{(a)}{=} \sum_{i=1}^{n} H(X_i, Y_i) - H(X_i, Y_i|U_i) \\
&\geq \sum_{i=1}^{n} I(X, Y; U_i), \qquad (30)
\end{aligned}$$

where equality (a) follows from definition of $U_i = (T_2, X^{i-1}, Y^{i-1})$. In addition, consider

$$\begin{aligned}
nR_3 &\geq H(T_3) \\
&\geq H(T_3|T_2) \\
&\geq I(X^n, Y^n; T_3|T_2) \\
&= \sum_{i=1}^{n} H(X_i, Y_i|T_2, X^{i-1}, Y^{i-1}) - H(X_i, Y_i|T_2, T_3, X^{i-1}, Y^{i-1}) \\
&\overset{(a)}{=} \sum_{i=1}^{n} H(X_i, Y_i|U_i) - H(X_i, Y_i|\hat{X}_{2,i}, U_i)
\end{aligned}$$

$$\geq \sum_{i=1}^{n} I(X,Y;\hat{X}_{2,i}|U_i)$$
$$\geq \sum_{i=1}^{n} I(X;\hat{X}_{2,i}|U_i), \tag{31}$$

where equality (a) follows from the definition of $U_i=(T_2,X^{i-1},Y^{i-1})$ and the fact that $\hat{X}_{2,i}$ is a deterministic function of (T_2,T_3).

The proof is concluded in the standard way by letting Q be a random variable independent of X^n,Y^n, uniformly distributed over the set $\{1,2,3,..,n\}$, and considering the joint distribution of $X_Q,Y_Q,U_Q,\hat{X}_{1,Q},\hat{X}_{2,Q}$. For this joint distribution, Inequalities (29), (30), (31) imply that (25), (26) and (27) hold, respectively, and the fact that the code we have fixed satisfies the distortion constraints implies that (28) holds.

To prove the cardinality bound of U, we invoke the support lemma [23, pp. 310]. The external random variable U must have $|\mathcal{X}||\mathcal{Y}||\hat{\mathcal{X}}_1||\hat{\mathcal{X}}_2|-1$ letters to preserve $P(x,y,\hat{x}_1,\hat{x}_2)$ plus three more to preserve the expressions $I(X;\hat{X}_1,U|Y)$, $I(Y,X;U)$, $I(X;\hat{X}_2|U)$. Note that preserving $P(x,y,\hat{x}_1,\hat{x}_2)$ implies that $\mathbb{E}d_i(X,\hat{X}_i)\leq D_i$ for $i=1,2$ is also preserved. ∎

For the achievability part, we first establish the following:

Lemma 5 (Optimality of $\hat{X}_2-(X,U)-(\hat{X}_1,Y)$): The rate region $\mathcal{R}_\triangle(D_1,D_2)$, which is defined by (25)-(28), does not decrease by restricting the joint distribution to the form $P(x,y)P(\hat{x}_1,u|x,y)P(\hat{x}_2|x,u)$.

Proof: For a fixed (D_1,D_2), let the rate-triple $(R_1,R_2,R_3)\in\mathcal{R}_\triangle(D_1,D_2)$. Then there exists a joint distribution

$$P(x,y,u,\hat{x}_1,\hat{x}_2)=P(x,y)P(\hat{x}_1,\hat{x}_2,u|x,y), \tag{32}$$

for which (25)-(28) hold. Let $P(\hat{x}_1,u|x,y)$ and $P(\hat{x}_2|x,u)$ be the conditional distribution induced by $P(x,y,u,\hat{x}_1,\hat{x}_2)$. We now claim that (25)-(28) are satisfied under the joint distribution

$$\tilde{P}(x,y,u,\hat{x}_1,\hat{x}_2)=P(x,y)P(\hat{x}_1,u|x,y)P(\hat{x}_2|x,u). \tag{33}$$

This is true, since the expressions (25)-(28) depend on $P(x,y,u,\hat{x}_1,\hat{x}_2)$ only through the marginals $P(x,y,u,\hat{x}_1)$ and $P(x,u,\hat{x}_2)$. Now notice that those marginals are the same whether the joint distribution is $P(x,y,u,\hat{x}_1,\hat{x}_2)$ or $\tilde{P}(x,y,u,\hat{x}_1,\hat{x}_2)$. ∎

Sketch of proof of Achievability part of Theorem 4: The achievability proof follows directly from the achievability of cascade source coding as given in Theorem 1. First, we fix a joint distribution of the form $P(x,y)P(\hat{x}_1,u|x,y)P(\hat{x}_2|x,u,y)$ such that (25)-(28) hold. Since $R_1>I(X;\hat{X}_1,U|Y)$ and $R_2>I(Y,X;U)$, then according to Theorem 1, we can generate $(\hat{X}_1^n,U^n)$ that with high probability would be jointly typical with (X^n,Y^n) according to the distribution $P(x,y)P(\hat{x}_1,u|x,y)$. Now, since U^n is known both to the encoder and to User 2, we need a rate $R_3>I(X;\hat{X}_2|U)$ to generate $\hat{X}_2^n$ such that with high probability it is jointly typical with X^n,U^n. Finally, because of the Markov relation $\hat{X}_2-(X,U)-(\hat{X}_1,Y)$, we can invoke the Markov lemma, and conclude that the sequences $X^n,Y^n,\hat{X}_1^n,,\hat{X}_2^n,U^n$ are jointly typical and therefore the distortion criteria are satisfied. ∎

B. The Gaussian triangular case

We now evaluate the rate region of the triangular network depicted in Fig. 5 for the quadratic Gaussian case, i.e., X, Y are jointly Gaussian and $d_1(x,\hat{x}_1) = (x-\hat{x}_1)^2$, $d_2(x,\hat{x}_2) = (x-\hat{x}_2)^2$. We first show that it suffices to consider only Gaussian joint distributions for exhausting the region, and then we show that by a small change in the Gaussian cascade region we obtain the Gaussian triangular region.

Theorem 6 (Optimality of jointly Gaussian distributions): For the quadratic Gaussian triangular rate-distortion problem with side information known to the source-encoder and to User 1, it suffices to consider only jointly Gaussian distributions $P_{X,Y,U,\hat{X}_1,\hat{X}_2}$ in order to exhaust the rate region $\mathcal{R}_\Delta(D_1,D_2)$ given in (25)-(28).

Before proving the theorem, let us introduce the Pareto frontier [24] of a region and show that if two rate-regions have the same Pareto frontier then they are identical. The *Pareto frontier* of a region $\mathcal{R}$, which we denote by $Par(\mathcal{R})$, is the set of all points for which there is no strictly better point in the region. Formally,

$$Par(\mathcal{R}) = \{R^n \in \mathcal{R} : \nexists \tilde{R}^n \in \mathcal{R} \text{ s.t. } \tilde{R}^n \prec R^n\}, \tag{34}$$

where $\tilde{R}^n \prec R^n$ denotes that $\tilde{R}_i \le R_i$ for all $1 \le i \le n$ and for some $1 \le i \le n$, $\tilde{R}_i < R_i$.

Lemma 7: If two rate-regions, $\mathcal{R}_1$ and $\mathcal{R}_2$, have the same Pareto frontier, then they are identical.

Proof: Let us show that the assumptions $R \in \mathcal{R}_1$ and $R \notin \mathcal{R}_2$ lead to a contradiction. If $R \in \mathcal{R}_1$, then there exists a point $R_p \in Par(\mathcal{R}_1)$ that satisfies $R_p \prec R$. Since $R_p \in Par(\mathcal{R}_1)$, it follows that $R_p \in Par(\mathcal{R}_2)$. Finally, since $R_p \in \mathcal{R}_2$ and $R_p \prec R$, then $R \in \mathcal{R}_2$, which contradicts the assumption. ■

Proof of Theorem 6: As a result of Lemma 7, we conclude that it suffices to prove Theorem 6 only for the points in the Pareto frontier. In addition, we notice that points that are Pareto optimal satisfy (25)-(27) with equality, which may be also written as

$$R_1 = I(X;\hat{X}_1,U|Y), \tag{35}$$

$$R_2 = I(Y,X;U), \tag{36}$$

$$R_3 + R_2 = I(Y,X;\hat{X}_2,U). \tag{37}$$

Finally, assuming without loss of generality U is real-valued and using similar arguments as in Lemma 2, we conclude that for any joint distribution $P_{X,Y,\hat{X}_1,\hat{X}_2,U}$ there exists a Gaussian joint distribution, $\tilde{P}_{X,Y,\hat{X}_1,\hat{X}_2,U}$, with the same covariance matrix as $P_{X,Y,\hat{X}_1,\hat{X}_2,U}$, for which the induced right hand sides of (35)-(37) do not increase. ■

Now, with a small change in the solution to the Gaussian cascade, we obtain the triangular Gaussian region. The proof is deferred to Appendix B.

Theorem 8 (Triangle Gaussian case): The rate region of the triangular source coding with side information at the first two nodes, where the source X and the side information $Y = X + Z$ are jointly Gaussian distributed, where X and Z are mutually independent, and the distortion is quadratic, is given by Eq. (16)-(17), where D_2 is replaced by $D_2 2^{2R_3}$ i.e., $R_1^{triangle}(D_1,D_2,R_2,R_3) = R_1^{cascade}(D_1,D_2 2^{2R_3},R_2)$.

VI. Extensions

Here we present two further extensions. The first is obtained by generalizing the triangular network results to more users. The second is obtained by considering a more general problem of empirical coordination rather than distortion criteria.

A. Multiple Users

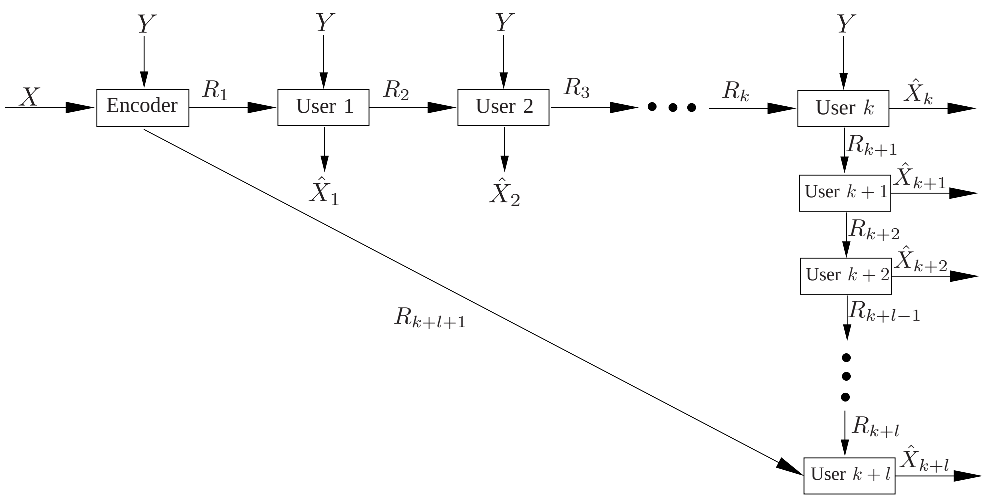

Fig. 6. A triangular rate distortion problem with $k+l$ users, where the side information Y is known to the encoder and to Users $1, 2, ..., k$, but not to Users $k+1, k+2, ..., k+l$.

The triangular problem depicted in Fig. 5 can be extended to $k+l$ users, where the side information is known to the source encoder and to Users $1, 2, ..., k$, but is not known to Users $k+1, k+2, ..., k+l$. This problem is depicted in Fig. 6, and its region is given by the next theorem.

Theorem 9: The achievable region for the problem depicted in Fig. 6 is given by the vector rates $(R_1, R_2, ..., R_{k+l+1})$ that satisfy

$$
\begin{aligned}
R_i &\geq I(X; \hat{X}_i, \hat{X}_{i+1}, ..., \hat{X}_{k+l-1}, U|Y), \;\; 1 \leq i \leq k \\
R_j &\geq I(X; \hat{X}_j, ..., \hat{X}_{k+l-1}, U), \qquad k+1 \leq j \leq k+l \\
R_{k+l+1} &\geq I(X; \hat{X}_{k+l}|U),
\end{aligned}
\tag{38}
$$

for some distribution $P(x,y)P(\hat{x}_1, \hat{x}_2, ..., \hat{x}_k, u|x,y)$ for which

$$
\mathbb{E}d_i(X, \hat{X}_i) \leq D_i, \quad 1 \leq i \leq k+l. \tag{39}
$$

where the cardinality of the auxiliary variable U may be bounded by $|U| \leq |\mathcal{X}||\mathcal{Y}||\hat{\mathcal{X}}_1||\hat{\mathcal{X}}_2|...|\hat{\mathcal{X}}_{k+l}| + k + l$. The proof of Theorem 9 follows similar steps as the proof of Theorem 4 and is therefore omitted.

B. Empirical coordination

In [25], two coordination problems were introduced: Empirical coordination, where the goal is to generate sequences with a specific empirical distribution, and strong coordination, where the goal is to generate sequences with a distribution that is close (in total variation) to a specific i.i.d. distribution. The empirical coordination problem is a generalization of the rate distortion problem, since a distortion constraint defines a half-plane in the empirical distribution space. Hence, if we find the optimal rate needed to generate a specific empirical distribution, we also find the optimal rate needed to generate a specific distortion constraint.

For the cascade rate distortion problem with side information at the first two nodes, the extension to the empirical coordination problem is straightforward.

Theorem 10 (Rate coordination in the cascade problem): The rate coordination region $R_{P_0}(P(\hat{x}_1,\hat{x}_2|x,y))$ of the cascade problem where side information is known to the first two nodes, where $X,Y \sim P_0(x,y)$, and an empirical distribution $P_0(x,y)P(\hat{x}_1,\hat{x}_2|x,y)$ is desired, is given by

$$\begin{aligned} R_2 &\geq I(Y,X;\hat{X}_2), \\ R_1 &\geq I(X;\hat{X}_1,\hat{X}_2|Y), \end{aligned} \tag{40}$$

where the joint distribution evaluating the mutual information expression is $P_0(x,y)P(\hat{x}_1,\hat{x}_2|x,y)$.

Proof: The achievability proof follows immediately from the achievability proof of Theorem 1, where we fixed an empirical distribution and showed that it can be achieved using the above rates. The converse also follows from the converse of Theorem 1, but in the last step we need to invoke [25, Proposition 2], which states that the expected empirical distribution equals the distribution of the random variables chosen uniformly over the time sequence $1,2,...,n$, i.e., $\mathbb{E}\left[P_{X^n,Y^n,\hat{X}_1^n,\hat{X}_2^n}(x,y,\hat{x}_1,\hat{x}_2)\right] = P_{X_Q,Y_Q,\hat{X}_{1,Q},\hat{X}_{2,Q}}(x,y,\hat{x}_1,\hat{x}_2)$. ■

However, the triangular coordination problem is an open problem, even without side information. The solution here is heavily based on the fact that in the achievability proof it suffices to consider only a specific empirical distribution (with a Markov structure), but for an arbitrary distribution the coordination problem remains open.

APPENDIX A
PROOF OF THEOREM 3

Following Lemma 2 we can rewrite the rate region for the Gaussian case as:

$$R_2 \geq I(Y,X;W), \tag{41}$$

$$R_1 \geq I(X;V,W|Y), \tag{42}$$

where the vector (X,Y,V,W) is jointly Gaussian distributed and satisfies

$$\sigma^2_{X|W} \leq D_2 \tag{43}$$

$$\sigma^2_{X|W,V,Y} \leq D_1, \tag{44}$$

where $\sigma^2_{A|B} \triangleq E[(A-E[A|B])^2]$.

Without loss of generality let us choose the following structure

$$\begin{aligned} Y &= X + Z, \\ W &= X + \alpha Y + Z_2 = (1+\alpha)X + \alpha Z + Z_2, \\ V &= X + \beta Y + \gamma Z_2 + Z_1, \end{aligned} \tag{45}$$

where the random variables X, Z, Z_1, Z_2 are jointly Gaussian and mutually independent, with variances $\sigma_X^2, \sigma_Z^2, \sigma_{Z_1}^2, \sigma_{Z_2}^2$, respectively, and the coefficients (α, β, γ) are real number scalars.

Equations (42)-(44) become

$$\begin{aligned} R_2 &\geq I(X,Y;W) \\ &= H(W) - H(W|X,Y) \\ &= \frac{1}{2}\log\frac{(1+\alpha)^2\sigma_X^2 + \alpha^2\sigma_Z^2 + \sigma_{Z_2}^2}{\sigma_{Z_2}^2} \end{aligned} \tag{46}$$

$$D_2 \geq \sigma_{X|W}^2 = \frac{\sigma_X^2(\alpha^2\sigma_Z^2 + \sigma_{Z_2}^2)}{(1+\alpha)^2\sigma_X^2 + \alpha^2\sigma_Z^2 + \sigma_{Z_2}^2} \tag{47}$$

$$R_1 = \frac{1}{2}\max\left(\log\frac{\sigma_{X|Y}^2}{\sigma_{X|W,Y}^2}, \log\frac{\sigma_{X|Y}^2}{D_1}\right), \tag{48}$$

where $\sigma_{X|Y}^2 = \frac{\sigma_X^2\sigma_Z^2}{\sigma_X^2+\sigma_Z^2}$ and $\sigma_{X|W,Y}^{-2} = \sigma_{Z_2}^{-2} + \sigma_X^{-2} + \sigma_Z^{-2}$.

Inequalities (46) and (47) follow directly from (41) and (43), respectively. Eq. (48) follows from combining the following two equations, (49)- (50). If $D_1 \geq \sigma_{X|W,Y}^2$, then (44) is automatically satisfied, and then V is not needed (may be independent of anything else) and therefore

$$\begin{aligned} R_1 &\geq I(X;W|Y) \\ &= H(X|Y) - H(X|Y,W) \\ &= H(X|Y) - H(X|Y,W) \\ &= \frac{1}{2}\log\frac{\sigma_{X|Y}^2}{\sigma_{X|W,Y}^2}. \end{aligned} \tag{49}$$

If $D_1 \leq \sigma_{X|W,Y}^2$, then

$$\begin{aligned} R_1 &\geq I(X;V,W|Y) \\ &= H(X|Y) - H(X|Y,V,W) \\ &= \frac{1}{2}\log\frac{\sigma_{X|Y}^2}{D_1}. \end{aligned} \tag{50}$$

The last equality is due to the fact that we can choose (β, γ, Z_1) such that $\sigma_{X|W,V,Y}^2 = D_1$.

Now let us fix $D_1 \geq 0$, $D_2 \geq 0$, and $R_2 \geq \frac{1}{2}\log\frac{\sigma_X^2}{D_2}$, and let us find the function $R_1(D_1, D_2, R_2)$, which defines the rate region. (The condition on R_2 is due to the fact that if $R_2 < \frac{1}{2}\log\frac{\sigma_X^2}{D_2}$ the rate will not be achievable for any R_1). To find R_1 we need to solve the following optimization problem

$$\text{maximize} \quad \sigma_{Z_2}^2 \tag{51}$$

$$\text{subject to} \quad (2^{2R_2}-1)\sigma_{Z_2}^2 \geq (1+\alpha)^2\sigma_X^2 + \alpha^2\sigma_Z^2 \tag{52}$$

$$\sigma_{Z_2}^2(\sigma_X^2 - D_2) \leq \alpha^2(\sigma_X^2 D_2 + \sigma_Z^2 D_2 - \sigma_X^2\sigma_Z^2) + 2\alpha\sigma_X^2 D_2 + D_2\sigma_X^2 \tag{53}$$

The objective (51) follows from the fact that R_1 depends only on $\sigma_{Z_2}^2$ and (52) and (53) follow from (46) and (47), respectively. To solve this optimization problem, we divide the problem into four cases, where each case has a simple solution (each case corresponds to a line in (17)).

Case 1: For this case we assume that

$$\sigma_X^2 D_2 + \sigma_Z^2 D_2 - \sigma_X^2\sigma_Z^2 < 0 \Rightarrow D_2 \leq \frac{\sigma_Z^2\sigma_X^2}{\sigma_Z^2 + \sigma_X^2} = \sigma_{X|Y}^2, \tag{54}$$

and

$$R_2 \quad \geq \quad \frac{1}{2}\log \frac{\sigma_Z^2(\sigma_X^2 - D_2)}{\sigma_Z^2\sigma_X^2 - D_2\sigma_Z^2 - D_2\sigma_X^2}\frac{\sigma_X^2}{D_2}. \tag{55}$$

Because of the assumption in (73), Eq. (53) holds with equality, since otherwise $\sigma_{Z_2}^2$ can be increased until it hits the boundary of (53).

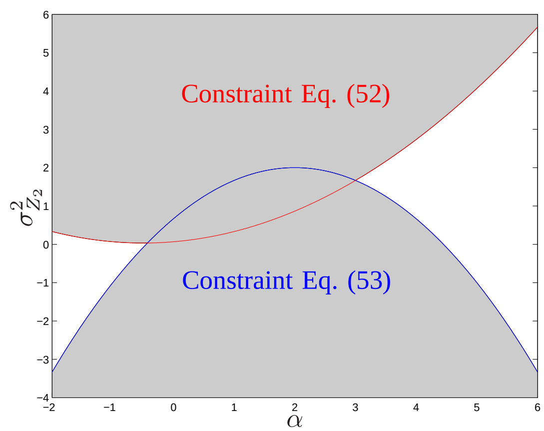

Fig. 7. Case 1: the maximum of $\sigma_{Z_2}^2$, where both constraints hold, is obtained at the maximum of Eq. (53).

The argument that achieves the maximum of a quadratic form $a\alpha^2 + g\alpha + c$ is $\frac{-b}{2a}$, hence the argument that maximizes (53) is

$$\overline{\alpha} = \frac{-\sigma_X^2 D_2}{\sigma_X^2 D_2 + \sigma_Z^2 D_2 - \sigma_X^2\sigma_Z^2}, \tag{56}$$

and the maximum is

$$\begin{aligned} \overline{\sigma}_{Z_2}^2 &= c - \frac{b^2}{4a} \\ &= \frac{\sigma_x^2 D_2}{\sigma_Z^2 D_2 - \sigma_Z^2\sigma_X^2}(\sigma_X^2 - D_2)(\sigma_X^2 D_2 + \sigma_Z^2 D_2 - \sigma_Z^2\sigma_X^2) \end{aligned}$$

$$= \quad \overline{\alpha}\sigma_Z^2. \tag{57}$$

Note that (57) can be also written as

$$\frac{1}{\overline{\sigma}_{Z_2}^2} = \frac{1}{D_2} - \frac{1}{\sigma_Z^2} - \frac{1}{\sigma_X^2}. \tag{58}$$

If $(\overline{\alpha}, \overline{\sigma}_{Z_2}^2)$ satisfy Eq. (52), then the solution to the optimization problem is simply $\overline{\sigma}_{Z_2}^2$ and using (48) we obtain

$$R_1 = \frac{1}{2}\max\left(\log\frac{\sigma_{X|Y}^2}{D_2}, \log\frac{\sigma_{X|Y}^2}{D_1}\right). \tag{59}$$

Now let us investigate when $(\overline{\alpha}, \overline{\sigma}_{Z_2}^2)$ satisfies Eq. (52) (or equivalently (46))

$$\begin{aligned} R_2 &\geq \frac{1}{2}\log\frac{(1+\overline{\alpha})^2\sigma_X^2 + \overline{\alpha}^2\sigma_Z^2 + \overline{\sigma}_{Z_2}^2}{\overline{\sigma}_{Z_2}^2} \\ &\overset{(a)}{=} \frac{1}{2}\log\frac{\sigma_X^2(\overline{\alpha}^2\sigma_Z^2 + \overline{\sigma}_{Z_2}^2)}{\overline{\sigma}_{Z_2}^2 D_2} \\ &\overset{(b)}{=} \frac{1}{2}\log\frac{\sigma_X^2(\overline{\alpha}^2\sigma_Z^2 + \overline{\alpha}\sigma_Z^2)}{\overline{\alpha}\sigma_Z^2 D_2} \\ &\overset{(c)}{=} \frac{1}{2}\log\frac{\sigma_Z^2(\sigma_X^2 - D_2)}{\sigma_Z^2\sigma_X^2 - D_2\sigma_Z^2 - D_2\sigma_X^2}\frac{\sigma_X^2}{D_2}, \end{aligned} \tag{60}$$

where (a) follows from Equality (47), (b) from (57) and (c) from (56).

Case 2: Assume that

$$D_2 \leq \frac{\sigma_Z^2\sigma_X^2}{\sigma_Z^2 + \sigma_X^2} = \sigma_{X|Y}^2, \tag{61}$$

and

$$R_2 \leq \frac{1}{2}\log\frac{\sigma_Z^2(\sigma_X^2 - D_2)}{\sigma_Z^2\sigma_X^2 - D_2\sigma_Z^2 - D_2\sigma_X^2}\frac{\sigma_X^2}{D_2}. \tag{62}$$

Now if (60) is not satisfied, then the maximum of $\sigma_{Z_2}^2$ should be on the boundary of the constraints, namely, both (52) and (53) should hold with equality. This is because the upper part of the intersection should be either increasing or decreasing. Such a case is shown in Fig. 8.

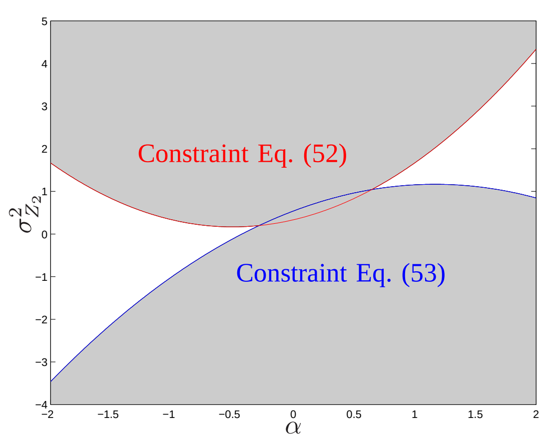

Fig. 8. Case 2: the maximum of $\sigma_{Z_2}^2$, where both constraints hold, is obtained at the intersection of (52) and (53).

Consider the case where (46) and (47) hold with equality. Then we obtain

$$2^{2R_2}\sigma_{Z_2}^2 = \frac{\sigma_X^2(\alpha^2\sigma_Z^2+\sigma_{Z_2}^2)}{D_2}, \tag{63}$$

which implies

$$\sigma_{Z_2}^2 = \frac{\sigma_Z^2\sigma_X^2}{2^{2R_2}D_2-\sigma_X^2}\alpha^2. \tag{64}$$

Now substituting $\sigma_{Z_2}^2$ given by (64) into (52) we obtain

$$\frac{\alpha^2\sigma_Z^2\sigma_X^2(2^{2R_2}-1)}{2^{2R_2}D_2-\sigma_X^2} = (1+\alpha)^2\sigma_X^2+\alpha^2\sigma_Z^2, \tag{65}$$

which simplifies to

$$\frac{\alpha^2\sigma_Z^2(\sigma_X^2-D_2)}{D_2-\sigma_X^2 2^{-2R_2}} = (1+\alpha)^2\sigma_X^2. \tag{66}$$

Taking the square-root on each side of the equation we obtain two possible solutions for α:

$$\frac{1}{\alpha} = \pm\frac{\sigma_Z}{\sigma_X}\sqrt{\frac{\sigma_X^2-D_2}{D_2-\sigma_X^2 2^{-2R_2}}}-1. \tag{67}$$

Since we need to maximize $\sigma_{Z_2}^2$, which is proportional to α^2 (see Eq. (64)), we choose the solution with the plus sign.

Case 3: Assume that

$$D_2 \geq \frac{\sigma_Z^2\sigma_X^2}{\sigma_Z^2+\sigma_X^2} = \sigma_{X|Y}^2, \tag{68}$$

and

$$R_2 \quad \geq \quad \frac{1}{2}\log\frac{\sigma_Z^2(\sigma_X^2-D_2)}{\sigma_Z^2\sigma_X^2-D_2\sigma_Z^2-D_2\sigma_X^2}\frac{\sigma_X^2}{D_2}. \tag{69}$$

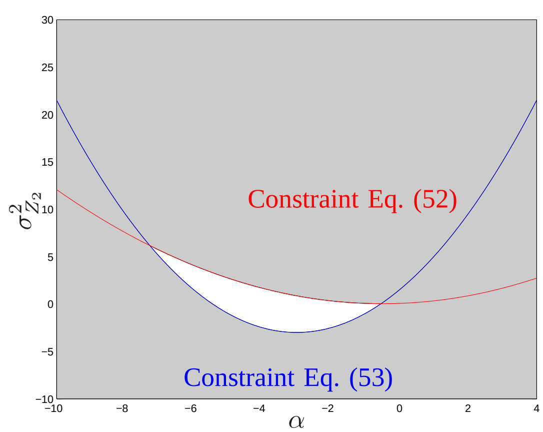

Fig. 9. Case 3: the maximum of $\sigma_{Z_2}^2$, where both constraints hold, is obtained at infinity, since there is a infinite overlap between the constraints.

If

$$\frac{(\sigma_X^2D_2+\sigma_Z^2D_2-\sigma_X^2\sigma_Z^2)}{\sigma_X^2-D_2} \geq \frac{\sigma_X^2+\sigma_Z^2}{2^{2R_2}-1}, \tag{70}$$

which is equivalent to

$$2^{2R_2} \geq \frac{\sigma_X^4}{\sigma_X^2 D_2 + \sigma_Z^2 D_2 - \sigma_X^2 \sigma_Z^2}, \tag{71}$$

then the maximum of $\sigma_{Z_2}^2$ is obtained at infinity (as illustrated in Fig. 9), which implies that

$$R_1 = \frac{1}{2} \max\left(0, \log \frac{\sigma_{X|Y}^2}{D_1}\right) = \frac{1}{2} \log \frac{\sigma_{X|Y}^2}{D_1}. \tag{72}$$

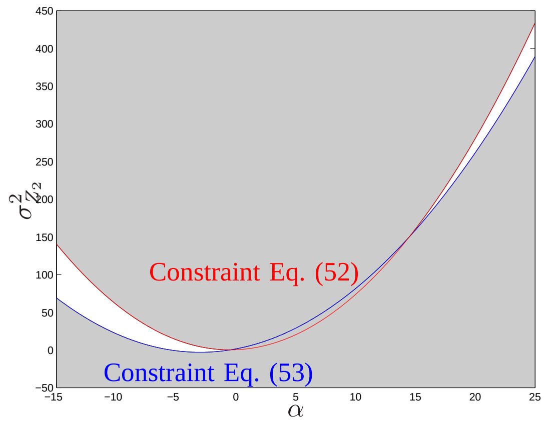

Fig. 10. Case 4: the maximum of $\sigma_{Z_2}^2$, where both constraints hold, is obtained at the intersection of (52) and (53).

Case 4: Assume that

$$D_2 \geq \frac{\sigma_Z^2 \sigma_X^2}{\sigma_Z^2 + \sigma_X^2} = \sigma_{X|Y}^2, \tag{73}$$

and

$$R_2 \leq \frac{1}{2} \log \frac{\sigma_Z^2(\sigma_X^2 - D_2)}{\sigma_Z^2 \sigma_X^2 - D_2 \sigma_Z^2 - D_2 \sigma_X^2} \frac{\sigma_X^2}{D_2}. \tag{74}$$

If (71) does not hold, then the maximum of $\sigma_{Z_2}^2$ should be at boundary of the constraint, namely, (52) and (53) should hold with equality. This is because the upper part of the intersection should be either increasing or decreasing. Such a case is shown in Fig. 10. ■

APPENDIX B
PROOF OF THEOREM 6

Let us rewrite the rate region equations similarly to (42)-(44) as,

$$R_1 \geq I(X; V, W|Y), \tag{75}$$

$$R_2 \geq I(Y, X; W), \tag{76}$$

$$R_3 \geq I(X; W'|W), \tag{77}$$

where the vector (X, Y, V, W) is jointly Gaussian distributed and satisfies

$$\sigma_{X|W,W'}^2 \leq D_2 \tag{78}$$

$$\sigma^2_{X|W,V,Y} \quad \leq \quad D_1, \tag{79}$$

Without loss of generality, we may assume that X, Y, W, V have the same structure as in (45) and $W' = X + \eta W + Z'$ where $Z' \sim N(0, \sigma^2_{Z'}$ is independent of X, Y, W, V. Furthermore, we note that we can assume that (77) holds with equality, since if not, we can change η and Z' such that equality will hold, and the change will only decrease $\sigma^2_{X|W,W'}$ - therefore (75)-(79) will continue to hold. Now, the equality in (77) implies that

$$\sigma^2_{X|W,W'} = \sigma^2_{X|W} 2^{-2R_3}. \tag{80}$$

Hence (78) becomes

$$\sigma^2_{X|W} \leq D_2 2^{2R_3}. \tag{81}$$

Now we note that we obtain the same optimization problem as in (46)-(48), just that D_2 is replaced by $D_2 2^{2R_3}$.■

References

[1] H. Yamamoto. Source coding theory for cascade and branching communication systems. *IEEE Trans. Inf. Theory*, 27:299–308, 1981.

[2] D. Vasudevan, C. Tian, and S N. Diggavi. Lossy source coding for a cascade communication system with side-informations. In *Proceeding of Allerton conference on Communication, Control and Computing*, 2006.

[3] H. Su P. Cuff, A. El Gamal. Cascade multiterminal source coding. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2009.

[4] H. Permuter, Y. Steinberg, and T. Weissman. Problems we can solve with a helper,. In *Proc. IEEE Information Theory Workshop (ITW), Volos*, June, 2009.

[5] A.D. Wyner. The rate-distortion function for source coding with side information at the decoder-II: General sources. *Information and Control*, 38:60–80, 1978.

[6] C. Heegard and T. Berger. Rate distortion when side information may be absent. *IEEE Trans. Inf. Theory*, IT-31:724–734, 1985.

[7] A. Kaspi. Rate-distortion function when side-information may be present at the decoder. *IEEE Trans. Inf. Theory*, IT-40:2031–2034, 1994.

[8] S. Diggavi E. Perron and E. Telatar. The Kaspi Rate-Distortion Problem with Encoder Side-Information: Gaussian Case. Technical report, 2005.

[9] A. El Gamal and T. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inf. Theory*, 28:851–867, 1982.

[10] L. Ozarow. On a source coding problem with two channels and three receivers. *IEEE Trans. Inf. Theory*, IT-59:1909–1921, 1980.

[11] A. H. Kaspi. Two-way source coding with a fidelity criterion. *IEEE Trans. Inf. Theory*, 31(6):735–740, 1985.

[12] Y. Oohama. Gaussian multiterminal source coding. *IEEE Trans. Inf. Theory*, 43:1912–1923, 1997.

[13] S. Tavildar A. B. Wagner and P. Viswanath. Rate region of the quadratic Gaussian two-encoder source-coding problem. *IEEE Trans. Inf. Theory*, IT-54:1938–1961, 2008.

[14] Z. Zhang T. Berger and H. Viswanathan. The CEO problem. *IEEE Trans. Inf. Theory*, IT-42:887902, 1996.

[15] H. Viswanathan and T. Berger. The quadratic Gaussian CEO problem. *IEEE Trans. Inf. Theory*, IT-43:15491559, 1997.

[16] Y. Oohama. The rate-distortion function for the quadratic Gaussian CEO problem. *IEEE Trans. Inf. Theory*, IT-44:10571070, 98.

[17] D. Vasudevan and E. Perron. Cooperative source coding with encoder breakdown. In *Proc. International Symposium on Information Theory (ISIT)*, Nice, France., June, 2007.

[18] H. Permuter, Y. Steinberg, and T. Weissman. Rate-distortion with a limited-rate helper to the encoder and decoder,. Availble at http://arxiv.org/abs/0811.4773v1, Nov. 2008.

[19] W. H. R. Equitz and T. M. Cover. Successive refinement of information. *IEEE Trans. Inf. Theory*, 37:269–275, 1991.

[20] Y. Steinberg and N. Merhav. On successive refinement for the wyner-ziv problem. *IEEE Trans. Inf. Theory*, IT-50:1636–1654, 2004.

[21] H. Yamamoto. Source coding theory for a triangular communication system. *IEEE Trans. Inf. Theory*, 42(3):848–853, 1996.

[22] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New-York, 2nd edition, 2006.

[23] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic, New York, 1981.

[24] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New-York, 2004.

[25] P. Cuff, H. H. Permuter, and T. M. Cover. Coordination capacity. submitted to *IEEE Trans. Inf. Theory.* Available at arxiv.org/abs/0909.2408v1, Sep. 2009.