Separation of Source-Network Coding and Channel Coding in Wireline Networks

Shirin Jalali
Center for Mathematics of Information
California Institute of Technology
Pasadena, California, 91125
Email: shirin@caltech.edu

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

***Abstract*—In this paper we prove the separation of source-network coding and channel coding in a wireline network, which is a network of memoryless point-to-point finite-alphabet channels used to transmit correlated sources. In deriving this result, we also prove that in a general memoryless network with correlated sources lossless and zero-distortion reconstruction are equivalent provided that the conditional entropy of each source given the other sources is non-zero. Furthermore, we extend the separation result to the case of well-behaved continuous-alaphabet point-to-point channels such as additive white Gaussian noise (AWGN) channels.**

I. Problem statement

Consider a network of memoryless noisy point-to-point channels with correlated sources and arbitrary demands. For achieving the optimal performance on such a network, i.e., achieving a point on the boundary of the set of achievable distortions, the nodes insides the network need to perform source coding, channel coding and network coding operations jointly. This makes both analyzing the performance of such networks and designing efficient codes for such networks very complicated. In this paper we prove that there is no loss in the asymptotic performance, by separating the source-network coding and channel coding operations. More precisely, we show that the set of achievable distortions in delivering a family of dependent sources across such a network $\mathcal{N}$ is equal to the set of achievable distortions for delivering the same sources across a distinct network $\mathcal{N}_b$. Network $\mathcal{N}_b$ is built by replacing each channel $p(y|x)$ in $\mathcal{N}$ by a noiseless, point-to-point bit-pipe of capacity $C = \max_{p(x)} I(X;Y)$.

Separation of source coding (lossy or lossless) and channel coding in a point-to-point communication network consisting of a transmitter and a receiver connected through a discrete memoryless channel (DMC) is a well-known result in the information theory literature [1]. The same result holds under a variety of source and channel distributions (see, for example [2] and the references therein). However, the separation does not necessarily hold in network systems. A classic example of this failure is due to Salehi, Cover and El Gamal on sending correlated sources over a multiple access channel [3]. In general, statistical dependencies between the sources at different network locations might be useful for increasing the rate across the channel. Since source codes tend to destroy such dependencies, joint source-channel codes can achieve better performance than separate source and channel codes in these scenarios.

The separation between network coding and channel coding in a wireline network with *independent* messages and lossless reconstructions have been proved for multicast networks in [4], [5] and for general demands in [6]. The separation of lossy source-network coding and channel coding in a wireline network with *correlated* sources has been proved independently in [7] and [8]. In this paper, we extend the result in [7] to the case where each reconstruction can be either lossless and lossy. Table I summarizes the setups under which separation has been proved in [6], [7], [8], and this paper.

TABLE I
Comparing the setups in [6], [7], [8] and this paper, under which separation is proved.

	correlated sources	lossless demands	lossy demands	continuous channels
[6]	no	yes	no	yes
[7], [8]	yes	no	yes	no
This work	yes	yes	yes	yes

The organization of this paper is as follows. Sections II and III describe the notation and problem setup, respectively. Section IV reviews the idea of constructing stacked networks that enable us to employ typicality across copies of a network rather than typicality across time. Section V proves the equivalence of zero-distortion reconstruction and lossless reconstruction in general multiuser memoryless networks. Section VI proves that separation of lossy source-network coding and channel coding holds in wireline networks with well-behaved continuous channels, such as AWGN, as well. Section VII concludes the paper.

II. Notation and definitions

Finite sets are denoted by script letters such as $\mathcal{X}$, $\mathcal{Y}$, etc. The size of a finite set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. For a random variable X, its alphabet set is represented by $\mathcal{X}$. Random variables are denoted by upper case letters such as X, Y, etc. Bold face letters represent vectors. A random vector is represented by upper case bold letters like $\mathbf{X}$, $\mathbf{Y}$, etc. The length of a vector is implied in the context, and its

ℓ^{th} element is denoted by X_ℓ. A vector $\mathbf{x} = (x_1, \ldots, x_n)$ ($\mathbf{X} = (X_1, \ldots, X_n)$) is sometimes represented as x^n (X^n). For $1 \leq i \leq j \leq n$, $x_i^j = (x_i, x_{i+1}, \ldots, x_j)$. For a set $\mathcal{A} \subseteq \{1, 2, \ldots, n\}$, $\mathbf{x}_\mathcal{A} = (x_i)_{i \in \mathcal{A}}$, where the elements are sorted in ascending order of their indices.

For two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^r$, $x \leq y$ iff $s_i \leq y_i$ for all $1 \leq i \leq r$ The ℓ_1 distance between the two vectors $\mathbf{x}$ and $\mathbf{y}$ of the same length r is denoted by $\|\mathbf{x} - \mathbf{y}\|_1 = \sum_{i=1}^r |x_i - y_i|$. If $\mathbf{x}$ and $\mathbf{y}$ represent pmfs, i.e., $\sum_{i=1}^r x_i = \sum_{i=1}^r y_i = 1$ and $x_i, y_i \geq 0$ for all $i \in \{1, \ldots, r\}$, then the total variation distance between $\mathbf{x}_1$ and $\mathbf{x}_2$ is defined as $\|\mathbf{x} - \mathbf{y}\|_{\text{TV}} = 0.5\|\mathbf{x} - \mathbf{y}\|_1$.

Definition 1: For a sequence $x^n \in \mathcal{X}^n$, the empirical distribution of x^n is defined as

$$\pi(x|x^n) \triangleq \frac{|\{i : x_i = x\}|}{n}, \tag{1}$$

for all $x \in \mathcal{X}$. Similarly, for $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$, the joint empirical distribution of (x^n, y^n) is defined as

$$\pi(x, y|x^n, y^n) \triangleq \frac{|\{i : (x_i, y_i) = (x, y)\}|}{n}, \tag{2}$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Definition 2: For a random variable $X \sim p(x)$, and $\epsilon > 0$, the set of ϵ-typical sequences of length n, $\mathcal{T}_\epsilon^{(n)}(X)$, is defined as [1]

$$\mathcal{T}_\epsilon^{(n)}(X) \triangleq \{x^n : |\pi(x|x^n) - p(x)| \leq \epsilon p(x) \text{ for all } x \in \mathcal{X}\}. \tag{3}$$

For $(X, Y) \sim p(x, y)$, the set of jointly ϵ-typical sequences is defined as

$$\mathcal{T}_\epsilon^{(n)}(X, Y) \triangleq \{(x^n, y^n) : |\pi(x, y|x^n, y^n) - p(x, y)| \leq \epsilon p(x, y), \text{ for all } (x, y) \in \mathcal{X} \times \mathcal{Y}\}. \tag{4}$$

We shall use $\mathcal{T}_\epsilon^{(n)}$ instead of $\mathcal{T}_\epsilon^{(n)}(X)$ or $\mathcal{T}_\epsilon^{(n)}(X, Y)$ when the random variable(s) clear from the context.
For $x^n \in \mathcal{T}_\epsilon^{(n)}$, let

$$\mathcal{T}_\epsilon^{(n)}(Y|x^n) \triangleq \{y^n : (x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}\}. \tag{5}$$

III. The setup

Consider a wireline network $\mathcal{N}$ consisting of m nodes interconnected via point-to-point, independent DMCs. The topology of the network is represented by a directed graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \ldots, m\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denote the set of nodes and the set of edges respectively. Each directed edge $e = [v_1, v_2] \in \mathcal{E}$ represents a point-to-point DMC between nodes v_1 (input) and v_2 (output). Each node a observes some source process $\mathbf{U}^{(a)} = \{U_k^{(a)}\}_{k=1}^\infty$, and is interested in reconstructing a subset of the processes observed by the other nodes. The alphabet of source $\mathbf{U}^{(a)}$, $\mathcal{U}^{(a)}$, can be either scalar or vector-valued. This allows node a to have a vector of sources. For achieving this goal in a block coding framework, source output symbols are divided into non-overlapping blocks of length L. Each block is described separately. At the beginning of the jth coding period, each node a has observed a length-L block of the process $\mathbf{U}^{(a)}$, i.e., $U_{(j-1)L+1}^{(a),jL} = (U_{(j-1)L+1}^{(a)}, \ldots, U_{jL}^{(a)})$. The blocks $\{U_{(j-1)L+1}^{(a),jL}\}_{a \in \mathcal{V}}$ observed at different nodes are described over the network in n uses of the network (The rate $\kappa \triangleq \frac{L}{n}$ is a parameter of the code). For those n time steps, at each step $t \in \{1, \ldots, n\}$, each node a generates its next channel inputs as a function of $U_{(j-1)L+1}^{(a),jL}$ and its channels' outputs up to time $t-1$, here denoted by $Y^{(a),t-1} = (Y_1^{(a)}, \ldots, Y_{t-1}^{(a)})$, according to

$$X_t^{(a)} : (\mathcal{Y}^{(a)})^{t-1} \times \mathcal{U}^{(a),L} \to \mathcal{X}^{(a)}. \tag{6}$$

Note that each node might be the input to more than one channel and/or the output of more than one channel. Hence, both $X_t^{(a)}$ and $Y_t^{(a)}$ might be vectors depending on the indegree and outdegree of node a. The reconstruction at node b of the block observed at node a is denoted by $\hat{U}^{(a \to b),L}$. This reconstruction is a function of the source observed at node b and node b's channel outputs, i.e., $\hat{U}^{(a \to b),L} = \hat{U}^{(a \to b)}(Y^{(b),n}, U^{(b),L})$, where

$$\hat{U}^{(a \to b)} : \mathcal{Y}^{(b),n} \times \mathcal{U}^{(a),L} \to \hat{\mathcal{U}}^{(a \to b),L}. \tag{7}$$

The performance criterion for a coding scheme is its induced expected average distortions between sources and reconstruction blocks, i.e., for all $a, b \in \mathcal{V}$

$$\mathrm{E}[d_L^{(a \to b)}(U^{(a),L}, \hat{U}^{(a \to b),L})] \triangleq \mathrm{E}[\frac{1}{L} \sum_{k=1}^L d^{(a \to b)}(U_k^{(a)}, \hat{U}_k^{(a \to b)})],$$

where $d^{(a \to b)} : \mathcal{U}^{(a)} \times \hat{\mathcal{U}}^{(a \to b)} \to \mathbb{R}^+$ is a per-letter distortion measure. As mentioned before $\mathcal{U}^{(a)}$ and $\hat{\mathcal{U}}^{(a \to b)}$ are either scalar or vector-valued. This allows the case where node a observes multiple sources and node b is interested in reconstructing a subset of them. Let

$$d_{\max} \triangleq \max_{a,b, \in \mathcal{V}, \alpha \in \mathcal{U}^{(a)}, \beta \in \hat{\mathcal{U}}^{(a \to b)}} d^{(a \to b)}(\alpha, \beta) < \infty.$$

If node b is not interested in reconstructing node a, then we simply let $d^{(a \to b)} \equiv 0$.

The distortion matrix $\mathbf{D}$ is said to be achievable at a rate κ in a network $\mathcal{N}$, if for any $\epsilon > 0$, there exists a pair (L, n), $L/n = \kappa$, and block length n coding scheme such that

$$\mathrm{E}[d_L^{(a \to b)}(U^{(a),L}, \hat{U}^{(a \to b),L})] \leq D(a, b) + \epsilon, \tag{8}$$

for any $(a, b) \in \mathcal{V} \times \mathcal{V}$.

Unlike [6], this paper uses strong typicality arguments to demonstrate the equivalence between noisy channels and noiseless bit-pipes of the same capacity. We first assume that the channel input and output alphabets are finite, and then extend the results to the case of continuous channels. The sources alphabets are always assumed to be discrete.

Throughout the paper, for a wireline network $\mathcal{N}$ represented by a directed graph $G = (\mathcal{V}, \mathcal{E})$, let $\mathcal{N}_b$ denote a network with identical topology such that each edge $e \in \mathcal{E}$ in $\mathcal{N}_b$ represents

[1] In this paper we only consider strong typicality, and use the definition introduced in [9].

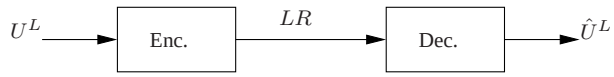

Fig. 1. Simple point-to-point channel

a bit-pipe of capacity C_e, where C_e is equal to the capacity of the point-to-point DMC corresponding to e in $\mathcal{N}$.

IV. Stacked network

For a given network $\mathcal{N}$, its N-fold stacked version $\underline{\mathcal{N}}$ is defined as N copies of the original network [6]. For each node (edge) in $\mathcal{N}$, there are N copies of the same node (edge) in $\underline{\mathcal{N}}$. A node in $\underline{\mathcal{N}}$ has access to the data available to all its copies, and together with its copies they performs the encoding and decoding operations. More precisely, in an N-fold stacked network,

$$\mathbf{X}_t^{(a)} : \mathcal{Y}^{(a),N(t-1)} \times \mathcal{U}^{(a),NL} \to \mathcal{X}^{(a),N}, \tag{9}$$

and

$$\hat{\mathbf{U}}^{(a\to b)} : \mathcal{Y}^{(b),nN} \times \mathcal{U}^{(b),NL} \to \hat{\mathcal{U}}^{(a\to b),NL}, \tag{10}$$

which correspond to (6) and (7) in the original network. Moreover, in such network, the distortion between the source observed at node a and its reconstruction at node b is defined as

$$D_N(a,b) = \mathrm{E}\left[d_{NL}^{(a\to b)}(U^{(a\to b),NL}, \hat{U}^{(a\to b),NL})\right], \tag{11}$$

for any $(a,b) \in \mathcal{V} \times \mathcal{V}$.

A distortion matrix D is said to be achievable in the stacked network at some rate κ if for any given $\epsilon > 0$, there exist N and n large enough, such that $D_N(a,b) \leq D(a,b) + \epsilon$, for all $(a,b) \in \mathcal{V} \times \mathcal{V}$. Note that the dimension of the distortion matrices in both single layer and multi-layer networks is $m \times m$. Let $\mathcal{D}(\kappa, \mathcal{N})$ and $\mathcal{D}_s(\kappa, \underline{\mathcal{N}})$ denote the closure of the set of achievable distortion matrices at some rate κ in a network $\mathcal{N}$ and its stacked version $\underline{\mathcal{N}}$ respectively. The following result from [7] shows that the two regions are equal.

Theorem 1 (Theorem 1 in [7]): At any rate κ,

$$\mathcal{D}(\kappa, \mathcal{N}) = \mathcal{D}_s(\kappa, \underline{\mathcal{N}}). \tag{12}$$

Employing Theorem 1 and the idea of channel simulation from [10], it is shown in [7] that in a wireline network with correlated sources and lossy reconstructions separation of source-network coding and channel coding is optimal.

Theorem 2 (Theorem 2 in [7]): At any rate κ,

$$\mathcal{D}(\kappa, \mathcal{N}) = \mathcal{D}(\kappa, \mathcal{N}_b). \tag{13}$$

In the following section we prove that the optimality of the separation of source-network coding and channel coding continues to hold in the case where each demand is either lossy or lossless.

V. Contiuity: zero-distortion versus lossless

Consider the simple point-to-point network shown in Fig. 1, where the source U is i.i.d. and distributed according to $p(u)$. In this simple network the minimum required rate for describing the source U at distortion D, $R(D)$, is known to be [11]

$$R(D) = \min_{p(\hat{u}|u):\mathrm{E}[d(U,\hat{U})]\leq D} I(U;\hat{U}).$$

Evaluating $R(D)$ at $D = 0$, it follows that

$$R(0) = \min_{p(\hat{u}|u):\mathrm{E}[d(U,\hat{U})]=0} I(U;\hat{U}) = I(U;U) = H(U), \tag{14}$$

where $H(U)$ is the entropy rate of the source U. On the other hand, it is known that the minimum required rate for lossless reconstruction of the source U is its entropy rate. Hence, in this simple setup, the zero-distortion and lossless reconstruction rate regions coincide. Explicit characterization of the multi-dimensional rate-distortion regions of general multiuser networks is unknown. Therefore, proving or disproving the equivalence of zero-distortion and lossless reconstruction rate-regions in such networks requires more elaborate analysis. W. Gu in his Ph.D. thesis proved that in noiseless networks consisting of point-to-point bit-pipies zero-distortion and lossless reconstruction rate-regions coincide [12].

In this section, we prove the equivalence of zero-distrotion reconstruction and lossless reconstruction in general multiuser discrete memoryless channels (DMC) with correlated sources. More precisely, we prove that in any DMC with correlated sources, achievability of zero-distortion reconstruction is equivalent to the achievability of lossless reconstruction. More precisely, if for a rate κ and $D \in \mathcal{D}(\kappa, \mathcal{N})$, $D(a,b) = 0$ for some $(a,b) \in \mathcal{V}^2$, then at the same rate, node b is able to reconstruct node a's data losslessly while keeping all the other reconstruction qualities unchanged.

Theorem 3: Let $D \in \mathcal{D}(\kappa, \mathcal{N})$. For any $a, b \in \mathcal{V}$ with $D(a,b) = 0$, if $H(U^{(a)}|(U^{(j)})_{j\neq a}) > 0$, zero-distortion reconstruction is equivalent to lossless reconstruction.

The proof of Theorem 3 is presented in Appendix A.

A direct implication of Theorem 3 is the extension of Theorem 2 to the case where each reconstruction can be either lossy or lossless.

VI. AWGN channels

So far the channels were all assumed to have discrete input and output alphabets. In this section, we prove that our results also hold for well-behaved continuous channels such as AWGN channels. In order to prove this we use the discretization method introduced in [13].

Consider a wireline network $\mathcal{N}$ with an AWGN channel from Node a to Node b with input X, output $Y = X + Z$, power constraint P and noise power N. Let $\mathcal{N}_b$ be a wireline network similar to the network $\mathcal{N}$, where the channel from a to b is replaced by a bit-pipe of capacity $C = 0.5\log(1 + P/N)$, but the rest of channels and sources are left intact. Theorem 4 shows, as in the case of discrete-valued channels, this change does not affect the set of achievable distortions.

Theorem 4: At any rate $\kappa > 0$,

$$\mathcal{D}(\kappa, \mathcal{N}) = \mathcal{D}(\kappa, \mathcal{N}_b). \tag{15}$$

Proof: Let $\mathcal{D} = \mathcal{D}(\kappa, \mathcal{N})$ and $\mathcal{D}_b = \mathcal{D}(\kappa, \mathcal{N}_b)$ denote the set of achievable distortion matrices at rate κ in networks $\mathcal{N}$ and $\mathcal{N}_b$ respectively. Following the same approach as in the first part of the proof of Theorem 2 in [7], we can show that $\mathcal{D}_b \subseteq \mathcal{D}$. Hence, in the rest of the proof we focus on the other direction showing that $\mathcal{D} \subseteq \mathcal{D}_b$ as well. To show this, as mentioned before, we employ the discretization method used in [13]. Let network $\mathcal{N}^{(\mathbf{j},\mathbf{k})}$, with $\mathbf{j} = (j_1, j_2, \ldots, j_n)$ and $\mathbf{k} = (k_1, k_2, \ldots, k_n)$, denote the network derived from network $\mathcal{N}$ by replacing its AWGN channel from a to b by the structure shown in Fig. 2, where at time $t \in \{1, 2, \ldots, n\}$, $j = j_t$ and $k = k_t$. Here $Q[i]$ denotes a quantizer with parameter i defined as follows. For $i \in \{1, 2, \ldots\}$, let $\Delta = 1/\sqrt{i}$, and define the quantizer $Q[i]$ with quantization levels $\mathcal{L}_i = \{-i\Delta, -(i-1)\Delta, \ldots, -\Delta, 0, \Delta, \ldots, (i-1)\Delta, i\Delta\}$. For $x \in \mathbb{R}$, $Q[i]$ maps x to $[x]_i$ which is the closest number to x in $\mathcal{L}_i$ such that $|[x]_i| \leq x$. Note that by this definition, if X is a random variable, $\mathrm{E}[[X]_i^2] \leq \mathrm{E}[X^2]$.

Lemma 1 in Appendix B shows that as the quantizations become finer, the set of achievable distortions on $\mathcal{N}^{(\mathbf{j},\mathbf{k})}$ becomes equal to the set of achievable distortions on the original network. More precisely

$$\overline{\limsup_{\mathbf{j},\mathbf{k}} \mathcal{D}(\kappa, \mathcal{N}^{(\mathbf{j},\mathbf{k})})} = \mathcal{D}(\kappa, \mathcal{N}), \tag{16}$$

where

$$\limsup_{\mathbf{j},\mathbf{k}} \mathcal{A}_{\mathbf{j},\mathbf{k}} \triangleq \bigcap_{\mathbf{j}_0,\mathbf{k}_0} \bigcup_{\substack{\mathbf{j}\geq\mathbf{j}_0 \\ \mathbf{k}\geq\mathbf{k}_0}} \mathcal{A}_{\mathbf{j},\mathbf{k}},$$

and $\overline{\mathcal{A}}$ denotes the closure of the set $\mathcal{A}$.

We next show that

$$\mathcal{D}(\kappa, \mathcal{N}^{(\mathbf{j},\mathbf{k})}) \subseteq \mathcal{D}_b. \tag{17}$$

This is sufficient to obtain the desired result since (16) and (17) together imply $\mathcal{D}(\kappa, \mathcal{N}) \subseteq \mathcal{D}_b$ by the closure in the definition of $\mathcal{D}_b$.

To prove that $\mathcal{D}(\kappa, \mathcal{N}^{(\mathbf{j},\mathbf{k})}) \subseteq \mathcal{D}_b$, note that, from the perspective of the network, the structure shown in Fig. 2 is, for each time t, equivalent to a DMC with input $[X]_{j_t}$ and output $[Y_{j_t}]_{k_t}$. Hence,

$$\mathcal{D}(\kappa, \mathcal{N}^{(\mathbf{j},\mathbf{k})}) \subseteq \mathcal{D}(\kappa, \mathcal{N}_b^{(\mathbf{j},\mathbf{k})}), \tag{18}$$

where $\mathcal{N}_b^{(\mathbf{j},\mathbf{k})}$ is identical to $\mathcal{N}^{(\mathbf{j},\mathbf{k})}$ except that the channel from a to b is replaced by a bit-pipe of capacity $C_{\mathbf{j},\mathbf{k}}$ equal to the maximum capacity of the n DMCs [7], i.e.,

$$C_{\mathbf{j},\mathbf{k}} \triangleq \max_{1\leq t\leq n} \max_{p_X : [X]_{j_t} \sim p_X} I([X]_{j_t}; [Y_{j_t}]_{k_t}).$$

By the data processing inequality [11],

$$\begin{aligned} I([X]_{j_t}; [Y_{j_t}]_{k_t}) &\leq I([X]_{j_t}; Y_{j_t}) \\ &= h(Y_j) - h(Z). \end{aligned} \tag{19}$$

On the other hand, by the construction of the quantizers,

$$\begin{aligned} \mathrm{E}[Y_{j_t}^2] &= \mathrm{E}[[X]_{j_t}^2] + N \\ &\leq \mathrm{E}[X^2] + N. \end{aligned} \tag{20}$$

Hence,

$$h(Y_{j_t}) \leq 0.5 \log(2\pi e(P + N)), \tag{21}$$

and as a result

$$I([X]_{j_t}; [Y_{j_t}]_{k_t}) \leq C. \tag{22}$$

Therefore, $\mathcal{D}(\kappa, \mathcal{N}^{(\mathbf{j},\mathbf{k})}) \subseteq \mathcal{D}_b$.

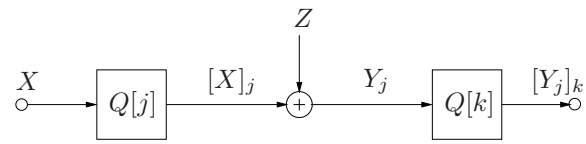

Fig. 2. Quantizing the input and output alphabets of an AWGN

■

VII. Conclusion

In this paper we proved separation of source-network and channel coding in general wireline networks of independent discrete point-to-point channels with correlated sources and arbitrary lossy or lossless reconstruction demands. We also proved that the same result continues to hold when each channel is either finite-alphabet or well-behaved continuous such as AWGN.

Acknowledgments

This work was supported in part by Caltech's Center for the Mathematics of Information (CMI) and DARPA ITMANET grant W911NF-07-1-0029.

Appendix A: Proof of Theorem 3

First assume that lossless reconstruction of source a at node b is achievable, i.e., there exist a family of codes at rate $\kappa = n/L$ such that

$$\mathrm{P}(U^{(a),L} \neq \hat{U}^{(a\to b),L})) \to 0, \tag{A-1}$$

as $L \to \infty$. Let $\mathcal{E} = \{U^{(a),L} \neq \hat{U}^{(a\to b),L})\}$. Hence, there exists a family for which $\mathrm{P}(\mathcal{E}) \to 0$, as $L \to \infty$. Now note that

$$\begin{aligned} &\mathrm{E}[d(U^{(a),L}, \hat{U}^{(a\to b),L})] \\ &\quad = \mathrm{E}[d(U^{(a),L}, \hat{U}^{(a\to b),L})|\mathcal{E}]\,\mathrm{P}(\mathcal{E}) \\ &\quad + \mathrm{E}[d(U^{(a),L}, \hat{U}^{(a\to b),L})|\mathcal{E}^c]\,\mathrm{P}(\mathcal{E}^c) \\ &\quad \leq d_{\max}\,\mathrm{P}(\mathcal{E}), \end{aligned} \tag{A-2}$$

since $\mathrm{E}[d(U^{(a),L}, \hat{U}^{(a\to b),L})|\mathcal{E}^c] = 0$. Hence, the same family of codes achieves zero-distortion reconstruction of source a at node b as well. Therefore we only need to prove the other direction.

For a given $D \in \mathcal{D}(\kappa, \mathcal{N})$, and any $\epsilon > 0$, by definition, there exists L (implying $n = \lceil \kappa L \rceil$), and encoding and decoding functions such that

$$\mathrm{E}[d(U^{(v_1),L}, \hat{U}^{(v_1 \to v_2),L})] \leq D(v_1, v_2) + \epsilon, \tag{A-3}$$

for any $(v_1, v_2) \in \mathcal{V}^2$. By assumption $D(a, b) = 0$. Therefore, for L sufficiently large,

$$\mathrm{E}[d(U^{(a),L}, \hat{U}^{(a \to b),L})] \leq \epsilon. \tag{A-4}$$

In the rest of the proof, for the ease of notation, we drop superscripts (a) and $(a \to b)$. For instance, $U^L = U^{(a),L}$ and $\hat{U}^L = \hat{U}^{(a \to b),L}$.

We now prove that with an asymptotically negligible increase in the rate κ, Node a can send Node b sufficient information to improve Node b's reconstruction of Node a's data from a zero-distortion reproduction to a lossless reconstruction. We further show that this change preserves the quality of all other reconstructions.

Assume that each node $v \in \mathcal{V}$ first observes a source block of length NL and breaks it into N non-overlapping blocks as

$$U^{(v),L}, U^{(v),2L}_{L+1}, \ldots, U^{(v),NL}_{(N-1)L+1}.$$

We use the encoding and decoding functions that achieve (A-3) N times to independently code each of these blocks. In total, this requires Nn channel uses, and achieves for each $v_1, v_2 \in \mathcal{V}$, a reconstruction of length NL such that

$$\mathrm{E}[d(U^{(v_1),jL}_{(j-1)L+1}, \hat{U}^{(v_1 \to v_2),jL}_{(j-1)L+1})] \leq D(v_1, v_2) + \epsilon, \tag{A-5}$$

for each $j = 1, 2, \ldots, N$. For each $\ell \in \{1, \ldots, N\}$, denote the input of Node a in session ℓ as $U^L(\ell) = U^{\ell L}_{(\ell-1)L+1}$, and the corresponding output at Node b as $\hat{U}^L(\ell) = \hat{U}^{\ell L}_{(\ell-1)L+1}$. By assumption,

$$\mathrm{E}[d(U^L(\ell), \hat{U}^L(\ell))] \leq \epsilon, \tag{A-6}$$

for all $\ell \in \{1, 2, \ldots, L\}$. Note that for any random vectors $U^L \in \mathcal{U}^L$ and $\hat{U}^L \in \hat{\mathcal{U}}^L$

$$\begin{aligned}\mathrm{E}[d(U^L, \hat{U}^L)] &= \frac{1}{L}\sum_{i=1}^{L} \mathrm{E}[d(U_i, \hat{U}_i)] \\ &\geq d_{\min} \frac{1}{L}\sum_{i=1}^{L} \mathrm{P}(U_i \neq \hat{U}_i).\end{aligned} \tag{A-7}$$

Therefore, since $\mathrm{E}[d(U^L(\ell), \hat{U}^L(\ell))] \leq \epsilon$,

$$\frac{1}{L}\sum_{i=1}^{L} \mathrm{P}(U_i(\ell) \neq \hat{U}_i(\ell)) \leq \frac{\epsilon}{d_{\min}} \tag{A-8}$$

for all $\ell \in \{1, 2, \ldots, N\}$, where $d_{\min} \triangleq \min_{u \in \mathcal{U}, \hat{u} \in \hat{\mathcal{U}}} d(u, \hat{u}) > 0$, since all alphabets are assumed to be discrete.

Since the sources and the channels are memoryless, $\{U^L(\ell), \hat{U}^L(\ell)\}_{\ell=1}^{L}$ is an i.i.d. sequence. (See Fig. 3.) Now consider the problem of lossless source coding with side information, as shown in Fig. 4. From [14], an extra rate of $R_0 = H(U^L|\hat{U}^L)$ suffices for losslessly reconstructing U^L at a receiver that knows $\hat{U}^L$. (Here lossless coding means that $\mathrm{P}(U^{LN} \neq \tilde{U}^{LN})$ can be made arbitrary small.) Using Fano's inequality [11] and the concavity of the entropy function,

$$\begin{aligned}R_0 = H(U^L|\hat{U}^L) &= \sum_{i=1}^{L} H(U_i|U^{i-1}, \hat{U}^L) \\ &\leq \sum_{i=1}^{L} [h(\mathrm{P}(U_i \neq \hat{U}_i)) + |\mathcal{U}|\, \mathrm{P}(U_i \neq \hat{U}_i)] \\ &\leq Lh(\frac{1}{L}\sum_{i=1}^{L} \mathrm{P}(U_i \neq \hat{U}_i)) + |\mathcal{U}| \sum_{i=1}^{L} \mathrm{P}(U_i \neq \hat{U}_i)] \\ &\leq L(h(\frac{\epsilon}{d_{\min}}) + \frac{|\mathcal{U}|\epsilon}{d_{\min}}) \\ &\triangleq Lf(\epsilon),\end{aligned} \tag{A-9}$$

where for $0 \leq p \leq 1$, $h(p) = -p \log p - (1-p)\log(1-p)$, and $f(\epsilon) = h(\frac{\epsilon}{d_{\min}}) + \frac{|\mathcal{U}|\epsilon}{d_{\min}}$, and $f(\epsilon) \to 0$ as $\epsilon \to 0$.

We send the rate-R_0 description of U^L across the communication channel between U^L and $\hat{U}^L$ that is implied by all the encoding and decoding functions. In the following, we show the existence of this channel and bound its capacity.

Since $d(U^L, \hat{U}^L) \geq 0$,

$$\begin{aligned}\epsilon &\geq \mathrm{E}[d(U^L, \hat{U}^L)] \\ &\geq \mathrm{E}[d(U^L, \hat{U}^L)|(U^{(j),L})_{j \neq a} \in \mathcal{T}_\delta^{(L)}]\, \mathrm{P}((U^{(j),L})_{j \neq a} \in \mathcal{T}_\delta^{(L)}).\end{aligned} \tag{A-10}$$

For L large enough, $\mathrm{P}((U^{(j),L})_{j \neq a} \in \mathcal{T}_\delta^{(L)}) > 1 - \delta$ which implies that

$$\mathrm{E}[d(U^L, \hat{U}^L)|(U^{(j),L})_{j \neq a} \in \mathcal{T}_\delta^{(L)}] \leq \frac{\epsilon}{1-\delta}. \tag{A-11}$$

Hence, there exists $(u^{(j),L})_{j \neq a} \in \mathcal{T}_\delta^{(L)}$ such that

$$\mathrm{E}[d(U^L, \hat{U}^L)|(U^{(j),L})_{j \neq a} = (u^{(j),L})_{j \neq a}] \leq \frac{\epsilon}{1-\delta}. \tag{A-12}$$

Following steps similar to those in (A-7) and (A-9) but here conditioning on $(U^{(j),L})_{j \neq a} = (u^{(j),L})_{j \neq a}$, we conclude that

$$H(U^L|\hat{U}^L, (U^{(j),L})_{j \neq a} = (u^{(j),L})_{j \neq a}) \leq Lf(\frac{\epsilon}{1-\delta}). \tag{A-13}$$

On the other hand, since $(u^{(j),L})_{j \neq a} \in \mathcal{T}_\delta^{(L)}$, for any $u^L \in \mathcal{T}_\delta^{(L)}(U|(u^{(j),L})_{j \neq a})$, [15]

$$p(u^L|(u^{(j),L})_{j \neq a}) \leq 2^{-(1-\delta)LH(U|(U^{(j)})_{j \neq a})}. \tag{A-14}$$

Hence, for L large enough,

$$\begin{aligned}
&H(U^L|(U^{(j),L})_{j\neq a}=(u^{(j),L})_{j\neq a})\\
&=\sum_{u^L} p(u^L|(u^{(j),L})_{j\neq a})\log p(u^L|(u^{(j),L})_{j\neq a})\\
&>\sum_{u^L\in\mathcal{T}_\delta(U|(u^{(j),L})_{j\neq a})} -p(u^L|(u^{(j),L})_{j\neq a})\log p(u^L|(u^{(j),L})_{j\neq a})\\
&\geq (1-\delta)LH(U|(U^{(j)})_{j\neq a})\,\mathrm{P}(U^L\in\mathcal{T}_\delta^{(L)}|(u^{(j),L})_{j\neq a}))\\
&\geq (1-\delta)^2LH(U|(U^{(j)})_{j\neq a}). \qquad \text{(A-15)}
\end{aligned}$$

Hence, fixing the input block of any node $j\in\mathcal{V}\backslash\{a\}$ to $u^{(j),L}$ determined by (A-12), and using a random codebook generated according to

$$p(u^L|(u^{(j),L})_{j\neq a}),$$

we can send data from node a to node b at any rate $R\leq C_0$, where

$$C_0\triangleq(1-\delta)^2LH(U|(U^{(j)})_{j\neq a})-Lf(\frac{\epsilon}{1-\delta}). \qquad \text{(A-16)}$$

Thus the rate required to losslessly describe U^L to a decoder with zero-distortion reproduction $\hat{U}^{LN}$ is at most R_0N, and the capacity per use of the given block length-L code with $(U^{(j),L})_{j\neq a}=(u^{(j),L})_{j\neq a})$ is at least C_0 bits per L network uses. We can therefore achieve the desired lossless description over a total of $(N+NR_0)/C_0=N(1+R_0/C_0)$ sessions, where

$$\begin{aligned}
\frac{R_0}{C_0}&=\frac{Lf(\epsilon)}{(1-\delta)^2LH(U|(U^{(j)})_{j\neq a})-Lf(\frac{\epsilon}{1-\delta})}\\
&=\frac{f(\epsilon)}{(1-\delta)^2H(U|(U^{(j)})_{j\neq a})-f(\frac{\epsilon}{1-\delta})}, \qquad \text{(A-17)}
\end{aligned}$$

approaches zero as ϵ approaches zero and δ approaches zero. The resulting coding rate is

$$\kappa'=\frac{\kappa}{1+R_0/C_0}, \qquad \text{(A-18)}$$

which approaches κ as ϵ approaches zero and δ approaches zero.

Appendix B: Lemma 1

Lemma 1: For any $\kappa>0$,

$$\overline{\limsup_{\mathbf{j},\mathbf{k}}\mathcal{D}(\kappa,\mathcal{N}^{(\mathbf{j},\mathbf{k})})}=\mathcal{D}(\kappa,\mathcal{N}), \qquad \text{(B-1)}$$

where $\overline{\mathcal{A}}$ denotes the closure of set $\mathcal{A}$.

Proof: Any performance achievable on $\mathcal{N}^{(\mathbf{j},\mathbf{k})}$ can be achieved on $\mathcal{N}$ by adding the proper quantizers to the input and output of the channel from a to b in $\mathcal{N}$. Hence,

$$\overline{\limsup_{\mathbf{j},\mathbf{k}}\mathcal{D}(\kappa,\mathcal{N}^{(\mathbf{j},\mathbf{k})})}\subseteq\mathcal{D}(\kappa,\mathcal{N}).$$

The nontrivial step is proving the other direction. Let $D\in\mathcal{D}(\kappa,\mathcal{N})$. For any $\epsilon>0$, and for L sufficiently large, there exist encoding and decoding schemes operating at rate κ with block length L such that

$$\mathrm{E}[d(U^{(v_1),L},\hat{U}^{(v_1\to v_2),L})]\leq D(v_1,v_2)+\epsilon, \qquad \text{(B-2)}$$

holds for any $(v_1,v_2)\in\mathcal{V}^2$. Let $U^L=U^{(v_1),L}$ and $\hat{U}^L=\hat{U}^{(v_1\to v_2),L}$ for some $(v_1,v_2)\in\mathcal{V}^2$.

Conditioning the expected average distortion between U^L and $\hat{U}^L$ on the input and output values of the AWGN channel at time $t=1$, it follows that

$$\begin{aligned}
\mathrm{E}[d(U^L,\hat{U}^L)]&=\mathrm{E}[\mathrm{E}[d(U^L,\hat{U}^L)|(X_1,Y_1)]]\\
&=\mathrm{E}[\delta^{(1)}(X_1,Y_1)]\\
&\leq D(v_1,v_2)+\epsilon, \qquad \text{(B-3)}
\end{aligned}$$

where $\delta^{(1)}(x_1,y_1)\triangleq\mathrm{E}[d(U^L,\hat{U}^L)|(X_1,Y_1)=(x_1,y_1)]$.

Now assume that the same code is applied to network $\mathcal{N}^{(j_1,k_1)}$, which is identical to $\mathcal{N}$ except for time $t=1$, where the AWGN channel is replaced by the structure shown in Fig. 2 with parameters $j=j_1$ and $k=k_1$. The expected average distortion between U^L and $\hat{U}^L$ in the modified network, $D^{(j_1,k_1)}(v_1,v_2)$, can be written as

$$\begin{aligned}
D^{(j_1,k_1)}(v_1,v_2)&=\mathrm{E}[\mathrm{E}[d(U^L,\hat{U}^L)]|(X_1,\tilde{Y}_1)]\\
&=\mathrm{E}[\delta^{(1)}(X_1,\tilde{Y}_1)], \qquad \text{(B-4)}
\end{aligned}$$

where $\tilde{Y}_1\triangleq[[X]_{j_1}+Z_1]_{k_1}$. Note that conditioned on the input and output values of the AWGN channel at time $t=1$, the two networks have identical performance.

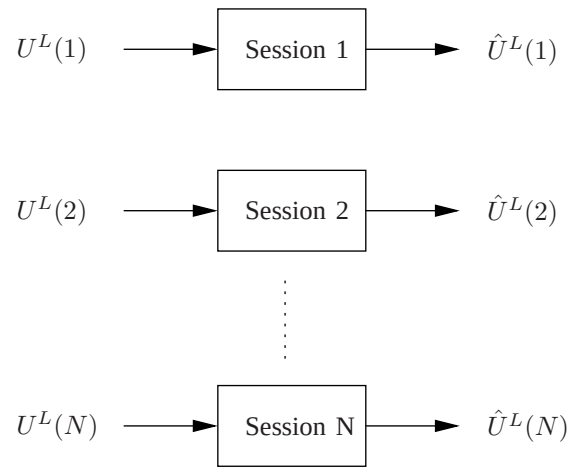

Fig. 3. Source U and its reconstruction at L parallel sessions.

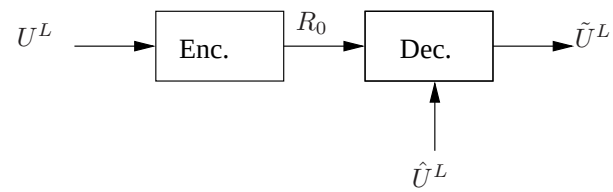

Fig. 4. Slepian-Wolf coding for converting zero-distortion reconstruction into lossless reconstruction.

On the other hand, $\tilde{Y}_1$ is converging pointwise to Y_1, i.e.,

$$\lim_{k_1\to\infty}\lim_{j_1\to\infty}\tilde{Y}_1 = Y_1, \tag{B-5}$$

where $Y_1 = X+Z$. Therefore, since $0 \le \delta^{(1)}(x_1,y_1) \le d_{\max}$, assuming that the set of discontinuity points of $\delta^{(1)}(x_1,y_1)$ is countable, by dominated convergence theorem [16], it follows that

$$\begin{aligned}\lim_{k_1\to\infty}\lim_{j_1\to\infty} D^{(j,k)}(v_1,v_2) &= \lim_{k_1\to\infty}\lim_{j_1\to\infty}\mathrm{E}[\delta^{(1)}(X_1,\tilde{Y}_1)]\\ &= \mathrm{E}[\delta^{(1)}(X_1,Y_1)]. \end{aligned}\tag{B-6}$$

The next step is to add the input and output quantizers both at time $t=1$ and $t=2$. Assume that the quantizers parameters are (j_1,j_2,k_1,k_2), and let $D^{(j^2,k^2)}(v_1,v_2)$ denote the corresponding performance. Define $\delta^{(2)}(x^2,y^2) = \mathrm{E}[d(U^L,\hat{U}^L)|(X^2,Y^2)=(x^2,y^2)]$. In the original network

$$\begin{aligned}D(v_1,v_2) &= \mathrm{E}[\delta^{(2)}(X^2,Y^2)]\\ &= \mathrm{E}[\mathrm{E}[\delta^{(2)}(X^2,Y^2)]|(X_1,Y_1)]], \end{aligned}\tag{B-7}$$

and in the modified network,

$$\begin{aligned}D^{(j^2,k^2)}(v_1,v_2) &= \mathrm{E}[\delta^{(2)}(X_1,\tilde{X}_2,\tilde{Y}_1,\tilde{Y}_2)],\\ &= \mathrm{E}[\mathrm{E}[\delta^{(2)}(X_1,\tilde{X}_2,\tilde{Y}^2)]|(X_1,\tilde{Y}_1)]], \end{aligned}\tag{B-8}$$

where $\tilde{Y}_1 = [[X_1]_{j_1} + Z_1]_{k_1}$ and $\tilde{Y}_2 = [[\tilde{X}_2]_{j_2} + Z_2]_{k_2}$. Note that while X_2 and $\tilde{X}_2$ might have different distributions due to the quantizations at time $t=1$, their conditional distributions given the input and output of the channel are identical in both networks, i.e.,

$$\begin{aligned}&\mathrm{P}(\tilde{X}_2 < x_2|(X_1,\tilde{Y}_1)=(x_1,y_1))\\ &= \mathrm{P}(X_2 < x_2|(X_1,Y_1)=(x_1,y_1)). \end{aligned}\tag{B-9}$$

Let

$$\begin{aligned}\gamma(x_1,y_1) &\triangleq \mathrm{E}[\delta^{(2)}(X^2,Y^2)]|(X_1,Y_1)=(x_1,y_1)]\\ &\mathrm{E}\left[\int \delta^{(2)}(x_1,x_2,y_1,Y_2)dF(x_2|(x_1,y_1))\right] \end{aligned}\tag{B-10}$$

and

$$\begin{aligned}\tilde{\gamma}^{(j_2,k_2)}(x_1,y_1) &\triangleq \mathrm{E}[\delta^{(2)}(\tilde{X}^2,\tilde{Y}^2)]|(X_1,\tilde{Y}_1)=(x_1,y_1)]\\ &= \mathrm{E}\left[\int \delta^{(j_2,k_2)}(x_1,x_2,y_1,\tilde{Y}_2)dF(x_2|(x_1,y_1))\right]. \end{aligned}\tag{B-11}$$

Since $\delta^{(2)}$ is a positive bounded function, assuming that the measure of its discontinuity points is zero, by the dominated convergence theorem,

$$\lim_{k_2\to\infty}\lim_{j_2\to\infty}\tilde{\gamma}^{(j_2,k_2)}(x_1,y_1) = \gamma(x_1,y_1). \tag{B-12}$$

Hence

$$\lim_{k_2\to\infty}\lim_{j_2\to\infty}\tilde{\gamma}^{(j_2,k_2)}(x_1,y_1) = \gamma(x_1,y_1). \tag{B-13}$$

Hence,

$$\begin{aligned}&\lim_{k_1\to\infty}\lim_{j_1\to\infty}\lim_{k_2\to\infty}\lim_{j_2\to\infty} D^{(j^2,k^2)}(v_1,v_2)\\ &= \lim_{k_1\to\infty}\lim_{j_1\to\infty}\lim_{k_2\to\infty}\lim_{j_2\to\infty}\mathrm{E}[\tilde{\gamma}^{(j_2,k_2)}(X_1,\tilde{Y}_1)]\\ &= \lim_{k_1\to\infty}\lim_{j_1\to\infty}\mathrm{E}[\lim_{k_2\to\infty}\lim_{j_2\to\infty}\tilde{\gamma}^{(j_2,k_2)}(X_1,\tilde{Y}_1)]\\ &= \lim_{k_1\to\infty}\lim_{j_1\to\infty}[\mathrm{E}\,\gamma(X_1,\tilde{Y}_1)]\\ &= \mathrm{E}[\lim_{k_1\to\infty}\lim_{j_1\to\infty}\gamma(X_1,\tilde{Y}_1)]\\ &= \mathrm{E}[\gamma(X_1,Y_1)]\\ &= D(v_1,v_2). \end{aligned}\tag{B-14}$$

The same procedure can be extended up to time $t=n$. ■

REFERENCES

[1] C. E. Shannon. A mathematical theory of communication: Parts I and II. *Bell Syst. Tech. J.*, 27:379–423 and 623–656, 1948.

[2] S. Vembu, S. Verdú, and Y. Steinberg. The source-channel separation theorem revisited. *IEEE Trans. Inform. Theory*, 41(1):44–54, January 1995.

[3] T. Cover, A.E. Gamal, and M. Salehi. Multiple access channels with arbitrarily correlated sources. *Information Theory, IEEE Transactions on*, 26(6):648 – 657, November 1980.

[4] S.P. Borade. Network information flow: limits and achievability. In *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, 2002.

[5] L. Song, R.W. Yeung, and N. Cai. A separation theorem for single-source network coding. *IEEE Trans. Inform. Theory*, 52(5):1861 –1871, May 2006.

[6] R. Koetter, M. Effros, and M. Médard. On the theory of network equivalence,. In *IEEE Inform. Theory Workshop (ITW)*, 2009.

[7] S. Jalali and M. Effros. On the separation of lossy source-network coding and channel coding in wireline networks. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 500 –504, June 2010.

[8] Chao Tian, Jun Chen, S.N. Diggavi, and S. Shamai. Optimality and approximate optimality of source-channel separation in networks. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 495–499, June 2010.

[9] A. Orlitsky and J.R. Roche. Coding for computing. *Information Theory, IEEE Transactions on*, 47(3):903 –917, March 2001.

[10] P.W. Cuff, H.H. Permuter, and T.M. Cover. Coordination capacity. *Information Theory, IEEE Transactions on*, 56(9):4181 –4206, Sept. 2010.

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

[12] W. Gu. *Achievable Rate Regions for Source Coding over Networks*. PhD thesis, California Institute of Technology, Pasadena, CA, 2008.

[13] R. J. McEliece. *The Theory of Information and Coding*. Addison-Wesley, Reading, MA, 1977.

[14] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory*, 19(4):471–480, 1973.

[15] A. El Gamal and Y. Kim. Lecture notes on information theory. *arXiv:1001.3404v4*.

[16] R. Durrett. *Probability: Theory and Examples*. Duxbury Press, Belmont, CA, 1996.