ISIT 2009, Seoul, Korea, June 28 - July 3, 2009

On Feedback in Network Source Coding

Mayank Bakshi, Michelle Effros
Department of Electrical Engineering
California Institute of Technology
Pasadena, California 91125, USA
Email: {mayank, effros}@caltech.edu

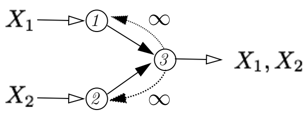


Fig. 1. Slepian-Wolf network with feedback

***Abstract*— We consider source coding over networks with unlimited feedback from the sinks to the sources. We first show examples of networks where the rate region with feedback is a strict superset of that without feedback. Next, we find an achievable region for multiterminal lossy source coding with feedback. Finally, we evaluate this region for the case when one of the sources is fully known at the decoder and use the result to show that this region is a strict superset of the best known achievable region for the problem without feedback.**

I. INTRODUCTION

The networks studied in the source coding literature are typically directed, acyclic graphs. Just as it is well known that feedback cannot increase the capacity of the canonical point-to-point channel [1], it is also evident that feedback cannot increase the rate region of the canonical point-to-point lossless and lossy source coding problems. We here examine the role of feedback in network source coding, demonstrating that feedback can increase the rate region for network source coding in some networks where the rate region is well understood and that feedback can increase the known set of achievable rates in one example network where the rate region remains unsolved. While we here focus on examples where feedback increases the achievable rate region, it is important to note that feedback does not increase the rate region for all networks or even all network topologies where feedback has the potential to increase the channel capacity. For example, in the Slepian-Wolf system[2] shown in Fig 1, the presence of feedback from receiver node 3 to source nodes 1 and 2 does not increase the min-cut and therefore does not enable operation at lower rates on the forward links. In contrast, it is well known that feedback can increase the capacity region of the multiple access channel [3].

Following the typical approach from channel coding, we here assume unbounded capacity on the feedback links and then consider the rate region for the forward links only. While this approach is chosen for its simplicity, the resulting insights may be directly applicable in networks where the cost of operating the feedback link is negligible compared to the cost of the forward links. For example, in sensor networks, where the central receiver node usually has much more power available than the remote sensors, the cost of sending information from the central processor back to the sensors may be far less than the cost of forward links. If transmitting information from the central processor to the sensors decreases the rate required on the forward links, then an overall system benefit might be realized.

This paper considers the problem of characterizing the set of achievable rates for source coding networks in the presence of unlimited feedback. In Section III, we show through several examples that feedback can enable operation at rate points that are not achievable otherwise. Examples 1 and 3 illustrate that codes that make available to the transmitter all the information that is known at the receiver require less rate form the transmitter. Example 2 demonstrates that even with independent sources in a multi-source, multi-sink network, feedback from the sinks to the sources increases the minimum-cut between sources and their sinks and results in an increased capacity region.

In Section IV, we examine a multiterminal lossy source coding problem with two encoders. While the rate region without feedback remains unsolved, we show that feedback enables lower rates than the best achievable rates known to date. The result of Example 3 is a special case of this network that demonstrates the tightness of our bound at the extreme points, showing that feedback strictly enlarges the rate region for this network.

We begin by describing our setup and introducing necessary notation in Section II.

II. PRELIMINARIES

Let V be the set of nodes and let $S = \{1, 2, \ldots, s\} \subseteq V$ be the set of source nodes. Source nodes are connected to the set of receiver nodes T forming an acyclic network with directed, lossless edges E. In addition, each receiver node is connected to each source node via a directed, lossless edge from a set F. The sets E and F are called the set of *forward links* and the set of *feedback links*, respectively. For a node $v \in V$, the sets $\Gamma_i(v)$ and $\Gamma_o(v)$ denote the incoming and outgoing forward inks, respectively. Similarly, $\Gamma_i(V')$ (resp. $\Gamma_o(V')$) denotes the set of incoming (resp. outgoing) forward links for a set of nodes $V' \subseteq V$.

[0]This work is partially supported by NSF Grant CCF-0325324, DARPA Grant W911NF-07-I0029, and the Lee Center for Advanced Networking at Caltech.

978-1-4244-4313-0/09/$25.00 ©2009 IEEE

Let $X_1, X_2, \ldots, X_s$ be discrete random variables distributed according to a joint probability distribution $P_{X_1 \ldots X_s}(\cdot)$ on a finite alphabet $\prod_{j=1}^{s} \mathcal{X}_j$. For $S' \subseteq S$, we denote the collection of random variables $(X_j : j \in S')$ by $X_{S'}$. Each source node $j \in S$ observes the random process $\{X_j(i)\}_{i=1}^{\infty}$, where, the random process $\{X_S(i)\}_{i=1}^{\infty}$ is drawn i.i.d from the joint distribution $P_{X_S}(\cdot)$. For an integer $n \geq 1$ and a subset $S' \subseteq S$, $X_{S'}[n]$ denotes the collection of random variables $\{X_{S'}(i)\}_{i=1}^{n}$.

Each receiver node $t \in T$ demands a reconstruction $(\widehat{X}_1^{(t)}, \widehat{X}_2^{(t)}, \ldots, \widehat{X}_s^{(t)}) \in \prod_{j=1}^{s} \widehat{\mathcal{X}}_j$ of X_S. For lossless source coding, the demand must be met with asymptotically negligible error probability. For lossy coding, the demand must be met subject to distortion criteria of the form $\{Ed_j(X_j, \widehat{X}_j^{(t)}) \leq D_j\}$ at each $t \in T$ for some finite-valued distortion measures $d_j : \mathcal{X}_j \times \widehat{\mathcal{X}}_j \rightarrow \mathbb{R}^+$, and distortion thresholds D_j, $j \in S$. When the network has only one sink t, we denote the reconstruction $(\widehat{X}_1^{(t)}, \widehat{X}_2^{(t)}, \ldots, \widehat{X}_s^{(t)})$ by $(\widehat{X}_1, \widehat{X}_2, \ldots, \widehat{X}_s)$.

For any collection of rates $(R_e : e \in E)$ with $R_e \geq 0$ for all $e \in E$, a $((2^{nR_e})_{e \in E}, n, L)$ network code $(f^n_{\{1,2,\ldots,L\} \times E}, g^n_T)$ defines a transmission strategy over L sessions with encoders

$$\begin{aligned} f^n_{l,vv'} \quad &: \quad \mathcal{X}_v^n \times \prod_{\substack{1 \leq r < l \\ e \in \Gamma_i(T \cup \{v\})}} \{1, \ldots, 2^{nR_e^{(r)}}\} \times \\ &\quad \prod_{e \in \Gamma_i(v)} \{1, \ldots, 2^{nR_e^{(l)}}\} \rightarrow \{1, \ldots, 2^{nR_{vv'}^{(l)}}\} \\ &\quad \forall v \in S, (v, v') \in E, l = 1, 2, \ldots, L \\ f^n_{l,vv'} \quad &: \quad \prod_{\substack{1 \leq r \leq l \\ e \in \Gamma_i(v)}} \{1, \ldots, 2^{nR_e^{(r)}}\} \rightarrow \{1, \ldots, 2^{nR_{vv'}^{(l)}}\} \\ &\quad \forall v \notin S, (v, v') \in E, l = 1, 2, \ldots, L \end{aligned}$$

such that for each $e \in E$,

$$\sum_{l=1}^{L} R_e^{(l)} = R_e,$$

and decoders

$$g_t^n : \prod_{\substack{1 \leq r \leq L \\ e \in \Gamma_i(t)}} \{1, \ldots, 2^{nR_e^{(r)}}\} \rightarrow \prod_{j=1}^{s} \widehat{\mathcal{X}}_j \quad \forall t \in T.$$

We say that a rate vector $R_E = (R_e : e \in E)$ is losslessly achievable with feedback if for any $\epsilon > 0$, there exists a $((2^{nR_e})_{e \in E}, n, L)$ code for some $n, L \geq 1$ such that

$$\Pr(\widehat{X}_S^{(t)}[n] \neq X_S[n]) < \epsilon \tag{1}$$

for all $t \in T$. For lossy coding, we say that the rate vector R_E achieves distortion constraints (D_1, D_2) with feedback if for any $\epsilon > 0$, there exists a $((2^{nR_e})_{e \in E}, n, L)$ code for some $n, L \geq 1$ such that

$$\frac{1}{n} Ed_j(X_j[n], \widehat{X}_j^{(t)}[n]) \leq D_j \tag{2}$$

for all $j \in S, t \in T$. The closure of the set of all achievable rate vectors with feedback is denoted by $\mathcal{R}^*_{\text{fb}}$.

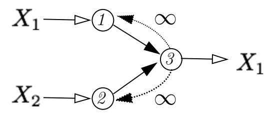


Fig. 2. Lossless source coding problem with coded side information and feedback

Codes without feedback differ from codes with feedback in two ways. First, only a single session is required ($L = 1$). Second, the encoders at nodes $j \in S$ rely only on their respective sources and incoming codewords giving encoders of the form

$$\begin{aligned} f^n_{vv'} \quad &: \quad \mathcal{X}_v^n \times \prod_{e \in \Gamma_i(v)} \{1, \ldots, 2^{nR_e}\} \rightarrow \{1, \ldots, 2^{nR_{vv'}}\} \\ &\quad \forall v \in S, (v, v') \in E \\ f^n_{vv'} \quad &: \quad \prod_{e \in \Gamma_i(v)} \{1, \ldots, 2^{nR_e}\} \rightarrow \{1, \ldots, 2^{nR_{vv'}}\} \\ &\quad \forall v \notin S, (v, v') \in E \end{aligned}$$

and decoders

$$g_t^n : \prod_{e \in \Gamma_i(t)} \{1, \ldots, 2^{nR_e}\} \rightarrow \prod_{j=1}^{s} \widehat{\mathcal{X}}_j \quad \forall t \in T.$$

The closure of the set of all achievable rate vectors without feedback is denoted by $\mathcal{R}^*$.

III. Networks where feedback helps

In this section, we give three examples of networks where $\mathcal{R}^*_{\text{fb}} \supsetneq \mathcal{R}^*$. The first two examples demonstrate that feedback can expand the rate region even for lossless coding. This is in contrast to the point-to-point case, where feedback cannot increase the rate region. The first example is the coded side information network, which has been studied previously without feedback in [4] and with partially separated encoders in [5].

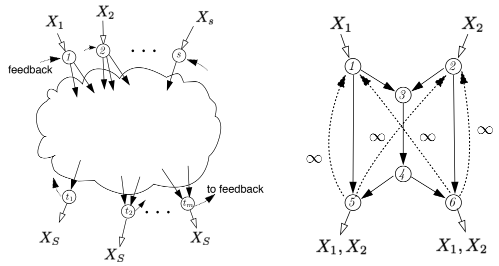
Example 1 (Source coding with coded side information): Consider the network shown in Fig 2. The encoders 1 and 2 observe sources X_1 and X_2 respectively, and the decoder at node 3 wishes to reconstruct X with arbitrarily small probability of error. Without feedback [4], the rate region, $\mathcal{R}^*$ is the collection of all rate pairs $R_E = (R_{13}, R_{23})$ that satisfy the following inequalities for some random variable U forming a Markov chain $U \rightarrow X_2 \rightarrow X_1$:

$$R_{13} \geq H(X_1|U) \tag{3}$$
$$R_{23} \geq I(X_2; U). \tag{4}$$

For some rate points on the boundary of this region, the sum rate $R_{13} + R_{23}$ can be strictly greater than $H(X_1)$ [6]. Feedback increases the rate region as follows:

Claim: For the network $\mathcal{N}$ shown in Fig 2, a rate pair $R_E = (R_{13}, R_{23})$ is achievable if and only if

$$R_{13} \geq H(X_1|X_2) \tag{5}$$
$$R_{13} + R_{23} \geq H(X_1). \tag{6}$$



(a) Multicast network (b) Butterfly network

Fig. 3. Multicast with feedback

Proof: The necessity of (5) and (6) follows from simple cutset arguments. The achievability of these rates can be shown using a transmission strategy over two sessions that relies on the partially separated encoding scheme of Kaspi and Berger [5]. We present a simpler proof of achievability of these rates by using a two-step Slepian-Wolf code. Let $R_E = (R_{13}, R_{23})$ satisfy (5) and (6). Given a block length n, let (f_1^n, f_2^n, g^n) be an n-length Slepian-Wolf code [2] at rate (R_{13}, R_{23}) for a network in which both encoders observe the source X_1 and independently describe X_1 to a shared decoder. Let $(\widetilde{f}_1^n, \widetilde{g}^n)$ be a rate R_{13} Slepian-Wolf code [2] for describing X_1 to a decoder that knows X_2. We design $\widetilde{f}_1^n$ by randomly binning X_1. We here set $f_1^n(X_1[n]) = \widetilde{f}_1^n(X_1[n])$ for all $X_1[n]$ and design $f_2(X_1[n])$ independently, again by random binning. Consider the following coding strategy for the network $\mathcal{N}$.

- Node 1 transmits $\widetilde{f}_1^n(X_1[n])$ to node 3 at a rate R_{13}. This description is then made available to node 2 via feedback.
- Node 2 performs a Slepian-Wolf decoding to create a reconstruction $\widehat{X}_1[n] = \widetilde{g}^n(\widetilde{f}_1^n(X_1[n]), X_2[n])$ of $X_1[n]$, and transmits the codeword $f_2^n(\widehat{X}_1[n])$ to node 3 using rate R_{23}.

Using the above strategy, a decoding error occurs if either node 2 is unable to reconstruct $X_1[n]$ using $(\widetilde{f}_1^n(X_1[n]), X_2[n])$, or node 3 is unable to decode $X_1[n]$ using $(f_1^n(X_1[n]), f_2^n(\widehat{X}_1[n]))$. Since R_E satisfies (5) and (6), following the arguments of [2], by choosing a large enough n, the probabilities for both these error events can be made arbitrarily small. Hence, the overall error probability can be designed to be as low as desired. ∎

Next, we consider networks with multicast demands, which have been studied in the context of network coding in [7], [8], [9]. Once again, we demonstrate that feedback enables rates which are not achievable otherwise.

Example 2 (Networks with multicast demands): Consider the network shown in Fig 3(a). Nodes in $T \subseteq V$ demand each of the sources $X_1, X_2, \ldots, X_s$ losslessly. We show that feedback expands the rate region to include all rates $R_E = (R_e : e \in E)$ satisfying:

$$\sum_{e \in \Gamma_o(C)} R_e \geq H(X_{C \cap S} | X_{S \setminus C}) \tag{7}$$

for each cut $C \subseteq V$ such that either $S \subseteq C$ or $C \cap T = \phi$.

Claim: For the network $\mathcal{N}$ shown in Figure 3(a), $\mathcal{R}^*_{\text{fb}} \supsetneq \mathcal{R}^*$.

Proof: The rate region with feedback is found by evaluating the characterization given in previous results on cyclic network [7], [8]. Since sources are connected to the sinks via the feedback links, the max flow calculation from a given source to a sink is modified by adding acyclic paths containing feedback links to other sources. The converse follows immediately by noting that (7) can be viewed as a collection of cut-set bounds on the cyclic network.

To verify the achievability of the above rates, note that in order to meet all the demands, it is sufficient that for any set $S' \subseteq S$ of source nodes, there exists a subset $T_{S'} \subseteq T$ of sink nodes for which the min-cut (without feedback) exceeds $H(X_{S'}|X_S)$ and by using feedback from $T_{S'}$ to $S \setminus S'$, the min-cut requirement from S' nodes in $T \setminus T_{S'}$ is also satisfied. This condition is satisfied for all rates satisfying (7).

Finally, the butterfly network shown in Fig 3(b) shows that the rate region with feedback may be strictly bigger than that without feedback. The rate vector $R_E = (R_e : e \in E)$, where $R_{15} = R_{26} = H(X_1, X_2), R_e = 0 \ \forall \ e \notin \{(1,5),(2,6)\}$, satisfies (7) but is not achievable without the feedback links. With feedback, it may be achieved by transmitting both X_1 and X_2 over the links $(1,5)$ and $(2,6)$. ∎

The next example is a lossy source coding problem where feedback can increase the rate region [10]. We use this result in order to prove Theorem 2.

Example 3 (Rate-Distortion coding with Side Information): Consider the network shown in Fig 4(a) The decoder at node 3 demands a lossy reconstruction $\widehat{X}_1$ of X_1 subject to a distortion criterion $Ed(X_1, \widehat{X}_1) \leq D$. Without feedback, the minimum rate achievable is described by the Wyner-Ziv region [10]:

$$R_{12} = \min_{\substack{(U,g): Ed(X_1, g(U, X_2)) \leq D \\ U \to X_1 \to X_2}} I(X_1; U | X_2). \tag{8}$$

When feedback is present in the network, both the encoder and the decoder have knowledge of X_2. In this case, the minimum achievable rate is given by the conditional rate-distortion function $R_{X_1|X_2}(D)$, given by [10]:

$$R_{X_1|X_2}(D) = \min_{U: Ed(X_1, U) \leq D} I(X_1; U | X_2). \tag{9}$$

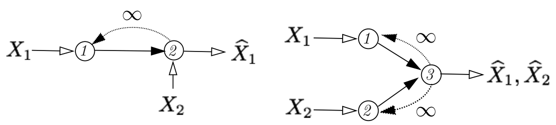
For some choices of sources X_1 and X_2 and distortion measure d, the expression in Eq (8) is strictly greater than $R_{X_1|X_2}(D)$ [10]. Thus, feedback increases the rate region for this network.

IV. Achievable Rates for Multiterminal Lossy Source Coding with Feedback

In this section, we examine the network shown in Fig. 4(b). Sources X_1 and X_2 are sources present at nodes 1 and 2 respectively. The receiver (node 3) wishes to reconstruct both sources subject to the fidelity criteria:

$$Ed_1(X_1, \widehat{X}_1) \leq D_1$$
$$\text{and} \quad Ed_2(X_2, \widehat{X}_2) \leq D_2$$

where, d_1 and d_2 are finite valued distortion measures, and D_1 and D_2 are the respective distortion thresholds. We derive an



(a) The Wyner-Ziv network with feedback (b) Multiterminal source coding with feedback

Fig. 4. Lossy source coding with feedback

achievable region $\mathcal{R}_{\rm in,fb}$ with feedback, and show that $\mathcal{R}_{\rm in,fb}$ is a strict superset of $\mathcal{R}_{\rm in}$, the best known achievable region without feedback. This is proved by evaluating both $\mathcal{R}_{\rm in,fb}$ and $\mathcal{R}_{\rm in}$ for the network considered in example 3, which is a special case. .

Let $\mathcal{D}(D_1, D_2)$ denote the set of pairs of random variables $(U, V) \in \mathcal{U}\times\mathcal{V}$ for which there exist functions $\alpha : \mathcal{U}\times\mathcal{V} \to \widehat{\mathcal{X}}_1$ and $\beta : \mathcal{U} \times \mathcal{V} \to \widehat{\mathcal{X}}_2$ such that. $Ed_1(X_1, \alpha(U, V)) \le D_1$ and $Ed_2(X_2, \beta(U, V)) \le D_2$. Define the set $\mathcal{R}_1$ to be the set of all rate pairs (R_{13}, R_{23}) that satisfy the conditions

$$R_{13} > I(X_1; U|V), \tag{10}$$

$$\text{and} \qquad R_{23} > I(X_2; V), \tag{11}$$

for some pair of random variables $(U, V) \in \mathcal{D}(D_1, D_2)$ for which $X_1 \to X_2 \to V$ and $U \to (X_1, V) \to X_2$ form Markov chains. In a symmetric fashion, define the set $\mathcal{R}_2$ to be the set of all rate pairs (R_{13}, R_{23}) that satisfy the conditions

$$R_{13} > I(X_1; U), \tag{12}$$

$$\text{and} \qquad R_{23} > I(X_2; V|U), \tag{13}$$

for some pair of random variables $(U, V) \in \mathcal{D}(D_1, D_2)$ for which $X_2 \to X_1 \to U$ and $V \to (X_2, U) \to X_1$ form Markov chains. Both $\mathcal{R}_1$ and $\mathcal{R}_2$ are non-empty since choosing $(U, V) = (X_1, X_2)$ satisfies all the Markov chain conditions. Finally, let $\mathcal{R}_{\rm in,fb}$ be the convex hull of $\mathcal{R}_1 \cup \mathcal{R}_2$, and again, let $\mathcal{R}^*_{\rm fb}$ denote the set of achievable rates with feedback for the network shown in Fig 4(b). The following theorem proves the achievability of $\mathcal{R}_{\rm in,fb}$.

Theorem 1: $\mathcal{R}^*_{\rm fb} \supseteq \mathcal{R}_{\rm in,fb}$.

The proof of this result relies on the notion of strong joint typicality [11], which is reviewed here briefly. Let $N_{abcd}(x_1[n], x_2[n], u[n], v[n])$ denote the number of occurrences of the quadruplet (a, b, c, d) in the sequence $(x_1[n], x_2[n], u[n], v[n])$. Define the strongly typical set:

$$A_\epsilon^{*(n)}(X_1, X_2, U, V) \triangleq \{(x_1[n], x_2[n], u[n], v[n]) : \left|\frac{1}{n}N_{abcd}(x_1[n], x_2[n], u[n], v[n]) - p(a, b, c, d)\right| < \frac{\epsilon}{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{U}||\mathcal{V}|} \forall\ (a, b, c, d) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{U} \times \mathcal{V}\}. \tag{14}$$

Similarly, for each subset W of $\{X_1, X_2, U, V\}$, define $A_\epsilon^{*(n)}(W)$ to be the typical sets corresponding to n-length sequences drawn from the distribution of W. The above definition implies that if a collection of sequences is jointly typical with respect to their joint distribution, then any subset of the collection is also jointly typical with respect to the joint distribution of that subset; for example, if $(x_1[n], x_2[n], u[n], v[n]) \in A_\epsilon^{*(n)}(X_1, X_2, U, V)$, then $(x_1[n], x_2[n], u[n]) \in A_\epsilon^{*(n)}(X_1, X_2, U)$. Therefore, whenever the set of underlying random variables is clear from the context, we denote the corresponding typical set by the simplified notation $A_\epsilon^{*(n)}$. Another useful property of this notion of typicality is that it implies distortion typicality; namely, if $(U, V) \in \mathcal{D}(D_1, D_2)$ and $(x_1[n], u[n], v[n]) \in A_\epsilon^{*(n)}$, then $\frac{1}{n}\sum_{i=1}^n d_1(x_1(i), \alpha(u(i), v(i))) < D_1 + d_{\max} \cdot \epsilon$.

Proof of Theorem 1: By the symmetry in the definition of $\mathcal{R}_{\rm in,fb}$ and the convexity of $\mathcal{R}^*_{\rm fb}$, it suffices to show that $\mathcal{R}_1 \subseteq \mathcal{R}^*_{\rm fb}$. Let $R_E = (R_{13}, R_{23}) \in \mathcal{R}_1$. By definition, there exists a pair $(U, V) \in \mathcal{D}(D_1, D_2)$ for which $X_1 \to X_2 \to V$ and $U \to (X_1, V) \to X_2$ form Markov chains and the inequalities (10) and (11) are satisfied.

Fix an integer n and an $\epsilon > 0$. Choose R'_{13} such that $I(X_1, V; U) < R'_{13} < R_{13} + I(U; V)$. The reason for this choice will become clear later. To show that the rate pair (R_{13}, R_{23}) is achievable, consider the following encoding and decoding strategy over a block of length n. *Codebook generation:* For encoder 1, first generate $2^{nR'_{13}}$ sequences $U_1[n], U_2[n], \ldots, U_{2^{nR'_{13}}}[n]$ drawn i.i.d. from the distribution $\Pi_{i=1}^n P_U(u_i)$. Uniformly bin these $2^{nR'_{13}}$ sequences into $2^{nR_{13}}$ bins. We use $B_n(j)$ to describe the index of the bin into which $U_j[n]$ falls. For the second encoder, generate $2^{nR_{23}}$ sequences $V_1[n], V_2[n], \ldots, V_{2^{nR'_{13}}}[n]$ drawn i.i.d. from the distribution $\Pi_{i=1}^n P_V(u_i)$. These codebooks are assumed known to both encoders and the decoder.

Encoding: Let $f_2^n(X_2[n]) = k$ if $(X_2[n], V_k[n]) \in A_\epsilon^{*(n)}$. Otherwise, let $f_1^n(X_2[n]) = 1$. Transmit $f_2^n(X_2[n])$ to node 3, and also to node 1 via the feedback link. Let $f_1^n(V_k[n], X_1[n]) = B_n(j)$ if $(X_1[n], V_k[n], U_j[n]) \in A_\epsilon^{*(n)}$. Otherwise, let $f_1^n(V_k[n], X_1[n]) = 1$.

Decoding: The decoder first decodes $f_2^n(X_2[n])$ to the sequence $\widehat{V}[n] = V_{f_1^n(X_2[n])}[n]$. Next, it looks for a sequence $\widehat{U}[n]$ in the bin $f_1^n(\widehat{V}[n], X_1[n])$ s.t. $(\widehat{U}[n], \widehat{V}[n]) \in A_\epsilon^{*(n)}$. Finally, it produces the reconstructions $\widehat{X}_1[n] = \alpha(\widehat{U}(1), \widehat{V}(1)), \ldots, \alpha(\widehat{U}(n)\widehat{V}(n))$ and $\widehat{X}_2[n] = \beta(\widehat{U}(1), \widehat{V}(1)), \ldots, \beta(\widehat{U}(n), \widehat{V}(n))$. Since ϵ can be made arbitrarily small, it is clear that the above coding scheme can encode at rates as close to $R_E = (R_{13}, R_{23})$ as desired. Further, since d_1 and d_2 are finite distortion measures, in order to show that the expected distortion of this code can be made arbitrarily close to (D_1, D_2), it suffices to show that $\Pr(\frac{1}{n}d_1(X_1[n], \alpha(\widehat{U}[n], \widehat{V}[n])) > D_1 + \delta)$ and $\Pr(\frac{1}{n}d_2(X_2[n], \beta(\widehat{U}[n], \widehat{V}[n])) > D_2 + \delta)$ can be made arbitrarily small for each $\delta > 0$. Thus, it is enough to prove that $\Pr(\{(X_1[n], X_2[n], \widehat{U}[n], \widehat{V}[n]) \notin A_\epsilon^{*(n)}\}$ can be made arbitrarily small for each $\epsilon > 0$. Note that

$$\{(X_1[n], X_2[n], \widehat{U}[n], \widehat{V}[n]) \notin A_\epsilon^{*(n)}\} \subseteq E_1 \cup E_2 \cup E_3 \cup E_4,$$

where, the events E_1, E_2, E_3, and E_4 are defined as follows:

- $E_1 = \{(X_1[n], X_2[n]) \notin A_\epsilon^{*(n)}\}$. By the Weak Law of Large Numbers, the probability of this event can be made arbitrarily small by choosing n large enough.
- $E_2 = E_1^c \cap \{(X_1[n], X_2[n], \widehat{V}[n] \notin A_\epsilon^{*(n)}\}$. By noting

that $X_1 \to X_2 \to V$ is a Markov chain, and using the Markov lemma [12], the probability of this event can be made to asymptotically vanish with n as long as $R_{23} > I(X_2; V)$ (see the proof of the rate distortion theorem in [11], [13] for further details on this argument).

- $E_3 = (E_1 \cup E_2)^c \cap \{(X_1[n], X_2[n], \widehat{V}[n], U_j[n]) \notin A_\epsilon^{*(n)} \ \forall\ j = 1, 2, \ldots, 2^{nR'_{13}}\}$. By following a similar reasoning as above, as long as $R'_{13} > I(X_1, V; U)$, the probability of this event can be made arbitrarily small.
- $E_4 = (E_1 \cup E_2 \cup E_3)^c \cap \{(u[n], \widehat{V}[n]) \in A_\epsilon^{*(n)}$ for some $u[n] \neq U_{f_1^n(\widehat{V}[n], X_1[n])}[n]$ s.t. $u[n]$ is in the bin $B_n(f_1^n(\widehat{V}[n], X_1[n]))\}$. The probability of this event can be made arbitrarily small too by choosing a large enough n, the number of elements in each bin is less than $2^{nI(U;V)}$ with probability approaching 1 as n grows without bound.

Thus, for any rate $R_E = (R_{13}, R_{23}) \in \mathcal{R}_1$, there exists a sequence of valid $((2^{nR_{13}}, 2^{nR_{23}}), n, 2)$ codes for this network. By a similar reasoning, $\mathcal{R}_2$ is achievable. By the convexity of $\mathcal{R}^*_{\text{fb}}$, $\mathcal{R}_{\text{in,fb}}$ is achievable. Hence, $\mathcal{R}_{\text{in,fb}} \subseteq \mathcal{R}^*_{\text{fb}}$. ■

Let $\mathcal{R}^*$ denote the set of achievable rate pairs for the network in Fig 4(b) without the feedback links. Example 3 demonstrates that $\mathcal{R}^* \subsetneq \mathcal{R}^*_{\text{fb}}$. It should be pointed out that a single letter characterization of $\mathcal{R}^*$ is not known. Berger and Tung proposed an inner bound [12], [14] $\mathcal{R}_{\text{in}} \subseteq \mathcal{R}^*$, which was shown to be optimal for Gaussian sources [15]. For other classes of sources, the question of tightness of this bound is still open. The inner bound is defined as follows:

Definition 1 (Berger-Tung inner bound): [12], [14] The Berger-Tung inner bound $\mathcal{R}_{\text{in}}$ is defined to be the set of all rate pairs (R_{13}, R_{23}) that satisfy the conditions

$$R_{13} > I(X_1; U|V), \tag{15}$$

$$R_{23} > I(X_2; V|U), \tag{16}$$

$$\text{and} \quad R_{13} + R_{23} > I(X_1, X_2; U, V), \tag{17}$$

for some random variables U and V taking values in alphabets $\mathcal{U}$ and $\mathcal{V}$ respectively, and satisfying the following properties:

1) $U \to X_1 \to X_2 \to V$ forms a Markov chain, and
2) $(U, V) \in \mathcal{D}(D_1, D_2)$.

Our next result relates $\mathcal{R}_{\text{in,fb}}$ to the Berger-Tung inner bound.

Theorem 2: For every source pair (X_1, X_2), $\mathcal{R}_{\text{in,fb}} \supseteq \mathcal{R}_{\text{in}}$. Further, there exists a source pair such that $\mathcal{R}_{\text{in,fb}} \supsetneq \mathcal{R}_{\text{in}}$.

Proof: In order to prove that $\mathcal{R}_{\text{in,fb}} \supseteq \mathcal{R}_{\text{in}}$, first note that $\mathcal{R}_{\text{in}}$ can be viewed as the convex hull of $\mathcal{R}_{1,\text{nf}} \cup \mathcal{R}_{2,\text{nf}}$, where $\mathcal{R}_{1,\text{nf}}$ (and in a similar manner, $\mathcal{R}_{2,\text{nf}}$) is defined as the set of all rate pairs $R_E = (R_{13}, R_{23}) \in \mathcal{R}_{\text{in}}$ satisfying

$$R_{13} \geq I(X_1; U), \tag{18}$$

$$R_{23} \geq I(X_2; V|U) \tag{19}$$

for some pair (U, V) of random variables that satisfy conditions 1) and 2) of Definition 1. To prove that $\mathcal{R}_{\text{in}} = conv(\mathcal{R}_{1,\text{nf}} \cup \mathcal{R}_{2,\text{nf}})$, note that for each $R_E \in \mathcal{R}_{\text{in}}$ and $\lambda \in [0, 1]$,

$$\begin{aligned} R_{13} + R_{23} \quad > \quad & (1-\lambda)(I(X_1; U) + I(X_2; U|X_1) \\ & + I(X_2; V|U) + I(X_1; V|X_2, U)) \\ & + \lambda(I(X_1; V|X_2) + I(X_2; V) \\ & + I(X_2; U|V, X_1) + I(X_1; U|V)). \end{aligned}$$

It follows that R_E can be written as a convex combination of points from $\mathcal{R}_{1,\text{nf}}$ and $\mathcal{R}_{2,\text{nf}}$. Therefore, it is sufficient to prove that $\mathcal{R}_{1,\text{nf}} \subseteq \mathcal{R}_1$. This is easy to see because the Markov condition $U \to X_1 \to X_2 \to V$ that is satisfied by every element in $\mathcal{R}_{1,\text{nf}}$ implies the Markov conditions $U \to X_1 \to X_2$ and $X_1 \to (X_2, U) \to V$. Hence, $\mathcal{R}_{1,\text{nf}} \subseteq \mathcal{R}_1$, and therefore, $\mathcal{R}_{\text{in}} \subseteq \mathcal{R}_{\text{in,fb}}$.

Finally, observe that for rates that allow X_2 to be known losslessly at the decoder, the network reduces to a single encoder source coding problem with side information at the decoder. As discussed in Example 3, the addition of feedback can lower the rate required by a non-zero quantity. Hence, $\mathcal{R}_{\text{in,fb}} \supsetneq \mathcal{R}_{\text{in}}$ for some choices of sources and distortion measures. ■

REFERENCES

[1] C. E. Shannon. The zero-error capacity of a noisy channel. *IRE Transactions on Information Theory*, IT-2:8–19, 1956.

[2] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.

[3] N.T. Gaarder and J. K. Wolf. The capacity region of a multiple-access discrete memoryless channel can increase with feedback. *IEEE Transactions on Information Theory*, IT-21:100–102, jan 1975.

[4] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *Information Theory, IEEE Transactions on*, 21(6):629–637, Nov 1975.

[5] A. Kaspi and T. Berger. Rate-distortion for correlated sources with partially separated encoders. *Information Theory, IEEE Transactions on*, 28(6):828–840, Nov 1982.

[6] D. Marco and M. Effros. A partial solution for lossless source coding with coded side information. In *Proceedings of the Information Theory Workshop*, Punta del Este, Uruguay, March 2006. IEEE.

[7] R. Ahlswede, N. Cai, S. Y. R. Li, and R. Yeung. Network information flow. *IEEE Transactions on Information Theory*, IT-46(4):1204–1216, July 2000.

[8] T. Ho, R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, October 2006.

[9] M. Bakshi and M. Effros. On achievable rates for multicast in the presence of side information. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 1661–1665, July 2008.

[10] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-22(1):1–10, January 1976.

[11] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.:Academic, 1981.

[12] T. Berger. Multiterminal source coding. in: The information theory approach to communications. In *CISM Courses and Lecture Notes No. 229*, pages 171–231. Springer-Verlag, 1977.

[13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[14] S.-Y. Tung. Multiterminal source coding. Technical report, Ph.D. dissertation, School of Electrical Engineering, Cornell University, Ithaca, NY, 1978.

[15] A. Wagner, S. Tavildar, and P. Viswanath. Rate region of the quadratic gaussian two-encoder source-coding problem. *IEEE Transactions on Information Theory*, IT-54:1938–1961, May 2008.