On Source-Channel Separation in Networks

Chao Tian*, Jun Chen†, Suhas N. Diggavi‡, and Shlomo Shamai§
*AT&T Labs-Research, Florham Park, NJ 07932, USA
†McMaster University, Hamilton, ON L8S 4K1, Canada
‡UCLA, Los Angeles, USA and EPFL, Switzerland
§Technion-Israel Institute of Technology, Haifa 32000, Israel

***Abstract*—We consider the source-channel separation architecture for lossy source coding in communication networks. It is shown that the separation approach is optimal when the memoryless sources at source nodes are arbitrarily correlated, each of which is to be reconstructed at possibly multiple destinations within certain distortions, and the channels in this network are synchronized, orthogonal and memoryless (i.e., noisy graphs). The extracted pure network source-coding problem has to incorporate user interactions and the corresponding causality constraints, which suggests a distinct research direction into interactive network source coding that has not received much attention in the literature. The separation result in this paper is a companion to results of a similar flavor established recently in [11] for the case of independent sources over communication networks with more general multiuser channels.**

I. Introduction

One of the fundamental observations in information theory is the Shannon's source-channel separation theorem [10]. The separation theorem is architecturally important because it demonstrates that in point-to-point communication systems the source coding and channel coding components can be designed separately essentially without loss in overall performance. Unfortunately, it has been shown that separation does hold in general multiuser networks. Given the architectural benefits of separation, it is important to understand network circumstances where it still retains optimality. In this paper we focus on this question for the special case of transmitting correlated sources over a "noisy" graph (a network of orthogonal memoryless channels). We refer to this as the *distributed network source coding* problem.

The difficulty in answering the separation question for the distributed network source coding problem, lies in the fact that we do not have explicit characterizations of the rate-distortion regions or the joint coding achievable distortion regions. Since the natural comparison between separation and joint coding is not a fruitful direction, we establish the answer to the separation question, without specifically solving the individual or joint coding problem.

In addition to source-channel separation, separation between network coding and channel coding, has received considerable attention in recent years [1]–[4]. In general, the approach based on separating network coding and channel coding is also not optimal, as shown by an example of deterministic broadcast channels with no interference [3]. However, a surprising result by Koetter, Effros and Medard [4] (see also [5]) essentially states that for general multicast on orthogonal, synchronized and memoryless channels (*i.e.,* "noisy" graphs), there is no loss of optimality by employing such a separation. This result does not rely on an explicit characterization of the general multicast network coding capacity region, which is well known to be equivalent to the extremely difficult problem of characterizing the entropy space [6].

The distributed network source coding can be thought of as a generalization of the problem studied by Koetter *et al* [4], [5]. The main difference between our work and [4], [5] is the following: (i) the sources in our setting are correlated instead of being independent messages; (ii) we allow for lossy delivery of the sources to the destinations. In this work, we adapt the technique developed in [4] to our distributed network source coding setting. The proof, in its simplest form, relies on a simulation of each orthogonal channel by an appropriately defined "rate-distortion" code. The example in Section II illustrates this idea for a simple network.

Though our results demonstrate the effectiveness of source-channel separation, a large portion of the difficulty in designing efficient codes remains in the extracted pure source coding problem. In particular, the network induces a pure source-coding problem with rather complex user interactions. The importance of such kind of interactive source coding in networks suggests a distinct and perhaps under-studied line of research in network source coding.

The distributed network source coding problem includes the problem of successive refinement source coding with degraded decoder side informations on orthogonal communication channels considered in [7], [8], and the problem of distributed source coding on orthogonal multiple access channel considered in [9].

The rest of this paper is organized as follows. In Section II we discuss an example to provide some intuitions for the solution. Necessary notation and definitions are given in Section III. The main result and the proof are given in Sections IV. Section V concludes the paper.

II. An Example

In this section we discuss an example that illustrates the intuitions of the main result established in and Sections IV. For simplicity, we do not attempt to keep the notation used in this section consistent with that given in III, which is more

The work of Shlomo Shamai has been supported by the FP7 Network of Excellence in Wireless COMmunications NEWCOM++.

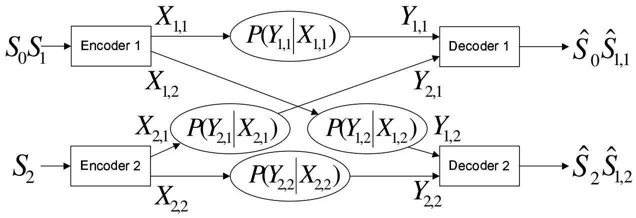

Fig. 1. Transmitting correlated sources on orthogonal interference networks.

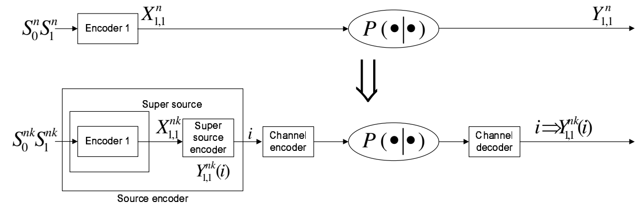

Fig. 2. Converting a joint source-channel code into a separation-based code on an individual orthogonal channel.

suitable for complex networks; we also assume the channel bandwidth and source bandwidth are matched for simplicity.

Consider the example depicted in Figure 1, where the discrete memoryless sources S_0, S_1, S_2 are correlated, and each discrete memoryless channel between a transmitter and a receiver is orthogonal to the other channels. Though the capacity region of this orthogonal interference channel is not difficult to establish, the rate-distortion region of the source coding problem is not known, which is at least as difficult as the well known lossy distributed source coding problem. Thus the conventional proof approach of characterizing separately the rate-distortion region, channel capacity region and the joint source-channel coding achievable distortion region, and then making comparison, does not lead to a meaningful result. Next, we illustrate through this example the methodology that enables us to prove the optimality of source-channel separation for the distributed network source coding problem.

Suppose there exists a length-n joint source-channel code that achieves the distortion quadruple $(D_0, D_{1,1}, D_{1,2}, D_2)$. The key observation is the following simple fact. If we fix this joint source-channel code, then the channel input for any given channel, for example $X_{1,1}^n$, can be viewed as a super (block) source, independent and identically distributed across blocks; see Fig. 2. Therefore, we can encode a length-k sequence of such blocks using a "rate-distortion" code of rate per block slightly greater than $I(X_{1,1}^n; Y_{1,1}^n)$, the codewords of which are generated using the distribution $Y_{1,1}^n$. It follows that with high probability we can find a $Y_{1,1}^{nk}$ codeword in the codebook that is jointly typical with a channel input sequence $X_{1,1}^{nk}$, (*i.e.*, a length-k vector of the super source samples), for sufficiently large k. This rate-distortion code essentially simulates the channel output over k length-n blocks, and we might as well send this digital rate-distortion codeword index across this channel using any good channel code, and perform the original joint source-channel decoding function on this simulated channel output, which eventually achieves the (asymptotically) same distortion as the original code. It is easy to see that since $I(X_{1,1}^n; Y_{1,1}^n) \leq nC_{1,1}$, where $C_{1,1}$ is the channel capacity of the channel between transmitter 1 and receiver 1, we expect that the channel should be able to deliver this rate-distortion codeword index reliably. Replacing all the channel outputs with such simulated outputs in this problem results in a new scheme. In this new coding scheme, the codeword indices of these "rate-distortion" codes are the only informational interface between the source coding component and the channel coding component, and this is a separation-based scheme which asymptotically achieves the same distortions $(D_0, D_{1,2}, D_{1,2}, D_2)$ originally achieved by the joint coding scheme. In other words, any distortions that are achievable by joint coding scheme can be achieved by a separation-based scheme.

The above observation largely reflects the intuition behind the the optimality proof of source-channel separation for the general distributed network source coding problem, however, some technical details (besides the asymptotically diminishing quantities omitted in the above discussion) need to be addressed: the main difficulty is that when the network has relays or cycles, the super source argument given above does not apply directly since we cannot wait for many long blocks due to causality and channel usage constraints. The proof given Section IV will resolve this difficulty through an intricate arrangement of the channel simulation.

III. Notation and Definitions

The network with a total of N nodes can be conveniently written as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\} \triangleq \mathcal{I}_N$ is the set of nodes, and $\mathcal{E}$ is the set of edges between nodes. Each edge $e = (i,j) \in \mathcal{E}$ is associated with a discrete memoryless channel, whose transition probability is given as $P(Y_{i,j}|X_{i,j})$ with input alphabet $\mathcal{X}_{i,j}$ and output alphabet $\mathcal{Y}_{i,j}$; these channels are assumed to be synchronized. Each node i has a discrete memoryless source S_i, distributed in the alphabet $\mathcal{S}_i$, and the collection of the sources are distributed according to the joint distribution $P(S_1, S_2, \ldots, S_M)$ at each time instance. For simplicity, we are inherently assuming these sources are synchronized (and thus the notation $P(S_1, S_2, \ldots, S_M)$ is meaningful), though this requirement can be relaxed to some extent at the expense of more complex notation. A length-n vector of a source S_i is written as S_i^n; we use upper case for random variables, and lower case for their realizations.

For each source, a distortion measure is defined in a general manner as $d : \mathcal{S}_i \times \hat{\mathcal{S}}_i \to [0, \infty)$ where $\hat{\mathcal{S}}_i$ is the reconstruction alphabet; for source vectors, the distortion is generalized to be the average of the single letter form in the usual way. Different distortion measures can in fact be used for multiple reconstructions of the same source, or even distortions defined on functions of several sources, without any essential difference, however for notational simplicity we do not consider such cases.

A node j may be interested in only a subset of the sources $\{S_i, i \in \mathcal{I}_N\}$; notationally, we may write the set of sources that the node j is interested in as $\mathscr{T}_j$. Next we define the class of codes being considered for the distributed network source coding problem, which are conventional block codes.

Definition 1: An $(m, n, \{d_{j,k}, k \in \mathscr{T}_j\})$ distributed network source code consists the following components:

- At each transmitter node j, for each k such that $(j,k) \in \mathcal{E}$, an encoding function for time instance t
$$\phi_{j,k}^{(t)} : \mathcal{S}_j^m \times \prod_{(i,j)\in\mathcal{E}} \mathcal{Y}_{i,j}^{t-1} \to \mathcal{X}_{j,k}, \quad t = 1, 2, \ldots, n. \quad (1)$$
- At each receiver node j, for each source $k \in \mathscr{T}_j$, a decoding function
$$\psi_{j,k} : \prod_{(i,j)\in\mathcal{E}} \mathcal{Y}_{i,j}^n \times \mathcal{S}_j^m \to \hat{\mathcal{S}}_k^m. \quad (2)$$

The encoding and decoding functions induce the distortions
$$d_{k,j} = \frac{1}{m}\sum_{t=1}^{m} d(S_k(t), \hat{S}_{k,j}(t)),$$
$$j = 1, 2, \ldots, M, \quad \text{and} \quad k \in \mathscr{T}_j,$$
where $\hat{S}_{k,j}$ is the reconstruction of source S_k at node j.

Note that if a node is not interested in a certain source, we may simply assume the distortion of the reconstruction at this node to be large. Thus we can write a distortion matrix without loss of generality, whose element $d_{k,j}$ is the distortion associated with the reconstruction of source S_k at node j. Clearly, without loss of generality, we can let the element $d_{i,i} = 0$ and simply define $d_{i,j} = d_i^{\max}$ for $i \notin \mathscr{T}_j$, where $d_i^{\max}$ is the distortion achievable at rate zero for source S_i. With this in mind, we next define the region of achievable distortion matrix.

Definition 2: A distortion matrix $\vec{D}$ is achievable for distributed network source coding with bandwidth mismatch factor κ, if for any $\epsilon > 0$ and sufficiently large m, there exist an integer $n \leq \kappa m$ and an $(m, n, \{d_{j,k}, k \in \mathscr{T}_j\})$ distributed network source code, such that $d_{i,j} \leq D_{i,j} + \epsilon$, $i, j = 1, 2, \ldots, N$. The collection of all such distortion matrices is the distributed network source coding achievable distortion region, denoted as $\mathcal{D}_{dis}$.

At this point, it is important to clarify the individual source code and channel code used in the separation-based approach. To this end, we essentially need to define the pure source coding problem and channel coding problem. The channel coding problem in the distributed network source coding problem is simply the point-to-point channel capacity problem, and the codes used are naturally block channel codes. The source coding problem is more complex: intuitively speaking, it is the original problem when the noisy channels are replaced by noise-free bit-pipes. However, the block source codes we use on this network need to incorporate the interactive communication aspect carefully. We next formally define such codes.

Definition 3: An $(m, l, \{M_{i,j}, (i,j) \in \mathcal{E}\}, \{d_{j,k}, k \in \mathscr{T}_j\})$ distributed network source *digital* code consists the following components:

- At each transmitter node j, for each k such that $(j,k) \in \mathcal{E}$, an encoding function for transmission session r
$$\tilde{\phi}_{j,k}^{(r)} : \mathcal{S}_j^m \times \prod_{(i,j)\in\mathcal{E}} \mathcal{I}_{M_{i,j}}^{r-1} \to \mathcal{I}_{M_{j,k}}, \ \ r = 1, 2, \ldots, l. \quad (3)$$
- At each receiver node j, for each source $k \in \mathscr{T}_j$, a decoding function
$$\tilde{\psi}_{j,k} : \prod_{(i,j)\in\mathcal{E}} \mathcal{I}_{M_{i,j}}^{l} \times \mathcal{S}_j^m \to \hat{\mathcal{S}}_k^m. \quad (4)$$

The encoding and decoding functions induce the distortions
$$d_{k,j} = \frac{1}{m}\sum_{t=1}^{m} d(S_k(t), \hat{S}_{k,j}(t)),$$
$$j = 1, 2, \ldots, M, \quad \text{and} \quad k \in \mathscr{T}_j,$$
where again $\hat{S}_{k,j}$ is the reconstruction of source S_k at node j.

In the above definition, $M_{i,j}$ is essentially the cardinality of the noise-free bit-pipe on edge $e = (i,j)$ per m source symbols. In this code, there are a total of l sessions of coding, and in each session the encoding functions need to observe the causality constraints on the communication session level. This kind of lossy source coding problems, though not often seen in the information theory literature, has in fact been considered for simple settings: Kaspi in 1985 considered the two-way source coding problem [12], which is defined in a very similar manner for a two-terminal setting.

Definition 4: A rate-distortion-matrix tuple $(\{R_{i,j}, (i,j) \in \mathcal{E}\}, \vec{D})$ is achievable on a given source communication network, if for any $\epsilon > 0$, there exists an integer l, such that for any sufficiently large m, there exists an $(m, l, \{M_{i,j}, (i,j) \in \mathcal{E}\}, \{d_{j,k}, k \in \mathscr{T}_j\})$ code such that
$$R_{i,j} + \epsilon \geq \frac{l}{m}\log M_{i,j}, \quad (i,j) \in \mathcal{E}$$
$$d_{j,k} \leq D_{j,k} + \epsilon, \quad k \in \mathscr{T}_j. \quad (5)$$
The collection of distortion matrix $\vec{D}$ for which the rate-distortion-matrix tuple $(\{R_{i,j}, (i,j) \in \mathcal{E}\}, \vec{D})$ is achievable for a given rate vector $\{R_{i,j}, (i,j) \in \mathcal{E}\}$ is denoted[1] as $\mathcal{D}_{dis}(\{R_{i,j}, (i,j) \in \mathcal{E}\})$.

With the above definition, it is clear that we can combine the digital source codes together with the capacity-achieving channel codes for each channel on the original communication network. More precisely, we can define the achievable distortion region using such a separation approach as
$$\mathcal{D}_{dis}^* = \mathcal{D}_{dis}(\{R_{i,j} = \kappa C_{i,j}, (i,j) \in \mathcal{E}\}). \quad (6)$$

IV. Optimality of Separation for Distributed Network Source Coding

Our main result is the following theorem.

Theorem 1: For any distributed network source coding problem (with orthogonal communication channels), we have $\mathcal{D}_{dis} = \mathcal{D}_{dis}^*$.

The condition on the synchronization of the channels in the network can be relaxed to some extent. Particularly, in a communication network without any relay or feedback, *i.e.*, on an interference channel without feedback, the orthogonal channels can have completely different bandwidth mismatch factors with the sources (thus asynchronized), and the source-channel separation architecture is still optimal.

[1]We have already used $\mathcal{D}_{dis}$ to denote the achievable distortion region for the joint coding problem, and here we slightly abuse the notation by using $\mathcal{D}_{dis}(\{R_{i,j}, (i,j) \in \mathcal{E}\})$ to denote the distortion-rate function in this pure source coding problem. This does not cause any confusion since the concept of rates does not naturally exist in the joint coding problem.

Proof: The forward part of the proof, i.e., $\mathcal{D}_{dis} \supseteq \mathcal{D}^*_{dis}$ is straightforward, and we omit it here due to space constraint.

Next we focus on the direction $\mathcal{D}_{dis} \subseteq \mathcal{D}^*_{dis}$. Consider an arbitrary joint source-channel coding scheme with $n \leq \kappa m$ channel uses, which achieves distortion matrix $\vec{D}$. It is instructive to first examine the joint distribution induced by this particular coding scheme. Let $X_{\mathcal{E}}(t)$ and $Y_{\mathcal{E}}(t)$ denote the collection of channel inputs and outputs at time t; similarly we use $S^m_{\mathcal{V}}$ to denote the collection of all the source vectors in the network. At $t = 1$, $X_{\mathcal{E}}(1)$ is a function of $S^m_{\mathcal{V}}$, *i.e.*, $X_{\mathcal{E}}(1) = \phi^{(1)}_{\mathcal{E}}(S^m_{\mathcal{V}})$ in the notation of (1); $X_{\mathcal{E}}(1)$ generates $Y_{\mathcal{E}}(1)$ via $|\mathcal{E}|$ orthogonal channels; note that we have $S^m_{\mathcal{V}} - X_{\mathcal{E}}(1) - Y_{\mathcal{E}}(1)$ form a Markov chain and $P(Y_{\mathcal{E}}(1)|X_{\mathcal{E}}(1)) = \prod_e P(Y_e(1)|X_e(1))$ since the channels are orthogonal. At $t = 2$, $X_{\mathcal{E}}(2)$ is a function of $S^m_{\mathcal{V}}$ and $Y_{\mathcal{E}}(1)$, *i.e.*, $X_{\mathcal{E}}(2) = \phi^{(2)}_{\mathcal{E}}(S^m_{\mathcal{V}}, Y_{\mathcal{E}}(1))$; $X_{\mathcal{E}}(2)$ further generates $Y_{\mathcal{E}}(2)$ via $|\mathcal{E}|$ orthogonal channels. Successively, at time t, we have

- Condition one: $X_{\mathcal{E}}(t) = \phi^{(t)}_{\mathcal{E}}(S^m_{\mathcal{V}}, Y^{t-1}_{\mathcal{E}})$;
- Condition two: $(S^m_{\mathcal{V}}, X^{t-1}_{\mathcal{E}}, Y^{t-1}_{\mathcal{E}}) - X_{\mathcal{E}}(t) - Y_{\mathcal{E}}(t)$ form a Markov chain, and $P(Y_{\mathcal{E}}(t)|X_{\mathcal{E}}(t)) = \prod_e P(Y_e(t)|X_e(t))$.

The basic idea following the example given in Section II is to simulate the channel output by a deterministic function of channel input, then send the output index of this function digitally through the channel via a channel code; this function should be constructed in such a way that the joint typicality is preserved. That is to say, in the evolution of the probability distribution induced by this deterministic substitution, condition one does not change while condition two is accurately simulated using a deterministic function. In order to accomplish this efficiently, the substitution has to be done over a long block. This however cannot be directly implemented because in a general network we cannot let the nodes wait for a long block while the channels idle[2]. To circumvent this difficulty, we use a technique adapted from [4]. The extracted pure source coding problem defined in Definition 3 and Definition 4 is rather crucial in the discussion below, and the readers are encouraged to familiarize with them before proceeding.

We next show that if a distortion matrix $\vec{D}$ is achievable in the joint coding problem defined in Definition 1 and Definition 2, which we denote as $_j$, then the rate distortion matrix pair $(\{\kappa C_e\}, \vec{D})$ is also achievable in the pure source coding problem defined in Definition 3 and Definition 4, which we denote as $_s$. For this purpose, we shall construct an n-session **digital** distributed network source code for $_s$ that operates on a source sequence of length mn' for some sufficiently large n', based on the original length-n joint source-channel code for $_j$, and upper-bound the rates used in the digital codes. We first partition the source sequence $S^{mn'}_i$, $i = 1, 2, \ldots, N$, into n' disjoint block components, each of length m. We shall write the l-th component of $S^{mn'}_i$, *i.e.*, $S_i((l-1)m+1), S_i((l-1)m+2), \ldots, S_i(lm)$, as $S^m_i \langle l \rangle$.

[2]This is in fact not an issue in the example given in Section II where no relay and feedback are present. In that setting, all the information is available before encoding starts, and thus the causality issue is significantly simplified.

For each $e \in \mathcal{E}$ and each *session* $t \in \mathcal{I}_n = \{1, \ldots, n\}$, we generate a source codebook $\mathcal{C}_{e,t}$ of size $\exp(n'(I(X_e(t); Y_e(t)) + \epsilon'))$ using $P_{Y_e(t)}$. This codebook is revealed to both the encoder and decoder on edge e in the problem $_s$.

Now consider encoding for session 1 at any given edge $e = (i, j) \in \mathcal{E}$. We first apply the original joint source channel encoding function $\phi^{(1)}_e$ on each block component $s^m_i \langle l \rangle$, $l = 1, 2, \ldots, n'$ and concatenate the outputs to produce a length-n' vector $x^{n'}_e(1)$. Note that after this is done at each node, $s^{mn'}_{\mathcal{V}}$ and $x^{n'}_{\mathcal{E}}(1)$ are strongly jointly typical with respect to $P_{S^m_{\mathcal{V}} X_{\mathcal{E}}(1)}$ with high probability. Then for each $e \in \mathcal{E}$, given a typical $x^{n'}_e(1)$, with high probability, we can find a $y^{n'}_e(1)$ codeword in $\mathcal{C}_{e,1}$ such that $x^{n'}_e(1)$ and $y^{n'}_e(1)$ are strongly jointly typical with respect to $P_{X_e(1) Y_e(1)}$; moreover, by the Markov lemma, $y^{n'}_e(1)$ is strongly jointly typical with $s^{mn'}_{\mathcal{V}}$ and $x^{n'}_{\mathcal{E}}(1)$ (with respect to $P_{S^m_{\mathcal{V}} X_{\mathcal{E}}(1) Y_{\mathcal{E}}(1)}$) with high probability.

In the problem $_s$, the index of the codeword $y^{n'}_e(1)$ is then delivered to the node j without error for each edge $e \in \mathcal{E}$ before the second session starts. The second session encoding starts for any given edge $e = (i, j) \in \mathcal{E}$ by applying the original joint source-channel encoding function $\phi^{(2)}_e$ on the following variables, for each l, respectively:

- The l-th component of $s^{mn'}_i$, *i.e.*, $s^m_i \langle l \rangle$;
- The l-th component of $y^{n'}_{e'}(1)$ for all $e' = (k, i) \in \mathcal{E}$, which is denoted as $y_{e'}(1, \langle l \rangle)$.

The outputs are concatenated, resulting in a length-n' vector $x^{n'}_e(2)$; see Figure 3. At this point, we can again find a $y^{n'}_e(2)$ codeword in $\mathcal{C}_{e,2}$ such that $x^{n'}_e(2)$ and $y^{n'}_e(2)$ are strongly jointly typical with respect to $P_{X_e(1) Y_e(1)}$, with high probability. An induction shows that by applying the above procedure, with high probability, the overall encoding procedure results in $s^{mn'}_{\mathcal{V}} x^{nn'}_{\mathcal{V}} y^{nn'}_{\mathcal{V}}$ strongly jointly typical with respective to the original distribution $P_{S^m_{\mathcal{V}} X^n_{\mathcal{V}} Y^n_{\mathcal{V}}}$; since n is finite in the original code, for sufficiently large n', the error in the jointly typicality definition is upper-bounded.

After the above n sessions of encoding, at node $v = j \in \mathcal{V}$, we apply the originally joint source-channel decoding function $\psi_{j,k}$ to reconstruct the l-th block component of source $s^{mn'}_k$, *i.e.*, $s_k \langle l \rangle$, for $k \in \mathscr{T}_j$ using the following variables:

- The l-th component of $s^{mn'}_j$, i.e, $s^m_j \langle l \rangle$;
- $y_{e'}(t, \langle l \rangle)$ for all $e' = (i, j) \in \mathcal{E}$ and all $t = 1, 2, \ldots, n$.

Denote the output of this function as $\hat{s}^m_{k,j} \langle l \rangle$. It follows that the vector pair sequence $(s^m_k \langle l \rangle, \hat{s}^m_{k,j} \langle l \rangle)$, $l = 1, 2, \ldots, n'$, is strongly jointly typical with respective to the distribution $P_{S^m \hat{S}^m_{k,j}}$, with high probability. This fact alone is sufficient to guarantee the average distortion over all the components is asymptotically close to $D_{k,j}$. Thus it is clear that this scheme achieves a distortion matrix $\vec{D} + \epsilon$, where $\epsilon > 0$ can be made arbitrarily small as $n' \to \infty$.

It remains to analyze the rates of this digital scheme, *i.e.*, the cardinalities of indices delivered during each session for

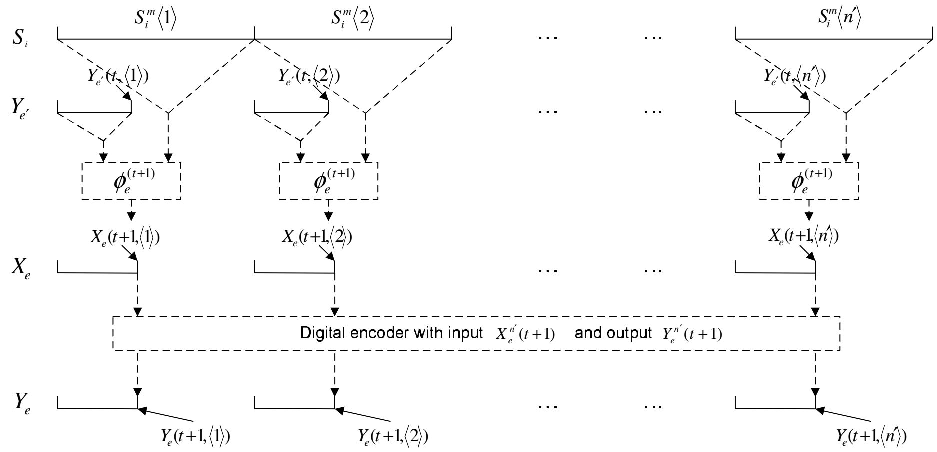

Fig. 3. Digital coding operations in session $t+1$ for a node i with an incoming link e' and an outgoing link e.

the problem $_s$. It is clear that for each link $e = (i,j) \in \mathcal{E}$, for each session t, we have

$$I(X_e(t); Y_e(t)) \leq C_e, \quad t = 1, 2, \ldots, n, \tag{7}$$

where C_e is the capacity of the channel on edge e, due to the conventional channel coding theorem. Thus it is seen that the cardinality of above digital codes for each session associated with any given link e is bounded as

$$\exp(n'(I(X_e(t); Y_e(t)) + \epsilon')) \leq \exp(n'(C_e + \epsilon')), \quad t = 1, 2, \ldots, n. \tag{8}$$

It follows that the following rate is achievable in problem $_s$

$$R_e = \frac{n}{mn'} \log \exp(n'(C_e + \epsilon')) = \kappa(C_e + \epsilon'), \tag{9}$$

according to Definition 4. Thus we have shown that the rate-distortion-matrix tuple $(\{R_e = \kappa(C_e + \epsilon')\}, \vec{D} + \epsilon)$ is achievable for the problem $_s$, where ϵ' and ϵ can be made arbitrarily small. Since the achievable rate-distortion-matrix region for $_s$ is a closed set, we can safely ignore the asymptotically small terms ϵ' and ϵ, and thus complete the proof for $\mathcal{D}_{dis} \subseteq \mathcal{D}^*_{dis}$ by applying (6). ■

V. CONCLUDING REMARKS

In a companion paper [11], we showed the optimality of separation for lossy coding of memoryless sources in a network with general multiuser communication channels, when the sources are mutually independent, and each source is needed only at one destination (or at multiple destinations at the same distortion level). For the same setting but each source is needed at multiple destinations under a restricted class of distortion measures, we also showed that the separation approach is approximately optimal, in the sense that the loss from optimum can be upper-bounded [11]. These results, together with the result in this paper, provide strong theoretical justification for using the separation approach in these scenarios. Our results are given in terms of rate-distortion tradeoff, however, they clearly also hold for lossless coding.

The results in this paper[3] and in [11] are obtained without explicit characterizations of the underlying regions. Such an approach of identifying properties without explicit individual component solutions is a valuable tool which may lead to further insights into network information theory problems. The source coding problem extracted from the distributed network source coding scenario implies that the interactive coding aspect needs to be carefully incorporated into this source coding problem, which suggests a distinct line of research direction into network source coding.

[3] After the completion of this work and [11], we became aware of the work in [13], which has independently arrived at the same result as presented in this paper.

REFERENCES

[1] L. Song, R. W. Yeung, and N. Cai, "A separation theorem for single-source network coding," *IEEE Trans. Information Theory*, vol. 52, no. 5, pp. 1861–1871, May 2006.

[2] S. P. Borade, "Network information flow: Limits and achievability," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 139.

[3] N. Ratnakar and G. Kramer, "The multicast capacity of deterministic relay networks with no interference," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2425–2432, Jun. 2006.

[4] R. Koetter, M. Effros, and M. Medard, "On a theory of network equivalence," in *Proc. IEEE Information Theory Workshop on Networking and Information Theory,* Jun. 2009, pp. 326–330.

[5] R. Koetter, "Separation for source, channel and network coding," plenary at IEEE Information Theory Workshop, Chengdu, China, Oct. 2006.

[6] R. Yeung, *Information theory and network coding.* Springer, 2008.

[7] Y. Steinberg and N. Merhav, "On hierarchical joint source-channel coding with degraded side information," *IEEE Trans. Information Theory*, vol. 52, no. 3, pp. 886–903, Mar. 2006.

[8] C. Tian and S. N. Diggavi, "On multistage successive refinement for Wyner-Ziv source coding with degraded side informations," *IEEE Trans. Information Theory*, vol. 53, no. 8, pp. 2946–2960, Aug. 2007.

[9] J.-J. Xiao and Z.-Q Luo, "Multiterminal source-channel communication over an orthogonal multiple-access channel," *IEEE Trans. Information Theory*, vol. 53, no. 9, pp. 3255–3264, Sep. 2007.

[10] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 1991.

[11] C. Tian, J. Chen, S N. Diggavi and S. Shamai, "Optimality and Approximate Optimality of Source-Channel Separation in Networks," to appear *Proc. IEEE Int. Symp. Information Theory*, Austin, Texas, June 2010. See also extended version submitted to IEEE Transactions on Information theory, April 15th 2010, http://arxiv.org/abs/1004.2648.

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

[13] S. Jalali and M. Effros, "On the separation of lossy source-network coding and channel coding in wireline networks," to appear *Proc. IEEE Int. Symp. Information Theory*, Austin, Texas, June 2010.