Interaction Strictly Improves the Wyner-Ziv Rate-distortion Function[1]

Nan Ma
ECE Dept, Boston University
Boston, MA 02215
nanma@bu.edu

Prakash Ishwar
ECE Dept, Boston University
Boston, MA 02215
pi@bu.edu

***Abstract*—In 1985 Kaspi provided a single-letter characterization of the sum-rate-distortion function for a two-way lossy source coding problem in which two terminals send multiple messages back and forth with the goal of reproducing each other's sources. Yet, the question remained whether more messages can strictly improve the sum-rate-distortion function. Viewing the sum-rate as a functional of the distortions and the joint source distribution and leveraging its convex-geometric properties, we construct an example which shows that two messages can strictly improve the one-message (Wyner-Ziv) rate-distortion function. The example also shows that the ratio of the one-message rate to the two-message sum-rate can be arbitrarily large and simultaneously the ratio of the backward rate to the forward rate in the two-message sum-rate can be arbitrarily small.**

I. Introduction

Consider the following two-way lossy source coding problem studied in [1]. Let $(X(1), Y(1)), \ldots, (X(n), Y(n))$ be n iid samples of a two-component discrete memoryless stationary source with joint pmf $p_{XY}(x, y)$, $(x, y) \in \mathcal{X} \times \mathcal{Y}$, $|\mathcal{X} \times \mathcal{Y}| < \infty$. Terminal A observes $\mathbf{X} := (X(1), \ldots, X(n))$ and terminal B observes $\mathbf{Y} := (Y(1), \ldots, Y(n))$. Terminal B is required to produce $\widehat{\mathbf{X}} := (\widehat{X}(1), \ldots, \widehat{X}(n)) \in \widehat{\mathcal{X}}^n$, where $\widehat{\mathcal{X}}$ is a reproduction alphabet with $|\widehat{\mathcal{X}}| < \infty$, such that the expected distortion $\mathbb{E}[d^{(n)}(\mathbf{X}, \widehat{\mathbf{X}})]$ does not exceed a desired level, where

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

and $d : \mathcal{X} \times \widehat{\mathcal{X}} \rightarrow \mathbb{R}^+ \bigcup \{\infty\}$ is a per-sample (single-letter) distortion function. Terminal A is likewise required to reproduce the source observed at terminal B within some distortion level with respect to another (possibly different) distortion function. To achieve this objective, the terminals are allowed to send a certain number of messages back and forth where each message sent from a terminal at any time only depends on the information available at the terminal up to that time. In [1], Kaspi provided a single-letter characterization of the sum-rate-distortion function for any finite number of messages. Yet, whether more messages can strictly improve the sum-rate-distortion function was left unresolved. If the goal is to reproduce both sources *losslessly* at each terminal (zero distortion) then there is no advantage in using multiple messages; two messages are sufficient and the minimum sum-rate cannot be reduced by using more than two messages.[2] If, however, the goal is changed to losslessly *compute functions* of sources at each terminal, then multiple messages can decrease the minimum sum-rate by an arbitrarily large factor [3], [4]. Therefore, the key unresolved question pertains to *lossy source reproduction*: can multiple messages strictly decrease the minimum sum-rate for a given (nonzero) distortion? This question is unresolved even when only one source needs to be reproduced with nonzero distortion.

In this paper, we construct the first example which shows that two messages can strictly improve the one-message (Wyner-Ziv) rate-distortion function. The example also shows that the ratio of the one-message rate to the two-message sum-rate can be arbitrarily large and simultaneously the ratio of the backward rate to the forward rate in the two-message sum-rate can be arbitrarily small. The key idea which enables the construction of this example is that the sum-rate is a *functional* of the distortion and the joint source distribution which has certain convex-geometric properties.

II. Problem setup and related prior results

A. One-message Wyner-Ziv rate-distortion function

Definition 1: A one-message distributed source code with parameters $(n, |\mathcal{M}|)$ is the tuple $(e^{(n)}, g^{(n)})$ consisting of an encoding function $e^{(n)} : \mathcal{X}^n \rightarrow \mathcal{M}$ and a decoding function $g^{(n)} : \mathcal{Y}^n \times \mathcal{M} \rightarrow \widehat{\mathcal{X}}^n$. The output of $g^{(n)}$, denoted by $\widehat{\mathbf{X}}$, is called the reproduction and $(1/n)\log_2 |\mathcal{M}|$ is called the block-coding rate (in bits per sample).

Definition 2: A tuple (R, D) is admissible for one-message distributed source coding if, $\forall \epsilon > 0$, $\exists\ \overline{n}(\epsilon)$ such that $\forall n > \overline{n}(\epsilon)$, there exists a one-message distributed source code with parameters $(n, |\mathcal{M}|)$ satisfying $\frac{1}{n}\log_2 |\mathcal{M}| \leq R + \epsilon$, and $\mathbb{E}[d^{(n)}(\mathbf{X}, \widehat{\mathbf{X}})] \leq D + \epsilon$.

The set of all admissible (R, D) tuples in Definition 2 is a closed subset of $\mathbb{R}^2$. For any $D \in \mathbb{R}$, the minimum value of R

[1]This material is based upon work supported by the US National Science Foundation (NSF) under award (CAREER) CCF–0546598 and CCF–0915389. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

[2]If only one of the sources is required to be losslessly reproduced at the other terminal then one message is sufficient and the minimum sum-rate cannot be improved by using more than one message. However, if $\mathbf{X}$ and $\mathbf{Y}$ are nonergodic, two-way interactive coding can be strictly better than one-way non-interactive coding [2].

such that (R, D) is admissible is the one-message Wyner-Ziv rate-distortion function [5] and will be denoted by $R_{sum,1}(D)$. The following single-letter characterization of $R_{sum,1}(D)$ was established in [5]:

$$R_{sum,1}(D) = \min_{p_{U|X}, g:\ \mathbb{E}[d(X, g(U,Y))] \leq D} I(X; U|Y), \tag{2.1}$$

where $U \in \mathcal{U}$ is an auxiliary random variable such that $U - X - Y$ is a Markov chain and $|\mathcal{U}| \leq |\mathcal{X}| + 1$, and $g : \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ is a deterministic single-letter decoding function.

B. Two-message sum-rate-distortion function

Definition 3: A two-message distributed source code with parameters $(n, |\mathcal{M}_1|, |\mathcal{M}_2|)$ is the tuple $(e_1^{(n)}, e_2^{(n)}, g^{(n)})$ consisting of encoding functions $e_1^{(n)} : \mathcal{Y}^n \to \mathcal{M}_1$, $e_2^{(n)} : \mathcal{X}^n \times \mathcal{M}_1 \to \mathcal{M}_2$ and a decoding function $g^{(n)} : \mathcal{Y}^n \times \mathcal{M}_1 \times \mathcal{M}_2 \to \widehat{\mathcal{X}}^n$. The output of $g^{(n)}$, denoted by $\widehat{\mathbf{X}}$, is called the reproduction and for $i = 1, 2$, $(1/n) \log_2 |\mathcal{M}_i|$ is called the i-th block-coding rate.

Definition 4: A tuple (R_1, R_2, D) is admissible for two-message distributed source coding if, $\forall \epsilon > 0$, $\exists\ \overline{n}(\epsilon)$ such that $\forall n > \overline{n}(\epsilon)$, there exists a two-message distributed source code with parameters $(n, |\mathcal{M}_1|, |\mathcal{M}_2|)$ satisfying $\frac{1}{n} \log_2 |\mathcal{M}_i| \leq R_i + \epsilon$, for $i = 1, 2$, and $\mathbb{E}[d^{(n)}(\mathbf{X}, \widehat{\mathbf{X}})] \leq D + \epsilon$.

The rate-distortion region, denoted by $\mathcal{RD}$, is defined as the set of all admissible (R_1, R_2, D) tuples and is a closed subset of $\mathbb{R}^3$. For any $D \in \mathbb{R}$, the minimum value of $(R_1 + R_2)$ such that $(R_1, R_2, D) \in \mathcal{RD}$ is the two-message sum-rate-distortion function and will be denoted by $R_{sum,2}(D)$. The following single-letter characterization of $\mathcal{RD}$ was established in [1]:

$$\begin{aligned} \mathcal{RD} = \ \{ \ & (R_1, R_2, D) \mid \exists\ p_{V_1|Y}, p_{V_2|XV_1}, g, s.t. \\ & R_1 \geq I(Y; V_1|X), \\ & R_2 \geq I(X; V_2|Y, V_1), \\ & \mathbb{E}[d(X, g(V_1, V_2, Y))] \leq D \ \}, \end{aligned} \tag{2.2}$$

where $V_1 \in \mathcal{V}_1$ and $V_2 \in \mathcal{V}_2$ are auxiliary random variables with bounded alphabets,[3] such that the Markov chains $V_1 - Y - X$ and $V_2 - (X, V_1) - Y$ hold, and $g : \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{Y} \to \widehat{\mathcal{X}}$ is a deterministic single-letter decoding function. From (2.2), it follows that

$$R_{sum,2}(D) = \min_{\substack{p_{V_1|Y}, p_{V_2|XV_1}, g: \\ \mathbb{E}[d(X, g(V_1, V_2, Y))] \leq D}} \{I(Y; V_1|X) + I(X; V_2|Y, V_1)\}. \tag{2.3}$$

Since a one-message code is a special case of a two-message code with $|\mathcal{M}_1| = 1$, the inequality $R_{sum,2}(D) \leq R_{sum,1}(D)$ holds for all $D \in \mathbb{R}$. Even though the single-letter characterizations of $R_{sum,1}(D)$ and $R_{sum,2}(D)$ are known, it has proved difficult to demonstrate the existence of p_{XY}, d, and D such that $R_{sum,2}(D) < R_{sum,1}(D)$. In the distributed source coding literature, to the best of our knowledge, there is neither an explicit example which shows that $R_{sum,2}(D) < R_{sum,1}(D)$ nor an implicit proof that such an example must exist nor a proof that there is no such example. In this paper we will construct an explicit example for which $R_{sum,2}(D) < R_{sum,1}(D)$.

[3]Bounds for the cardinalities of $\mathcal{V}_1$ and $\mathcal{V}_2$ can be found in [1].

In [6], [7], for a general $t \in \mathbb{Z}^+$, we established a connection between the t-message sum-rate-distortion function and the $(t-1)$-message sum-rate-distortion function using the rate reduction functional defined in the next subsection. This connection and the properties of the rate reduction functional allows one to compare $R_{sum,2}(D)$ and $R_{sum,1}(D)$ without having to explicitly solve the optimization problem in (2.3).

C. Key tool: rate reduction functionals

Generally speaking, for $i = 1, 2$, $R_{sum,i}$ depends on (p_{XY}, d, D). As in [6], [7], we fix d and view $R_{sum,i}$ as a functional of (p_{XY}, D). The sum-rate needed to reproduce only terminal A's source at terminal B with nonzero distortion can only be smaller than the sum-rate needed to losslessly reproduce both sources at both terminals which is equal to $H(X|Y) + H(Y|X)$. The reduction in the rate for lossy source reproduction in comparison to lossless source reproduction of both sources at both terminals is the rate-reduction functional. Specifically, the rate reduction functionals [7] are defined as follows. For $i = 1, 2$,

$$\rho_i(p_{XY}, D) := H(X|Y) + H(Y|X) - R_{sum,i}(p_{XY}, D). \tag{2.4}$$

Since $R_{sum,1} \geq R_{sum,2}$ and $\rho_1 \leq \rho_2$ always hold, $R_{sum,1}(p_{XY}, D) > R_{sum,2}(p_{XY}, D)$ if, and only if, $\rho_1(p_{XY}, D) < \rho_2(p_{XY}, D)$, i.e., if, and only if, $\rho_1(p_{XY}, D) \neq \rho_2(p_{XY}, D)$. The following key lemma provides a means for testing whether or not $\rho_1 = \rho_2$ without ever having to evaluate or work with ρ_2, i.e., without explicitly constructing auxiliary variables V_1, V_2 and the decoding function g in (2.3).

Lemma 1: The following two conditions are equivalent: (1) For all p_{XY} and D, $\rho_1(p_{XY}, D) = \rho_2(p_{XY}, D)$. (2) For all $p_{X|Y}$, $\rho_1(p_{X|Y} p_Y, D)$ is concave with respect to (p_Y, D).

In simple terms, $\rho_1 = \rho_2$ if, and only if, ρ_1 is concave under Y-marginal and distortion perturbations. The proof of Lemma 1 is along the lines of the proof of part (i) of Theorem 2 in [7] and is omitted. In fact, it can be proved that if for some $t \in \mathbb{Z}^+$, the t-message rate-reduction functional is identically equal to the $(t+1)$-message rate reduction functional, i.e., $\rho_t = \rho_{t+1}$, then $\rho_t = \rho_\infty$, the infinite-message rate-reduction functional. As discussed in [7, Remark 6], Lemma 1 does not hold if all the rate reduction functionals are replaced by the sum-rate-distortion functionals. Therefore the rate reduction functional is the key to the connection between a one-message distributed source coding scheme and a two-message distributed source coding scheme.

The remainder of this paper is organized as follows. In Theorem 1, we will use Lemma 1 to show that there exist p_{XY}, d, and D for which $R_{sum,1}(p_{XY}, D) > R_{sum,2}(p_{XY}, D)$. We will do this by (i) choosing $p_{X|Y}$ so that X and Y are symmetrically correlated binary random variables with $\mathbb{P}(Y \neq X) = p$, (ii) taking $d(x, \hat{x})$ to be the binary erasure distortion function, (iii) selecting a value for D, and (iv) showing that $\rho_1(p_{X|Y} p_Y, D)$ is not concave with respect to p_Y. By Lemma 1, this would imply that $\rho_1(p_{XY}, D) \neq \rho_2(p_{XY}, D)$ which, in turn, would imply that $R_{sum,1}(p_{XY}, D) > R_{sum,2}(p_{XY}, D)$. In Theorem 2 we will show that for certain values of parameters p and D, the two-message sum-rate can be split in such a way that the ratio R_1/R_2 is arbitrarily small and simultaneously the ratio $R_{sum,1}/(R_1+R_2)$ is arbitrarily large. This will be proved by explicitly constructing auxiliary variables V_1, V_2 and decoding function g in (2.3). While the explicit construction of V_1, V_2 and g in the proof of Theorem 2 may make the implicit proof of Theorem 1 seem redundant, it is unclear how the explicit construction can be generalized to other families of source distributions and distortion functions. The approach followed in the proof of Theorem 1, on the other hand, provides an efficient method to *test* whether the best two-message scheme can strictly outperform the best one-message scheme for *more general* distributed source coding and function computation problems. The implicit proof naturally points to an explicit construction and was, in fact, the path taken by the authors to arrive at the explicit construction.

III. Main results

Theorem 1: There exists a distortion function d, a joint distribution p_{XY}, and a distortion level D for which $R_{sum,1}(p_{XY}, D) > R_{sum,2}(p_{XY}, D)$.

Proof: In the light of the discussion in Section II-C, to prove Theorem 1, it is sufficient to show there exist $p_{X|Y}$, d, and D for which $\rho_1(p_{X|Y}p_Y, D)$ is not concave with respect to p_Y. In particular, it is sufficient to show that there exist $p_{Y,1}$ and $p_{Y,2}$ such that

$$\rho_1\left(p_{X|Y}\frac{p_{Y,1}+p_{Y,2}}{2}, D\right) < \frac{\rho_1(p_{X|Y}p_{Y,1}, D) + \rho_1(p_{X|Y}p_{Y,2}, D)}{2}. \tag{3.5}$$

Let $\mathcal{X} = \mathcal{Y} = \{0,1\}$, and $\widehat{\mathcal{X}} = \{0,1,e\}$. Let d be the binary erasure distortion function, i.e., $d: \{0,1\}\times\{0,e,1\} \to \{0,1,\infty\}$ and for $i = 0,1$, $d(i,i) = 0$, $d(i,1-i) = \infty$, and $d(i,e) = 1$. Let $p_{Y,1}(1) = 1 - p_{Y,1}(0) = p_{Y,2}(0) = 1 - p_{Y,2}(1) = q$, i.e., $p_{Y,1} = \text{Bernoulli}(q)$ and $p_{Y,2} = \text{Bernoulli}(\overline{q})$.[4] Let $p_{X|Y}$ be the conditional pmf of the binary symmetric channel with crossover probability p, i.e., $p_{X|Y}(1|0) = p_{X|Y}(0|1) = p$. Let $p_Y := (p_{Y,1}+p_{Y,2})/2$ which is Bernoulli$(1/2)$. The joint distribution $p_{XY} = p_Y p_{X|Y}$ is the joint pmf of a pair of doubly symmetric binary sources (DSBS) with parameter p, i.e., if p_{xy} denotes $p_{XY}(x,y)$, then $p_{00} = p_{11} = \overline{p}/2$ and $p_{01} = p_{10} = p/2$. For these choices of $p_{X|Y}$, $p_{Y,1}$, $p_{Y,2}$, p_Y, and d, we will analyze the left and right sides of (3.5) step by step through a sequence of definitions and propositions and establish the strict inequality for a suitable choice of D. The proofs of all the propositions are given in Section IV.

• *Left-side of (3.5):* From (2.1) and (2.4) we have

$$\rho_1(p_{XY}, D) = \max_{p_{U|X}, g:\ \mathbb{E}[d(X,g(U,Y))]\le D} \{H(X|Y,U) + H(Y|X)\}. \tag{3.6}$$

For the binary erasure distortion and a full support joint source pmf taking values in binary alphabets, (3.6) simplifies to the expression given in Proposition 1.

Proposition 1: If $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $\text{supp}(p_{XY}) = \{0,1\}^2$, d is the binary erasure distortion, and $D \in \mathbb{R}$, then $\rho_1 = \max_{p_{U|X}}(H(X|Y,U) + H(Y|X))$, where $\mathcal{U} = \{0,e,1\}$ and

$$p_{U|X}(u|x) = \begin{cases} \alpha_{0e}, & \text{if } x=0, u=e,\\ 1-\alpha_{0e}, & \text{if } x=0, u=0,\\ \alpha_{1e}, & \text{if } x=1, u=e,\\ 1-\alpha_{1e}, & \text{if } x=1, u=1,\\ 0, & \text{otherwise,}\end{cases} \tag{3.7}$$

where $\alpha_{0e}, \alpha_{1e} \in [0,1]$ satisfy $\mathbb{E}[d(X,U)] = p_X(0)\alpha_{0e} + p_X(1)\alpha_{1e} \le D$.

The expression for ρ_1 further simplifies to the one in Proposition 2 by using $p_{U|X}$ given by (3.7) in (3.6).

Proposition 2: If $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $\text{supp}(p_{XY}) = \{0,1\}^2$, d is the binary erasure distortion, and $D \in \mathbb{R}$, then

$$\rho_1(p_{XY}, D) = \max_{\substack{\alpha_{0e},\alpha_{1e}\in[0,1]:\\ \phi(p_{XY},\alpha_{0e},\alpha_{1e})\le D}} \psi(p_{XY}, \alpha_{0e}, \alpha_{1e}), \tag{3.8}$$

where

$$\begin{aligned}\psi(p_{XY}, \alpha_{0e}, \alpha_{1e})&\\ := \ & (p_{00}\alpha_{0e} + p_{10}\alpha_{1e})h\left(\frac{p_{00}\alpha_{0e}}{p_{00}\alpha_{0e}+p_{10}\alpha_{1e}}\right)\\ &+(p_{01}\alpha_{0e} + p_{11}\alpha_{1e})h\left(\frac{p_{01}\alpha_{0e}}{p_{01}\alpha_{0e}+p_{11}\alpha_{1e}}\right)\\ &+(p_{00}+p_{01})h\left(\frac{p_{00}}{p_{00}+p_{01}}\right) + (p_{11}+p_{10})h\left(\frac{p_{11}}{p_{11}+p_{10}}\right),\end{aligned}$$

$\phi(p_{XY}, \alpha_{0e}, \alpha_{1e}) := p_X(0)\alpha_{0e} + p_X(1)\alpha_{1e}$, and h is the binary entropy function: $h(\theta) := -\theta\log_2\theta - \overline{\theta}\log_2\overline{\theta}, \theta \in [0,1]$.

Finally, for a DSBS with parameter p and the binary erasure distortion, ρ_1 reduces to the compact expression in Proposition 3.

Proposition 3: If d is the binary erasure distortion, $D \in [0,1]$, and p_{XY} is the joint pmf of a DSBS with parameter p, then

$$\rho_1(p_{XY}, D) = (1+D)h(p). \tag{3.9}$$

• *Right-side of (3.5):* Solving the rate reduction functionals in the right-side of (3.5) requires solving the maximization problem (3.8) for asymmetric distributions $p_{X|Y}p_{Y,1}$ and $p_{X|Y}p_{Y,2}$. Exactly solving this problem is cumbersome but it is easy to provide a lower bound for the maximum as follows.

Proposition 4: If d is the binary erasure distortion, $p_{Y,1}$ is Bernoulli(q), $p_{Y,2}$ is Bernoulli$(\overline{q})$, and $p_{X|Y}$ is the conditional pmf of the binary symmetric channel with crossover probability p, then the inequality

$$\frac{\rho_1(p_{X|Y}p_{Y,1}, D) + \rho_1(p_{X|Y}p_{Y,2}, D)}{2} \ge C(p, q, \alpha_{0e}, 1) \tag{3.10}$$

holds for $D = \eta(p, q, \alpha_{0e}, 1)$, where

$$\begin{aligned} C(p,q,\alpha_{0e},\alpha_{1e}) &:= \psi(p_{X|Y}p_{Y,1}, \alpha_{0e}, \alpha_{1e}),\\ \eta(p,q,\alpha_{0e},\alpha_{1e}) &:= \phi(p_{X|Y}p_{Y,1}, \alpha_{0e}, \alpha_{1e}).\end{aligned}$$

Remark 1: The rate-distortion tuple $(H(X|Y) + H(Y|X) - C(p,q,\alpha_{0e},1), \eta(p,q,\alpha_{0e},1))$ is admissible for one-message source coding for joint source distribution $p_{X|Y}p_{Y,1}$ and corresponds to choosing $p_{U|X}$ given by (3.7) with $\alpha_{1e} = 1$ and the decoding function $g(u,y) = u$. Since this choice of $p_{U|X}$ and g

[4] For any $a \in [0,1]$, $\overline{a} := 1 - a$. For the erasure symbol e, $\overline{e} := e$.

may be suboptimal, $C(p, q, \alpha_{0e}, 1)$ is only a lower bound for the rate reduction functional.

• *Comparing left and right sides of (3.5):* The left-side of (3.5) and the lower bound of the right-side of (3.5) can be compared as follows.

Proposition 5: Let d be the binary erasure distortion, p_Y be Bernoulli(1/2), and $p_{X|Y}$ be the binary symmetric channel with parameter p. For all $q \in (0, 1/2)$ and all $\alpha_{0e} \in (0, 1)$, there exists $p \in (0, 1)$ such that the strict inequality $\rho_1(p_{XY}, D) < C(p, q, \alpha_{0e}, 1)$ holds for $D = \eta(p, q, \alpha_{0e}, 1)$.

Since the left-side of (3.5) is strictly less than a lower bound of the right-side of (3.5), the strict inequality (3.5) holds, which completes the proof of Theorem 1. ■

Theorem 2 quantifies the multiplicative reduction in the sum-rate that is possible with two messages.

Theorem 2: If d is the binary erasure distortion and p_{XY} the joint pmf of a DSBS with parameter p, then for all $L > 0$ there exists an admissible two-message rate-distortion tuple (R_1, R_2, D) such that $R_{sum,1}(p_{XY}, D)/(R_1 + R_2) > L$ and $R_1/R_2 < 1/L$.

Proof: We will explicitly construct $p_{V_1|Y}, p_{V_2|XV_1}$, and g in (2.2) which lead to an admissible tuple (R_1, R_2, D). Let $p_{V_1|Y}$ be the conditional pmf of the binary symmetric channel with crossover probability q. Let the conditional distribution $p_{V_2|XV_1}(v_2|x, v_1)$ have the form described in Table I and let $g(v_1, v_2, y) := v_2$.

TABLE I

CONDITIONAL DISTRIBUTION $p_{V_2|XV_1}$

$p_{V_2\|XV_1}$	$v_2 = 0$	$v_2 = e$	$v_2 = 1$
$x = 0, v_1 = 0$	$1 - \alpha$	α	0
$x = 1, v_1 = 0$	0	1	0
$x = 0, v_1 = 1$	0	1	0
$x = 1, v_1 = 1$	0	α	$1 - \alpha$

The corresponding rate-distortion tuple can be shown to satisfy the following property.

Proposition 6: Let d be the binary erasure distortion and let p_{XY} be the joint pmf of a DSBS with parameter p. For $p_{V_1|Y}, p_{V_2|XV_1}$, and g as described above, and all $L > 0$, there exist parameters p, q, α such that the two-message rate-distortion tuple (R_1, R_2, D) given by $R_1 = I(Y; V_1|X)$, $R_2 = I(X; V_2|Y, V_1)$, $D = \mathbb{E}[d(X, V_2)]$ satisfies $R_{sum,1}(p_{XY}, D)/(R_1 + R_2) > L$ and $R_1/R_2 < 1/L$.

This completes the proof of Theorem 2. ■

The conditional pmfs $p_{V_1|Y}$ and $p_{V_2|XV_1}$ in the proof of Theorem 2 are related to the conditional pmf $p_{U|X}$ in the proof of Theorem 1 as follows. Given $V_1 = 0$, the conditional distribution $p_{XYV_2|V_1}(x, y, v_2|0) = p_{Y,1}(y)p_{X|Y}(x|y)p_{U|X}(v_2|x)$, where $p_{U|X}$ is given by (3.7) with $\alpha_{0e} = \alpha$ and $\alpha_{1e} = 1$. Given $V_1 = 1$, the conditional distribution $p_{XYV_2|V_1}(x, y, v_2|1) = p_{Y,2}(y)p_{X|Y}(x|y)p_{U|X}(v_2|x)$, where $p_{U|X}$ is given by (3.7) with $\alpha_{1e} = \alpha$ and $\alpha_{0e} = 1$. Conditioning on V_1, in effect, decomposes the two-message problem into two one-message problems that were analyzed in the proof of Theorem 1.

IV. PROOFS

Proof of Proposition 1: Given a general $p_{U|X}$ and g satisfying the original constraint in (3.6), we will construct U^* satisfying the stronger constraints in Proposition 1 with an objective function that is not less than the original one as follows.

Without loss of generality, we assume $\mathrm{supp}(p_U) = \mathcal{U}$. For $i = 0, 1$, let $\mathcal{U}_i := \{u \in \mathcal{U} : p_{X|U}(i|u) = 1\}$. Let $\mathcal{U}_e := \{u \in \mathcal{U} : p_{X|U}(1|u) \in (0, 1)\}$. Then $\{\mathcal{U}_1, \mathcal{U}_0, \mathcal{U}_e\}$ forms a partition of $\mathcal{U}$. For each $u \in \mathcal{U}_e$, since $p_{XY|U}(x, y|u) > 0$ for all $(x, y) \in \{0, 1\}^2$, it follows that $g(u, y = 0) = g(u, y = 1) = e$ must hold, because otherwise $\mathbb{E}(d(X, g(U, Y))) = \infty$. But for every $u \in \mathcal{U}_i$, $i = 0, 1$, $g(u, y)$ may equal i or e but not $(1 - i)$ to get a finite distortion. When we replace g by

$$g^*(u, y) = \begin{cases} i, & \text{if } u \in \mathcal{U}_i, i = 0, 1, \\ e, & \text{if } u \in \mathcal{U}_e, \end{cases}$$

the distortion for $u \in \mathcal{U}_i, i = 0, 1$, is reduced to zero, and the distortion for $u \in \mathcal{U}_e$ remains unchanged. Therefore we have $\mathbb{E}(d(X, g^*(U, Y))) \le \mathbb{E}(d(X, g(U, Y))) \le D$. Note that $g^*(U, Y)$ is completely determined by U. Let $U^* := g^*(U, Y)$. Then $U^* = i$ iff $U \in \mathcal{U}_i, i = \{0, 1, e\}$. The objective function $H(X|Y, U) + H(Y|X) = H(X|Y, U, U^*) + H(Y|X) \le H(X|Y, U^*) + H(Y|X)$, which completes the proof. ■

Proof of Proposition 3:

For a fixed p_{XY}, $H(X|Y, U) + H(Y|X)$ is concave with respect to p_{XYU} and therefore also $p_{U|X}$. Since $p_{U|X}$ is linear with respect to $(\alpha_{0e}, \alpha_{1e})$, $\psi(p_{XY}, \alpha_{0e}, \alpha_{1e}) = H(X|Y, U) + H(Y|X)$ is concave with respect to $(\alpha_{0e}, \alpha_{1e})$.

The maximum in (3.8) can be achieved along the axis of symmetry given by $\alpha_{1e} = \alpha_{0e}$ because (i) ψ and ϕ are both symmetric with respect to α_{0e} and α_{1e}, i.e., $\psi(p_{XY}, \alpha_{0e}, \alpha_{1e}) = \psi(p_{XY}, \alpha_{1e}, \alpha_{0e})$ and $\phi(p_{XY}, \alpha_{0e}, \alpha_{1e}) = \phi(p_{XY}, \alpha_{1e}, \alpha_{0e})$, and (ii) $\psi(p_{XY}, \alpha_{0e}, \alpha_{1e})$ is a concave function of $(\alpha_{0e}, \alpha_{1e})$. When $D \in [0, 1]$, ρ_1 can be further simplified as follows.

$$\rho_1(p_{XY}, D) = \max_{\alpha_{0e} = \alpha_{1e} \in [0, D]} \psi(p_{XY}, \alpha_{0e}, \alpha_{1e}) = (1 + D)h(p),$$

which completes the proof. ■

Proof of Proposition 4: For the joint pmf $p_{X|Y}p_{Y,1}$ summarized in Table II, functions ψ and η simplify even further to special functions of $(p, q, \alpha_{0e}, \alpha_{1e})$ as follows:

TABLE II

JOINT DISTRIBUTION $p_{X|Y}p_{Y,1}$

$p_{X\|Y}p_{Y,1}$	$y = 0$	$y = 1$
$x = 0$	$\bar{p}\bar{q}$	pq
$x = 1$	$p\bar{q}$	$\bar{p}q$

$$\begin{aligned} C(p, q, \alpha_{0e}, \alpha_{1e}) &= \psi(p_{X|Y}p_{Y,1}, \alpha_{0e}, \alpha_{1e}) \\ &= \bar{q}(\bar{p}\alpha_{0e} + p\alpha_{1e})h\left(\frac{\bar{p}\alpha_{0e}}{\bar{p}\alpha_{0e} + p\alpha_{1e}}\right) \\ &\quad + q(p\alpha_{0e} + \bar{p}\alpha_{1e})h\left(\frac{p\alpha_{0e}}{p\alpha_{0e} + \bar{p}\alpha_{1e}}\right) \\ &\quad + (\bar{p}\bar{q} + pq)h\left(\frac{\bar{p}\bar{q}}{\bar{p}\bar{q} + pq}\right) \\ &\quad + (\bar{p}q + p\bar{q})h\left(\frac{\bar{p}q}{\bar{p}q + p\bar{q}}\right), \qquad (4.11) \\ \eta(p, q, \alpha_{0e}, \alpha_{1e}) &= \phi(p_{X|Y}p_{Y,1}, \alpha_{0e}, \alpha_{1e}) \\ &= (\bar{p}\bar{q} + pq)\alpha_{0e} + (\bar{p}q + p\bar{q})\alpha_{1e}. \end{aligned}$$

Observe that $C(p,q,\alpha_{0e},\alpha_{1e}) = C(p,\bar{q},\alpha_{1e},\alpha_{0e})$, and $\eta(p,q,\alpha_{0e},\alpha_{1e}) = \eta(p,\bar{q},\alpha_{1e},\alpha_{0e})$ hold. Therefore we have

$$\begin{aligned}\rho_1(p_{X|Y}p_{Y,2},D) &= \max_{\substack{\alpha_{0e},\alpha_{1e}\in[0,1]:\\ \eta(p,\bar{q},\alpha_{0e},\alpha_{1e})\le D}} C(p,\bar{q},\alpha_{0e},\alpha_{1e}) \\ &= \max_{\substack{\alpha_{0e},\alpha_{1e}\in[0,1]:\\ \eta(p,q,\alpha_{1e},\alpha_{0e})\le D}} C(p,q,\alpha_{1e},\alpha_{0e}) \\ &= \rho_1(p_{X|Y}p_{Y,1},D).\end{aligned}$$

It follows that

$$\begin{aligned}\frac{\rho_1(p_{X|Y}p_{Y,1},D)+\rho_1(p_{X|Y}p_{Y,2},D)}{2} &= \rho_1(p_{X|Y}p_{Y,1},D) \\ &\ge C(p,q,\alpha_{0e},1)\end{aligned}$$

holds for $D = \eta(p,q,\alpha_{0e},1)$. ■

Proof of Proposition 5:

Since $D = \eta(p,q,\alpha_{0e},1) \in [0,1]$ always holds, we have $\rho_1(p_{XY},D) = (1+D)h(p)$ due to (3.9). We will show that for any fixed $q \in (0,1/2)$ and $\alpha_{0e} \in (0,1)$, $\lim_{p\to 0} C(p,q,\alpha_{0e},1)/h(p) > \lim_{p\to 0}(1+D)$ holds, which implies that $\exists p \in (0,1)$ such that $C(p,q,\alpha_{0e},1)/h(p) > (1+D)$, which, in turn, implies Proposition 5. It is convenient to use the following lemma to analyze the limits.

Lemma 2: Let $f(p)$ be a function differentiable around $p = 0$ such that $f(0) = 0$ and $f'(0) > 0$. Then

$$\lim_{p\to 0}\frac{h(f(p))}{h(p)} = f'(0)$$

Proof: Applying the l'Hôpital rule several times, we have

$$\begin{aligned}\lim_{p\to 0}\frac{h(f(p))}{h(p)} &= \lim_{p\to 0}\frac{\ln(1-f(p))-\ln f(p)}{\ln(1-p)-\ln p}f'(0) \\ &= \lim_{p\to 0}\frac{\ln f(p)}{\ln p}f'(0) \\ &= \lim_{p\to 0}\frac{p}{f(p)}(f'(0))^2 \\ &= f'(0),\end{aligned}$$

which completes the proof of Lemma 2. ■

Applying Lemma 2, we have

$$\lim_{p\to 0}\frac{C(p,q,\alpha_{0e},1)}{h(p)} = 2-q(1-\alpha_{0e}), \tag{4.12}$$

$$\lim_{p\to 0}(1+D) = 2-\bar{q}(1-\alpha_{0e}), \tag{4.13}$$

$$\lim_{p\to 0}\left(\frac{C(p,q,\alpha_{0e},1)}{h(p)}-(1+D)\right) = (1-2q)(1-\alpha_{0e}).$$

Therefore for any $\alpha_{0e} \in (0,1)$ and $q \in (0,1/2)$, there exists a small enough p such that $C(p,q,\alpha_{0e},1) > (1+D)$ holds, which completes the proof. ■

Proof of Proposition 6:

For the rate-distortion tuple (R_1,R_2,D) corresponding to the choice of $p_{V_1|Y}$, $p_{V_2|XV_1}$ and g described in the proof of Theorem 2, we have (i) $R_1 = I(Y;V_1|X) = H(Y|X) - C_2(p,q)$, where $C_2(p,q)$ is the sum of the last two terms in (4.11); (ii) $R_2 = I(X;V_2|Y,V_1) = 2h(p) - C(p,q,\alpha,1) - R_1$; and (iii) $D = \eta(p,q,\alpha,1)$. It follows that

$$\begin{aligned}\lim_{p\to 0}\frac{R_1}{h(p)} &= 0, \\ \lim_{p\to 0}\frac{R_2}{h(p)} &= 2-\lim_{p\to 0}\frac{C(p,q,\alpha,1)}{h(p)}-\lim_{p\to 0}\frac{R_1}{h(p)} = q(1-\alpha).\end{aligned}$$

Therefore for all $q > 0$ and $\alpha \in (0,1)$, we have

$$\lim_{p\to 0}\frac{R_1}{R_2} = 0. \tag{4.14}$$

For the one-message rate-distortion function, we have $R_{sum,1}(p_{XY},D) = 2h(p) - \rho_1(p_{XY},D)$, where $\rho_1(p_{XY},D)$ is given by (3.9). Therefore we have

$$\lim_{p\to 0}\frac{R_{sum,1}(p_{XY},D)}{h(p)} = 2-\lim_{p\to 0}\frac{\rho_1(p_{XY},D)}{h(p)} = \bar{q}(1-\alpha),$$

which implies that

$$\lim_{p\to 0}\frac{R_{sum,1}(p_{XY},D)}{R_1+R_2} = \frac{\bar{q}}{q}. \tag{4.15}$$

For any $L > 0$, we can always find a small enough $q > 0$ such that $\bar{q}/q > L+1$. Due to (4.14) and (4.15), there exists $p > 0$ such that $R_1/R_2 < 1/L$ and $R_{sum,1}/(R_1+R_2) > L$. ■

Remark 2: The convergence of the limit analyzed in Lemma 2 is actually slow, because the logarithm function increases to infinity slowly. The consequence is that if one chooses a small q to get $R_{sum,1}/(R_1+R_2)$ close to the limit $\bar{q}/q$, then p needs to be very small. For example, when $q = 1/10, \alpha_{0e} = 1/2$, $\bar{q}/q = 9$, with $p = 10^{-200}$, we get $R_{sum,1}/R^*_{sum,2} \approx 8.16$. This, however, does not mean that the benefit of multiple messages only occurs in extreme cases. In numerical computations we have observed that for the erasure distortion, the gain for certain asymmetric sources can be much more than that for the DSBS example analyzed in this paper. The DSBS example was chosen in this paper only because it is easy to analyze.

Acknowledgment

The second author would like to thank Dr. Vinod Prabhakaran for introducing him to the unresolved question in [1].

References

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

[2] E. Yang and D. He, "On interactive encoding and decoding for lossless source coding with decoder only side information," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, Canada, Jul. 6–11, 2008, pp. 419–423.

[3] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[4] N. Ma and P. Ishwar, "Two-terminal distributed source coding with alternating messages for function computation," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, Canada, Jul. 6–11, 2008, pp. 51–55, arXiv:0801.0756v4.

[5] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–11, Jan. 1976.

[6] N. Ma and P. Ishwar, "Infinite-message distributed source coding for two-terminal interactive computing," in *Proc. 47th Annu. Allerton Conf. Communication, Control, Computing*, Monticello, IL, Sep. 30–Oct. 7, 2009.

[7] ——, "Infinite-message distributed source coding for two-terminal interactive computing," *arXiv:0908.3512v2*.