

Optimality and Approximate Optimality of Source-Channel Separation in Networks

Chao Tian, *Senior Member, IEEE*, Jun Chen, *Member, IEEE*, Suhas N. Diggavi, *Fellow, IEEE*,
and Shlomo Shamai (Shitz), *Fellow, IEEE*

Abstract—We consider the source-channel separation architecture for lossy source coding in communication networks. It is shown that the separation approach is optimal in two general scenarios and is approximately optimal in a third scenario. The two scenarios for which separation is optimal complement each other: the first is when the memoryless sources at source nodes are arbitrarily correlated, each of which is to be reconstructed at possibly multiple destinations within certain distortions, but the channels in this network are synchronized, orthogonal, and memoryless point-to-point channels; the second is when the memoryless sources are mutually independent, each of which is to be reconstructed only at one destination within a certain distortion, but the channels are general, including multi-user channels, such as multiple access, broadcast, interference, and relay channels, possibly with feedback. The third scenario, for which we demonstrate approximate optimality of source-channel separation, generalizes the second scenario by allowing each source to be reconstructed at multiple destinations with different distortions. For this case, the loss from optimality using the separation approach can be upper-bounded when a difference distortion measure is taken, and in the special case of quadratic distortion measure, this leads to universal constant bounds.

Index Terms—Joint source-channel coding, separation.

I. INTRODUCTION

SHANNON'S source-channel separation theorem asserts that there is no essential loss asymptotically in point-to-point communication systems, when the source coding

Manuscript received April 15, 2010; revised August 9, 2013; accepted September 4, 2013. Date of publication November 21, 2013; date of current version January 15, 2014. J. Chen was supported in part by an Early Researcher Award through the Province of Ontario and in part by the Natural Science and Engineering Research Council of Canada under a Discovery Grant. S. Diggavi was supported in part by the NSF under Award 1136174 and in part by MURI under Award AFOSR FA9550-09-064. S. Shamai (Shitz) was supported in part by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMUNICATIONS NEWCOM++ and NEWCOM# and in part by the Israel Science Foundation. This paper was presented at the 2010 IEEE International Symposium on Information Theory and the 2010 IEEE International Conference on Signal Processing and Communications.

C. Tian is with the AT&T Labs-Research, Florham Park, NJ 07932 USA (e-mail: tian@research.att.com).

J. Chen is with the Electrical and Computer Engineering Department, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: junchen@ece.mcmaster.ca).

S. N. Diggavi is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA (e-mail: suhas@ee.ucla.edu).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

Communicated by Y.-H. Kim, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2013.2291787

component and the channel coding component are designed and operated separately [1]. This separation architecture simplifies the overall communication system tremendously, because the decoupled subsystems are much easier to design and implement, with the codeword index as the only interface between the two components. Unfortunately, it has been shown that the separation approach is not optimal in very simple multiuser scenarios (e.g., [2]), which suggests that the optimality of source-channel separation may not hold beyond the conventional point-to-point case.

Because of the clear benefits of the source-channel separation architecture, it is important to understand the issue better. In this work, we seek to answer the following sequence of questions: is there a general class of multiuser communication systems for which

- The separation approach is optimal?
- If separation is not optimal, then is it at least approximately optimal?

The difficulty in answering these questions lies in the fact that in most multiuser communication scenarios, we do not have explicit characterizations of the rate-distortion regions, the channel capacity regions, or the joint coding achievable distortion regions; however, in order to determine whether the separation approach is optimal, it is natural to first couple the rate-distortion region and the channel capacity region, then compare it with the joint coding achievable distortion region. With at least one region unknown in most cases, it seems impossible to answer the above questions even in some of the simplest settings (e.g., communicating sources on an interference channel), let alone in more complex networks. In this work, we show that this difficulty in determining the optimality of source-channel separation can in fact be circumvented completely in several important settings, and the answers to the sequence of questions posed earlier are indeed positive.

More precisely, we show that for lossy coding of memoryless sources in a network, the source-channel separation approach is optimal for the following two general scenarios: the first scenario, referred to as *distributed network joint source-channel coding* (DNJSCC), is when the sources are arbitrarily correlated, each of which is to be reconstructed at possibly multiple destinations within certain distortions, but the channels between any pair of nodes in this network are synchronized, orthogonal, and memoryless; the second scenario, referred to as *joint source-channel multiple unicast with distortions* (JSCMUD), is when the sources are mutually

independent, each of which is to be reconstructed only at one destination within a certain distortion, but the channels can be general, including multi-user channels such as multiple access, broadcast, interference and relay channels, possibly with feedback.

The third scenario is a natural extension of the second one by allowing a source to be reconstructed at multiple destinations with different distortions; this case is referred to as *joint source-channel multiple multicast with distortions* (JSCMMD). For this scenario, the classical example of sending a Gaussian source over a Gaussian broadcast channel [3] reveals that the source-channel separation approach is not optimal in general. Thus we turn our attention to whether the separation approach is approximately optimal, and show that under a “difference” distortion measure, it is indeed so in the sense that the loss from the optimum can be upper-bounded. In the important special case of quadratic distortion measure, the upper bound is at most 0.5 bit per (additional) user which reconstructs the same source.

The optimality of source-channel separation beyond point-to-point communications has been considered in the past for more restricted classes of sources and channels [4]–[9], usually by taking advantage of the problem-specific structures and applying conventional techniques. The first scenario we consider, *i.e.*, the DNJSCC problem, is closely related to the problem treated in [10], where the optimality of the separation between channel coding and network coding [11] was established. In fact, our interest in the DNJSCC problem was motivated by the success in this work, from which we also borrow the ideas of channel simulation and sample interleaving; by applying these ideas directly, we obtain a concise proof for the DNJSCC problem without relying on the full-fledged stacked network as in [10], and our approach has the additional benefit of making explicit the underlying interactive source coding component. The result in [10] was extended to the DNJSCC scenario in [12] independently from and concurrently with our work [13]–[15]. Another relevant work is [16] where the super-channel view similar to what we use in the JSCMUD problem was applied to non-ergodic point-to-point channels. Also notable is the “information separation” discovered by Tuncel [17], which is a notion of separation weaker than the classical source-channel separation, and thus not the focus of this work.

The rest of this paper is organized as follows. Examples are provided in Section II to illustrate the underlying intuitions, and necessary definitions are given in Section III. The main results and the proofs on DNJSCC, JSCMUD and JSCMMD are given in Sections IV, V and VI, respectively. Section VII finally concludes the paper.

II. THREE EXAMPLES

In this section three examples are discussed in the context of sending sources on interference channels to provide some intuitions for the optimality or approximate optimality of source-channel separation in DNJSCC, JSCMUD and JSCMMD. The main results of this work are built on these intuitions, and Sections IV, V and VI essentially make them more precise

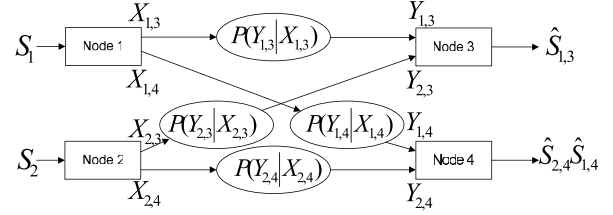


Fig. 1. Transmitting correlated sources on an interference network.

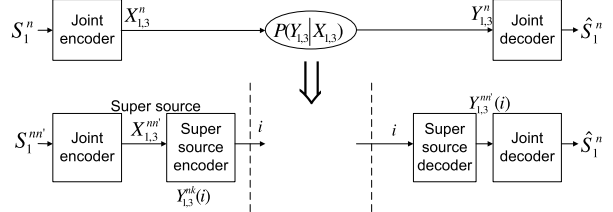


Fig. 2. Extracting a super-source from a joint source-channel code.

and rigorous. For simplicity, the channel bandwidth and the source bandwidth are assumed to match in this section.

A. An Example for Distributed Network Joint Source-Channel Coding

Consider the example in Fig. 1, where the discrete-time finite-alphabet memoryless sources S_1 and S_2 are correlated. Each discrete-time finite-alphabet memoryless channel between a transmitter and a receiver is orthogonal to the other channels: the channel from node i to node j has transition probability $P(Y_{i,j}|X_{i,j})$ and channel capacity $C_{i,j}$, and the overall transition probability of the channel network is $\prod_{(i,j)} P(Y_{i,j}|X_{i,j})$. Both node 3 and node 4 require a lossy reconstruction of source S_1 , denoted as $\hat{S}_{1,3}$ and $\hat{S}_{1,4}$, respectively. Node 4 also requires a lossy reconstruction of source S_2 , denoted as $\hat{S}_{2,4}$. The rate-distortion region of the underlying source coding problem is unknown, characterizing which is at least as difficult as the distributed source coding problem [18].

Suppose there exists a length- n joint source-channel code that achieves the distortion triple $(D_{1,3}, D_{1,4}, D_{2,4})$. 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 $X_{1,3}^n$, can be viewed as a super (block) source, independent and identically distributed across blocks; see Fig. 2. Therefore, we can encode a length- n' sequence of such blocks using a “rate-distortion” code of rate per block slightly exceeding $I(X_{1,3}^n; Y_{1,3}^n)$, the codewords of which are generated using the distribution $P(Y_{1,3}^n)$. It follows that with probability approaching one (as n' goes to infinity) a $Y_{1,3}^{nn'}$ codeword can be found in the codebook that is jointly typical with a channel input sequence $X_{1,3}^{nn'}$, *i.e.*, a length- n' vector of the super source samples. This lossy source code essentially simulates the channel output over n' length- n blocks, and only the codeword index needs to be known at node 3 to reconstruct the simulated channel output $Y_{1,3}^{nn'}$. Note that the rate of this code is $I(X_{1,3}^n; Y_{1,3}^n) \leq nC_{1,3}$; a similar argument

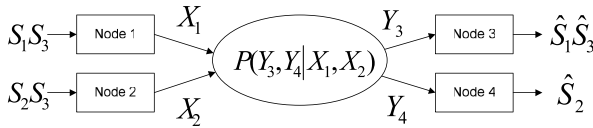


Fig. 3. Transmitting mutually independent S_1, S_2, S_3 on an interference channel.

holds for all other links. The original joint source-channel code decoders can now be applied on the simulated channel outputs to yield the reconstructions. This intuitively implies that the underlying *source coding problem* is guaranteed to achieve the distortion $(D_{1,3}, D_{1,4}, D_{2,4})$ at rates $(C_{1,3}, C_{1,4}, C_{2,3}, C_{2,4})$, which would further imply the optimality of the separation approach.

The above observation largely reflects the intuition behind the proof of source-channel separation for the DNJSCC 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 since channel usage constraints prevent coding over long super-channel blocks directly. The proof given in Section IV will resolve this difficulty through an intricate arrangement of channel simulation.

B. An Example for Joint Source-Channel Multiple Unicast With Distortions

Consider the problem depicted in Fig. 3, where the sources S_1, S_2 and S_3 are mutually independent; here the interference channel is more generally given by the transition probability $P(Y_3, Y_4 | X_1, X_2)$, where X_1, X_2 are the channel inputs by node 1 and node 2, respectively, and Y_3, Y_4 are the channel outputs at node 3 and node 4, respectively. Since the capacity region of the interference channel is unknown, it is infeasible to explicitly characterize the achievable distortion region of the separation approach.

Suppose a distortion triple (D_1, D_2, D_3) is achievable using some joint source-channel code of length- n . The key observation is now the following simple fact: if we fix this joint source-channel code, the transition probability of $P(\hat{S}_1^n, \hat{S}_2^n, \hat{S}_3^n | S_1^n, S_2^n, S_3^n)$ can be viewed as that of an alternative super interference channel with three users. On this super channel, the individual mutual information guarantee $I(S_i^n; \hat{S}_i^n) \geq nR_i(D_i)$ holds for $i = 1, 2, 3$, due to the conventional rate-distortion theorem [19]. Thus intuitively, this super channel is “good” since the mutual information $I(S_i^n; \hat{S}_i^n)$ terms are lower bounded, and the rate triple $(nR_1(D_1), nR_2(D_2), nR_3(D_3))$ should be in its capacity region, which would further imply that the achievable distortion triple (D_1, D_2, D_3) is achievable by the separation approach.

In order to show that the super interference channel can indeed support the rate triple $(nR_1(D_1), nR_2(D_2), nR_3(D_3))$, we essentially need to construct (random) codes over large super-channel blocks, and prove that the error probability can be made small, just as in conventional channels. The proof

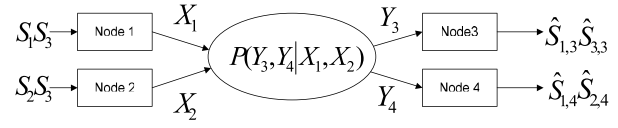


Fig. 4. Transmitting mutually independent S_1, S_2, S_3 on an interference channel to multiple destinations, *i.e.*, source S_1 is required at both destination node 3 and node 4.

in Section V follows this approach and makes the above intuitive argument more rigorous.

C. An Example for Joint Source-Channel Multiple Multicast With Distortions

Consider the problem depicted in Fig. 4, which is only slightly different from that in Fig. 3 in that source S_1 is to be reconstructed at both node 3 and node 4, denoted as $\hat{S}_{1,3}$ and $\hat{S}_{1,4}$, respectively; the reconstruction of source S_3 at node 3 is denoted¹ as $\hat{S}_{3,3}$ and the reconstruction of source S_2 at node 4 is denoted as $\hat{S}_{2,4}$. Taking a similar view as in the previous example, the abstracted channel now has transition probability $P(\hat{S}_{1,3}^n, \hat{S}_{1,4}^n, \hat{S}_{2,4}^n, \hat{S}_{3,3}^n | S_1^n, S_2^n, S_3^n)$. However, the mutual information bounds by the conventional rate-distortion theorem cannot be directly used as in the previous case. A moment of thought should convince the readers that the broadcast nature of the marginal transition probability $P(\hat{S}_{1,3}^n, \hat{S}_{1,4}^n | S_1^n)$ is the culprit, and some additional coding component is needed.

A natural separation architecture here is to use a successive refinement source code [20] to produce descriptions satisfying the distortion requirements for each destination and couple it to a superposition broadcast code [19] to deliver reliably these messages in the degraded message set [21]. More precisely, in the example of Fig. 4, assume without loss of generality that the distortion for source S_1 at node 3 is greater than that at node 4. A successive refinement code for S_1 can be used to produce messages $(W_{1,1}, W_{1,2})$ such that $W_{1,1}$ is to be delivered to node 3 and both $(W_{1,1}, W_{1,2})$ are to be delivered to node 4. Node 1 also produces a message $W_{3,1}$ to encode source S_3 , and node 2 produces a message $W_{2,1}$ to encode source S_2 . The messages $(W_{1,1}, W_{3,1})$ need to be reliably transmitted to node 3, and the messages $(W_{1,1}, W_{1,2}, W_{2,1})$ to node 4.

Let us for the moment isolate source S_1 and focus on the super block broadcast channel $P(\hat{S}_{1,3}^n, \hat{S}_{1,4}^n | S_1^n)$ with the messages $(W_{1,1}, W_{1,2})$, since it is the main difficulty in generalizing the proof approach for JSCMUD. We can show that this broadcast channel can support a certain rate pair for degraded message set broadcast, by introducing an additional auxiliary random variable. The same auxiliary random variable is also used to construct successive refinement source code for S_1 . The afore-mentioned broadcast channel code rates are however insufficient to support this successive refinement source code; nevertheless, the shortfall can be upper-bounded by comparing the channel code rates and the source code rates. This upper

¹The notation used here may seem unnatural initially, however it will become clear that this notation is convenient when generalizing to more complex networks.

bound implies the approximate optimality of source-channel separation in JSCMMMD.

III. NOTATION AND DEFINITIONS

In this section, notation and necessary definitions are provided. The notation would become rather unwieldy if a unified framework were used for all the problems treated in this work, therefore we forgo this ambitious goal and define the problems separately. We focus on the problems with discrete-time finite-alphabet memoryless sources, discrete-time finite-alphabet memoryless channels and bounded distortion measures, unless stated otherwise explicitly. It should be noted that it is often assumed that the sources are independent of the channels in such separation problems, which is also assumed in this work; this is because otherwise, even if the encoding and the decoding functions are designed separately, the inherent dependence between the source and the channel will render such a separation rather meaningless even in a point-to-point setting.

A. Definitions for the Distributed Network Joint Source-Channel Coding Problem

For this case, the network with a total of N nodes can be conveniently written as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \dots, N\}$ is the set of nodes, and \mathcal{E} is the set of edges between any two nodes; from here on, for any positive integer M , we use \mathcal{I}_M to denote the set $\{1, 2, \dots, M\}$.

Each edge $e = (i, j) \in \mathcal{E}$ is associated with a channel, whose transition probability is given as $P(Y_{i,j}|X_{i,j})$ with input alphabet $\mathcal{X}_{i,j}$ and output alphabet $\mathcal{Y}_{i,j}$ where the input and the output are not always independent, *i.e.*, the capacity of the channel on this link is non-zero; these channels are assumed to be synchronized. Each node i has a source S_i , distributed in the alphabet \mathcal{S}_i , and the collection of the sources are distributed according to the joint distribution $P(S_1, S_2, \dots, S_N)$ at each time instance. We have inherently assumed these sources are synchronized for simplicity, and thus the notation $P(S_1, S_2, \dots, S_N)$ is meaningful. A length- n vector of a source S_i is written as S_i^n , and the t -th symbol in this vector is written as $S_i(t)$; *i.e.*, $S_i^n = (S_i(1), S_i(2), \dots, S_i(n))$. A set of sources $\{S_i, i \in \mathcal{A}\}$ may be written as $\mathcal{S}_{\mathcal{A}}$; similarly, $\{X_{i,j}, (i, j) \in \mathcal{A}\}$ may be written as $X_{\mathcal{A}}$. Upper case is used for random variables, and lower case for their realizations. For any set \mathcal{S} , its r -th order product set is written as \mathcal{S}^r .

For each source, a distortion measure is defined as $d : \mathcal{S}_i \times \hat{\mathcal{S}}_i \rightarrow [0, \infty)$ where $\hat{\mathcal{S}}_i$ is the reconstruction alphabet. Nodes that are interested in a given source S_i may use different reconstruction alphabets and distortion measures, however, we do not distinguish them for notational simplicity. A node j may be interested in only a subset of the sources $\{S_i, i \in \mathcal{I}_N\}$; notationally, the set of sources that node j is interested in is written as \mathcal{T}_j . The class of codes being considered for the distributed network source coding problem are conventional block codes defined below.

Definition 1: An $(m, n, \{d_{k,j}, k \in \mathcal{T}_j\})$ distributed network joint source-channel code on a joint source-channel network

$(\mathcal{V}, \mathcal{E}, \{\mathcal{T}_j, j \in \mathcal{I}_N\}, P(S_1, S_2, \dots, S_N), \prod_{(i,j) \in \mathcal{E}} P(Y_{i,j}|X_{i,j}))$ consists of the following components:

- At each transmitter node i , for each j such that $(i, j) \in \mathcal{E}$, an encoding function for time instance t

$$\phi_{i,j}^{(t)} : \mathcal{S}_i^m \times \prod_{(k,i) \in \mathcal{E}} \mathcal{Y}_{k,i}^{t-1} \rightarrow \mathcal{X}_{i,j}, \quad t = 1, 2, \dots, n. \quad (1)$$

- At each receiver node j , for each source $k \in \mathcal{T}_j$, a decoding function

$$\psi_{k,j} : \prod_{(i,j) \in \mathcal{E}} \mathcal{Y}_{i,j}^n \times \mathcal{S}_j^m \rightarrow \hat{\mathcal{S}}_k^m. \quad (2)$$

The encoding and the decoding functions induce the distortions

$$d_{k,j} = \frac{1}{m} \sum_{t=1}^m \mathbb{E} d(S_k(t), \hat{S}_{k,j}(t)), \\ j = 1, 2, \dots, N, \quad \text{and } k \in \mathcal{T}_j,$$

where $\hat{S}_{k,j}$ is the reconstruction of source S_k at node j .

Here m is the source block length and n is the channel block length, which imply that there is a source-channel bandwidth mismatch factor of $\kappa = n/m$ (channel uses per source sample). If a node is not interested in a certain source, the distortion of the reconstruction at this node can simply be assumed to be large. Thus we can write a distortion matrix, whose element $d_{k,j}$ is the distortion associated with the reconstruction of source S_k at node j . Without loss of generality,² let the element $d_{i,i} = 0$ and define $d_{i,j} = d_i^{\max}$ for $i \notin \mathcal{T}_j$, where d_i^{\max} is the distortion achievable at rate zero for source S_i . The region of achievable distortion matrices can be defined as follows.

Definition 2: A distortion matrix \bar{D} is achievable for distributed network joint source-channel coding with bandwidth mismatch factor κ on a joint source-channel network $(\mathcal{V}, \mathcal{E}, \{\mathcal{T}_j, j \in \mathcal{I}_N\}, P(S_1, S_2, \dots, S_N), \prod_{(i,j) \in \mathcal{E}} P(Y_{i,j}|X_{i,j}))$, if for any $\epsilon > 0$ and sufficiently large m , there exist an integer $n \leq \kappa m$ and an $(m, n, \{d_{k,j}, k \in \mathcal{T}_j\})$ distributed network joint source-channel code, such that $d_{i,j} \leq \bar{D}_{i,j} + \epsilon$, $i, j = 1, 2, \dots, N$. The collection of all such distortion matrices is the distributed network joint source-channel coding achievable distortion region, denoted as \mathcal{D}_{dis} .

To discuss source-channel separation, it is important to define the source coding problem and the channel coding problem that are being separated into. The channel coding problem in DNJSCC is simply the point-to-point channel capacity problem. The source coding problem is more complex, which requires the incorporation of interactive coding.

Definition 3: An $(m, l, \{L_{i,j}, (i, j) \in \mathcal{E}\}, \{d_{k,j}, k \in \mathcal{T}_j\})$ distributed network source code with a total of l sessions on a source communication network $(\mathcal{V}, \mathcal{E}, \{\mathcal{T}_j, j \in \mathcal{I}_N\}, P(S_1, S_2, \dots, S_N))$ consists of the following components:

- At each (transmitter) node i , for each j such that $(i, j) \in \mathcal{E}$, an encoding function for transmission session

²Without loss of generality, we can always assume the minimum distortion for a given distortion measure is zero; see [22].

$t = 1, 2, \dots, l$,

$$\tilde{\phi}_{i,j}^{(t)} : \mathcal{S}_i^m \times \prod_{(k,i) \in \mathcal{E}} \mathcal{I}_{L_{k,i}}^{t-1} \rightarrow \mathcal{I}_{L_{i,j}}, \quad (3)$$

where $L_{i,j}$ and $L_{k,i}$'s are positive integers.

- At each receiver node j , for each source $k \in \mathcal{T}_j$, a decoding function

$$\tilde{\psi}_{k,j} : \prod_{(i,j) \in \mathcal{E}} \mathcal{I}_{L_{i,j}}^l \times \mathcal{S}_j^m \rightarrow \hat{\mathcal{S}}_k^m. \quad (4)$$

The encoding functions and the decoding functions induce the distortions

$$d_{k,j} = \frac{1}{m} \sum_{t=1}^m \mathbb{E}d(S_k(t), \hat{S}_{k,j}(t)),$$

$$j = 1, 2, \dots, N, \quad \text{and} \quad k \in \mathcal{T}_j,$$

where again $\hat{S}_{k,j}$ is the reconstruction of source S_k at node j .

Definition 4: A rate-distortion-matrix tuple $(\{R_{i,j}, (i,j) \in \mathcal{E}\}, \vec{D})$ is achievable on a source communication network $(\mathcal{V}, \mathcal{E}, \{\mathcal{T}_j, j \in \mathcal{I}_N\}, P(S_1, S_2, \dots, S_N))$, if for any $\epsilon > 0$, there exists an integer l , such that for any sufficiently large m , there exists an $(m, l, \{L_{i,j}, (i,j) \in \mathcal{E}\}, \{d_{k,j}, k \in \mathcal{T}_j\})$ distributed network source code such that

$$R_{i,j} + \epsilon \geq \frac{l}{m} \log L_{i,j}, \quad (i,j) \in \mathcal{E}$$

$$d_{i,j} \leq D_{i,j} + \epsilon, \quad i, j = 1, 2, \dots, N. \quad (5)$$

The collection of distortion matrices \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 denoted³ as $\mathcal{D}_{dis}(\{R_{i,j}\}_{(i,j) \in \mathcal{E}})$.

Note that in the above definition, m grows to infinity for any fixed value of l . One may alternatively define the region to allow m and l to grow in a more general manner. However, this alternative definition will only enlarge the region $\mathcal{D}_{dis}(\{R_{i,j}\}_{(i,j) \in \mathcal{E}})$, and thus does not affect the optimality result. In other words, the separation result we shall present is in fact stronger with the restrictions in Definitions 3 and 4 than that under a more general version of these definitions.

Roughly speaking, $\frac{l}{m} \log L_{i,j}$ is the rate of the noiseless channel on edge (i,j) per source symbol in each session. There are a total of l sessions, and on each edge the same rate is used in all sessions. At the end of each session, the index $w_{j,k} \in \mathcal{I}_{L_{j,k}}$ in this session becomes available at destination node k , which can be used by node k in the next session. In other words, the encoding functions observe the causality constraints on the session level. Note that the region $\mathcal{D}_{dis}(\{R_{i,j}\}_{(i,j) \in \mathcal{E}})$ is convex by a time-sharing argument. Definitions 3 and 4 specify a special class of interactive source coding problem, which appears particularly important given the result presented in this work.

We can now combine the source codes together with the capacity-achieving channel codes for each channel on the original communication network. More precisely, we can

³ \mathcal{D}_{dis} has already been used in 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 the source coding problem.

define the achievable distortion region using such a separation approach as

$$\mathcal{D}_{dis}^* = \mathcal{D}_{dis}(\{\kappa C_{i,j}\}_{(i,j) \in \mathcal{E}}), \quad (6)$$

where $C_{i,j}$ is the channel capacity between node i and node j , sometimes written as C_e with $e = (i,j) \in \mathcal{E}$.

B. Definitions for Joint Source-Channel Multiple Unicast and Multiple Multicast With Distortions

There are M mutually independent sources, denoted as S_i , distributed in the alphabet \mathcal{S}_i according to some distribution $P(S_i)$, $i = 1, 2, \dots, M$; note that the index i here is not related to the index of the node, unlike in Section III.A. For simplicity, we assume all the sources are synchronized. The distortion measures are defined similarly as in the last subsection, however we do not allow the existence of multiple distortion measures for the same source. Let the number of nodes be N . For simplicity we treat the overall communication network as a single memoryless channel, with inputs (X_1, X_2, \dots, X_N) over the alphabets $\mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_N$ and outputs (Y_1, Y_2, \dots, Y_N) over the alphabets $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \dots \times \mathcal{Y}_N$, and transition probability given by $P(Y_1^N | X_1^N)$; X_i and Y_i are the channel input and output at node i , respectively.

Each source S_i can be present at several nodes, and for each node $j \in \mathcal{I}_N$, we denote the sources present at node j as \mathcal{S}_j . The receiver demands are defined as follows:

- **Joint source-channel multiple unicast with distortions:** each source is to be reconstructed at a single destination. Again denote for receiver node j the set of the sources it is interested in as \mathcal{T}_j , then $\mathcal{T}_j \cap \mathcal{T}_k = \emptyset$ for any $j \neq k$.
- **Joint source-channel multiple multicast with distortions:** each source is to be reconstructed at multiple destinations, *i.e.*, it is possible that $\mathcal{T}_j \cap \mathcal{T}_k \neq \emptyset$.

Definition 5: An $(m, n, d_1, d_2, \dots, d_M)$ JSCMUD code on a source-channel communication network $(\{\mathcal{S}_j, j \in \mathcal{I}_N\}, \{\mathcal{T}_j, j \in \mathcal{I}_N\}, \prod_{i=1}^M P(S_i), P(Y_1^N | X_1^N))$ consists of the following components:

- At each transmitter node j , an encoding function for (time) index t

$$\phi_j^{(t)} : \prod_{i \in \mathcal{S}_j} \mathcal{S}_i^m \times \mathcal{Y}_j^{t-1} \rightarrow \mathcal{X}_j, \quad t = 1, 2, \dots, n. \quad (7)$$

- At each receiver node j , for each source $k \in \mathcal{T}_j$, a decoding function

$$\psi_{k,j} : \mathcal{Y}_j^n \times \prod_{i \in \mathcal{S}_j} \mathcal{S}_i^m \rightarrow \hat{\mathcal{S}}_k^m. \quad (8)$$

The encoding functions and decoding functions induce the distortion

$$d_k = \frac{1}{m} \sum_{t=1}^m \mathbb{E}d(S_k(t), \hat{S}_k(t)), \quad k = 1, 2, \dots, M,$$

where $\hat{S}_k(t)$ is the reconstruction of source S_k at a node j such that $k \in \mathcal{T}_j$.

Definition 6: A distortion vector (D_1, D_2, \dots, D_M) is achievable for JSCMUD on a source-channel communication

network $(\{\mathcal{S}_j, j \in \mathcal{I}_N\}, \{\mathcal{T}_j, j \in \mathcal{I}_N\}, \prod_{i=1}^M P(S_i), P(Y_1^N | X_1^N))$ with a bandwidth mismatch factor κ , if for any $\epsilon > 0$ and sufficiently large m , there exist an integer $n \leq \kappa m$ and an $(m, n, d_1, d_2, \dots, d_M)$ JSCMUD code, such that $d_i \leq D_i + \epsilon$, $i = 1, 2, \dots, M$. The collection of all such distortion vectors is the achievable JSCMUD distortion region, denoted as \mathcal{D}_{uni} .

Next we define the source coding problem and the channel coding problem that are being separated into. For the JSCMUD problem, the source codes are conventional lossy source codes. The channel coding problem is more involved: each source S_i is replaced with a message W_i of cardinality L_i with a uniform distribution; moreover, these messages are mutually independent. The precise channel code definition is as follows.

Definition 7: An $(n, L_1, L_2, \dots, L_M, P_{err})$ multiple unicast channel code on a channel communication network $(\{\mathcal{S}_j, j \in \mathcal{I}_N\}, \{\mathcal{T}_j, j \in \mathcal{I}_N\}, P(Y_1^N | X_1^N))$ consists of the following components:

- At each transmitter node j , an encoding function for (time) index t

$$\tilde{\phi}_j^{(t)} : \prod_{i \in \mathcal{S}_j} \mathcal{I}_{L_i} \times \mathcal{Y}_j^{t-1} \rightarrow \mathcal{X}_j, \quad t = 1, 2, \dots, n. \quad (9)$$

- At each receiver node j , for each message W_k where $k \in \mathcal{T}_j$, a decoding function

$$\tilde{\psi}_{k,j} : \mathcal{Y}_j^n \times \prod_{i \in \mathcal{S}_j} \mathcal{I}_{L_i} \rightarrow \mathcal{I}_{L_k}. \quad (10)$$

Denote the decoded message as \hat{W}_i at node j where $i \in \mathcal{T}_j$. The encoding functions and decoding functions induce the average decoding error probability

$$P_{err} = \Pr\left(\bigcup_{i=1}^M W_i \neq \hat{W}_i\right). \quad (11)$$

Definition 8: A rate vector (R_1, R_2, \dots, R_M) is achievable for multiple unicast channel coding on a channel communication network $(\{\mathcal{S}_j, j \in \mathcal{I}_N\}, \{\mathcal{T}_j, j \in \mathcal{I}_N\}, P(Y_1^N | X_1^N))$, if for any $\epsilon > 0$ and sufficiently large n , there exists an $(n, L_1, L_2, \dots, L_M, \epsilon)$ multiple unicast channel code, such that $R_i \leq \frac{1}{n} \log L_i + \epsilon$, $i = 1, 2, \dots, M$. The collection of such achievable rate vectors is the achievable capacity region of the network, denoted as \mathcal{C}_{uni} .

Using conventional rate-distortion codes on each source and then combining it with the above defined multiple unicast channel codes, an achievable distortion region is immediate, which will be denoted as \mathcal{D}_{uni}^* . More precisely, we can write

$$\mathcal{D}_{uni}^* = \bigcup_{(R_1, R_2, \dots, R_M) \in \mathcal{C}_{uni}} \left\{ (D_1, D_2, \dots, D_M) : \begin{array}{l} D_i \geq D_i(\kappa R_i), \\ i = 1, 2, \dots, M \end{array} \right\} \quad (12)$$

where $D_i(\cdot)$ is the distortion-rate function of the source S_i .

In the case of JSCMMD, a source is to be reconstructed with possibly different distortions at multiple destinations. The JSCMMD codes are defined in the same manner as in the case of JSCMUD, and thus the detailed definitions are omitted here. The achievable distortion matrix and the achievable distortion region \mathcal{D}_{mul} can also be defined accordingly.

The source-channel separation scheme for JSCMMD is slightly more involved. Consider first source S_i , and assume it is to be reconstructed in a lossy manner at nodes in the set $\mathcal{Q}_i = \{j : i \in \mathcal{T}_j\}$. The source codes we shall consider are successive refinement codes [20], and source S_i is encoded in $|\mathcal{Q}_i|$ stages, where the operator $|\cdot|$ denotes the cardinality of a set. For the channel codes in the separation approach, we consider the degraded message set problem [21]. More precisely, in the given communication network, fix an order O_i for the elements in the set \mathcal{Q}_i for each $i = 1, 2, \dots, M$. The source S_i is replaced with a total of $|\mathcal{Q}_i|$ messages, denoted as $W_{i,j}$, whose rate is $R_{i,O_i(j)}$, $j = 1, 2, \dots, |\mathcal{Q}_i|$, where $O_i(j)$ is the j -th element in the order O_i . The k -th node in this given order O_i is required to reconstruct the first k messages, $W_{i,j}$, $j = 1, 2, \dots, k$. We can now define the achievable capacity region $\mathcal{C}_{mul}(O_1, O_2, \dots, O_M)$ for this degraded message set problem, which depends on the set of orders $\vec{O} = (O_1, O_2, \dots, O_M)$; see the JSCMMD example in Section II-B, where $\mathcal{Q}_1 = \{3, 4\}$ and the specific order discussed is $O_1 = (3, 4)$.

The degraded message set problem naturally sets the stage for the successive refinement source codes, and by combining these two components, we arrive at an achievable distortion region using the separation approach for a given set of orders \vec{O} . We shall denote this achievable region as $\mathcal{D}_{mul}^*(\vec{O})$.

IV. OPTIMALITY OF SEPARATION FOR DISTRIBUTED NETWORK JOINT SOURCE-CHANNEL CODING

Our first main result formally states the optimality of source-channel separation in the DNJSCC problem. Recall \mathcal{D}_{dis} and \mathcal{D}_{dis}^* given in Definition 2 and Eqn. (6), respectively.

Theorem 1: $\mathcal{D}_{dis} = \mathcal{D}_{dis}^*$.

The uniform Markov lemma [23], [24] is needed in the proof of this theorem, which is an alternative version of the Markov lemma in [18], [25]. It is rewritten below using notation more convenient to us.

Lemma 1: Let $X \leftrightarrow Y \leftrightarrow Z$ be a Markov string in finite alphabets. For any fixed strongly jointly typical sequence pair (x^n, y^n) , let Z^n be chosen uniformly at random from the set which consists of all sequences that are strongly typical with y^n . Let $Q(\cdot)$ be the probability measure induced by this random choice. Then $\lim_{n \rightarrow \infty} Q(\{(x^n, y^n, Z^n) \text{ are not strongly jointly typical}\}) = 0$, and the convergence is uniform over the set of strongly jointly typical (x^n, y^n) sequence pairs.

Proof of Theorem 1:

Proof for the direction $\mathcal{D}_{dis} \supseteq \mathcal{D}_{dis}^*$: To prove this direction, it suffices to show

$$\mathcal{D}_{dis} \supseteq \bigcup_{\{R_{i,j} : (i,j) \in \mathcal{E}\} : R_{i,j} < \kappa C_{i,j}} \mathcal{D}_{dis}(\{R_{i,j}\}_{(i,j) \in \mathcal{E}}). \quad (13)$$

This is because the achievable distortion region \mathcal{D}_{dis} is closed, and the distortion-matrix-rate function $\mathcal{D}_{dis}(\{R_{i,j}\}_{(i,j) \in \mathcal{E}})$ is continuous in the relative interior of the non-negative quadrant (implied by its convexity), from which it follows that the condition $R_{i,j} < \kappa C_{i,j}$ can be replaced by $R_{i,j} \leq \kappa C_{i,j}$ in (13), implying that $\mathcal{D}_{dis} \supseteq \mathcal{D}_{dis}^*$ where \mathcal{D}_{dis}^* is defined in (6).

To show (13), let $\epsilon > 0$ be some quantity such that

$$\epsilon \leq \min_{(i,j) \in \mathcal{E}} (\kappa C_{i,j} - R_{i,j}) \quad (14)$$

for any chosen set of $\{R_{i,j}, (i,j) \in \mathcal{E}\}$ such that $R_{i,j} < \kappa C_{i,j}$, $(i,j) \in \mathcal{E}$. For any distortion vector $\{D_{k,j}, k \in \mathcal{T}_j\} \in \mathcal{D}_{dis}(\{R_{i,j}, (i,j) \in \mathcal{E}\})$, since it is achievable with $\{R_{i,j}, (i,j) \in \mathcal{E}\}$, there exists an l such that for any sufficiently large m , there exists an $(m, l, \{L_{i,j}, (i,j) \in \mathcal{E}\}, \{D_{k,j} + \frac{1}{4}\epsilon, k \in \mathcal{T}_j\})$ distributed network source code (see Definition 3) with a total of l sessions, where

$$R_{i,j} + \frac{\epsilon}{4} \geq \frac{l}{m} \log L_{i,j}. \quad (15)$$

We utilize this source code together with a good channel code for each channel in the original network. More precisely, there are at least a total of $\kappa m - 1$ channel uses available, and we shall partition them into l channel sessions, each with at least $\lfloor \frac{\kappa m - 1}{l} \rfloor$ channel uses. Thus the channel on edge (i,j) in each session can support a message of cardinality $\left\lfloor 2^{\lfloor \frac{\kappa m - 1}{l} \rfloor (C_{i,j} - \frac{1}{4}\epsilon)} \right\rfloor$, with maximum error probability (among all messages for each channel code) less than ϵ , by choosing m sufficiently large. Each session of the pure source code has a message output of cardinality no larger than $L_{i,j}$. Thus as long as

$$L_{i,j} \leq \left\lfloor 2^{\lfloor \frac{\kappa m - 1}{l} \rfloor (C_{i,j} - \frac{1}{4}\epsilon)} \right\rfloor, \quad (16)$$

we can use the digital channel codes to transmit the source code indices with vanishing error probability. It follows that for any $\epsilon > 0$, there exists a sufficiently large m such that the total error probability over l -sessions is less than $l|\mathcal{E}|\epsilon$ in this network. For (16) to hold under the condition (15), it suffices to have

$$2^{\frac{m}{l}(R_{i,j} + \frac{1}{4}\epsilon)} \leq 2^{(\frac{\kappa m - 1}{l} - 1)(C_{i,j} - \frac{1}{4}\epsilon)} - 1. \quad (17)$$

Eqn. (14) implies that for any $\epsilon > 0$, (17) is true for any sufficiently large m and the fixed l afore-mentioned in the $(m, l, \{L_{i,j}, (i,j) \in \mathcal{E}\}, \{D_{k,j} + \frac{1}{4}\epsilon, k \in \mathcal{T}_j\})$ distributed network source code, and subsequently (16) holds. Thus for any $\epsilon > 0$, by choosing m sufficiently large, the separation based scheme is able to achieve the distortion vector $\{D_{k,j} + \epsilon, k \in \mathcal{T}_j\}$ for any $\{D_{k,j}, k \in \mathcal{T}_j\} \in \mathcal{D}_{dis}(\{R_{i,j}, (i,j) \in \mathcal{E}\})$ with probability greater than or equal to $(1 - l|\mathcal{E}|\epsilon)$, and distortion D_{\max} with probability less than or equal to $l|\mathcal{E}|\epsilon$, for any $R_{i,j}$ such that $\kappa C_{i,j} - R_{i,j} > \epsilon$, $(i,j) \in \mathcal{E}$, where D_{\max} is the maximum distortion value for all the finite-alphabet sources in the network. Since ϵ can be made arbitrarily small and D_{\max} is finite, and moreover \mathcal{D}_{dis} is a closed set, (13) is indeed true.

Proof for the direction $\mathcal{D}_{dis} \subseteq \mathcal{D}_{dis}^*$: We wish to show that if a distortion matrix \vec{D} is achievable in the joint coding problem \mathbb{P}_j (Definitions 1 and 2), then the rate distortion matrix pair $(\{\kappa C_e\}, \vec{D})$ is also achievable in the source coding problem \mathbb{P}_s (Definitions 3 and 4). For this purpose, we construct an n -session distributed network source code for \mathbb{P}_s that operates on a source sequence of length mn' from a joint coding code. For any achievable distortion matrix \vec{D} and any $\epsilon > 0$, there exists an $(m, n, \{D_{k,j} + \epsilon, k \in \mathcal{T}_j\})$ distributed network joint

source-channel code (see Definitions 1 and 2), where $n \leq \kappa m$. Let us fix this joint coding code, and use it to construct a source code for \mathbb{P}_s .

First partition the source sequence $S_i^{mn'}$, $i = 1, 2, \dots, N$, into n' disjoint block components, each of length m . The v -th block component of $S_i^{mn'}$ is written as $S_i^m(v)$, i.e.,

$$S_i^m(v) \triangleq \left(S_i((v-1)m+1), S_i((v-1)m+2), \dots, S_i(vm) \right), \\ v = 1, 2, \dots, n'.$$

To make this partition explicit, $S_i^{mn'}$ is written in the sequel as $S_i^{m,(n')}$.

Codebook generation: For each $(i,j) \in \mathcal{E}$ and each session $t = 1, 2, \dots, n$, a source coding codebook $\mathcal{C}_{(i,j),t}$ of size $2^{n'(I(X_{i,j}(t); Y_{i,j}(t)) + \delta)}$ is generated by choosing from the strongly typical set of the random variable $Y_{i,j}(t)$ uniformly at random with replacement, where $\delta > 0$ is a small quantity $\delta \rightarrow 0$ as $n' \rightarrow \infty$. This codebook is revealed to both the encoder and the decoder on edge (i,j) in the problem \mathbb{P}_s .

Encoding and decoding: For session $t = 1$ at any given edge $(i,j) \in \mathcal{E}$, we first apply the chosen joint source channel encoding function $\phi_{i,j}^{(1)}$ on each block component $S_i^m(v)$, $v = 1, 2, \dots, n'$; denote the output $\phi_{i,j}^{(1)}(S_i^m(v))$ as $x_{i,j}(1, (v))$. The following length- n' vector is formed by concatenating them

$$x_{i,j}^{(n')}(1) \triangleq (x_{i,j}(1, \langle 1 \rangle), x_{i,j}(1, \langle 2 \rangle), \dots, x_{i,j}(1, \langle n' \rangle)). \quad (18)$$

For each $(i,j) \in \mathcal{E}$, if $x_{i,j}^{(n')}(1)$ is strongly typical, we find a codeword $y_{i,j}^{(n')}(1)$ in $\mathcal{C}_{(i,j),1}$ such that $x_{i,j}^{(n')}(1)$ and $y_{i,j}^{(n')}(1)$ are strongly jointly typical with respect to $P(X_{i,j}(1), Y_{i,j}(1))$; if there does not exist such a codeword, an error is declared. Denote the index of this chosen $y_{i,j}^{(n')}(1)$ codeword as $w_{i,j}(1)$; the v -th location in the vector $y_{i,j}^{(n')}(1)$ is written as $y_{i,j}(1, (v))$. The encoding functions $\tilde{\phi}_{i,j}^{(1)}$ for \mathbb{P}_s are given by

$$\tilde{\phi}_{i,j}^{(1)}(S_i^{m,(n')}) = w_{i,j}(1), \quad (i,j) \in \mathcal{E}. \quad (19)$$

In the t -th session, for any given edge $(i,j) \in \mathcal{E}$, the chosen joint source-channel encoding function $\phi_{i,j}^{(t)}$ is applied, and the outputs are concatenated (see Fig. 5), i.e.,

$$x_{i,j}^{(n')}(t) = \left(\phi_{i,j}^{(t)}(S_i^m(1), \{y_{k,i}^{t-1}(1), (k,i) \in \mathcal{E}\}), \right. \\ \left. \phi_{i,j}^{(t)}(S_i^m(2), \{y_{k,i}^{t-1}(2), (k,i) \in \mathcal{E}\}), \dots, \right. \\ \left. \phi_{i,j}^{(t)}(S_i^m(n'), \{y_{k,i}^{t-1}(n'), (k,i) \in \mathcal{E}\}) \right), \quad (20)$$

where

$$y_{i,j}^{t-1}(v) \triangleq (y_{i,j}(1, (v)), y_{i,j}(2, (v)), \dots, y_{i,j}(t-1, (v))). \quad (21)$$

For any $(i,j) \in \mathcal{E}$, if $x_{i,j}^{(n')}(t)$ is strongly typical, find a codeword $y_{i,j}^{(n')}(t)$ in $\mathcal{C}_{(i,j),t}$ such that $x_{i,j}^{(n')}(t)$ and $y_{i,j}^{(n')}(t)$ are strongly jointly typical with respect to $P(X_{i,j}(t), Y_{i,j}(t))$; if there does not exist such a codeword, an error is declared. The index of the chosen codeword $y_{i,j}^{(n')}(t)$ in $\mathcal{C}_{(i,j),t}$ is

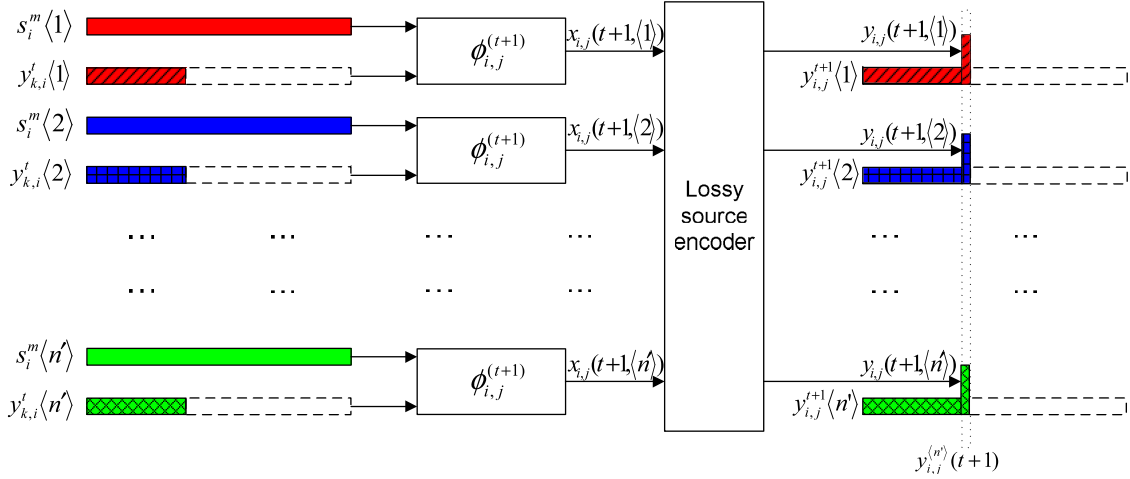


Fig. 5. Coding operation of \mathbb{P}_s in session $t+1$ for node i with an incoming link (k, i) and an outgoing link (i, j) . Each narrow horizontal box represents a vector; the vectors $y_{k,i}^t(v)$'s and $y_{i,j}^{t+1}(v)$'s are shaded partially because at this point, the later parts have not been generated. Each component of the lossy encoder output, i.e., $y_{i,j}^{t+1}(v)$, is appended to the existing $y_{i,j}^t(v)$ to form $y_{i,j}^{t+1}(v)$.

denoted as $w_{i,j}(t)$, and thus the encoding functions $\tilde{\phi}_{i,j}^{(t)}$ for \mathbb{P}_s are

$$\tilde{\phi}_{i,j}^{(t)}\left(s_i^{m,\langle n' \rangle}, \{w_{k,i}^{t-1}, (k, i) \in \mathcal{E}\}\right) = w_{i,j}(t), \quad (i, j) \in \mathcal{E}. \quad (22)$$

After n sessions of encoding, at node $j \in \mathcal{V}$, the chosen joint source-channel decoding function $\psi_{k,j}$ is applied to reconstruct the v -th block component of source $s_k^{m,\langle n' \rangle}$, $k \in \mathcal{T}_j$, i.e.,

$$\hat{s}_{k,j}^m\langle v \rangle = \psi_{k,j}(s_j^m\langle v \rangle, \{y_{i,j}^n\langle v \rangle, (i, j) \in \mathcal{E}\}), \quad v = 1, 2, \dots, n', \quad (23)$$

which are then concatenated to form $\hat{s}_{k,j}^{m,\langle n' \rangle}$, i.e., the length- mn' reconstruction of source k at node j . Thus the decoding functions $\tilde{\psi}_{k,j}$ for \mathbb{P}_s are given as

$$\tilde{\psi}_{k,j}(s_j^{m,\langle n' \rangle}, \{w_{i,j}^n, (i, j) \in \mathcal{E}\}) = \hat{s}_{k,j}^{m,\langle n' \rangle}, \quad k \in \mathcal{T}_j. \quad (24)$$

Error probability and distortion analysis: There are three kinds of error events in session- t

- $E_t^{(1)}$: $(s_{\mathcal{V}}^{m,\langle n' \rangle}, x_{\mathcal{E}}^{t,\langle n' \rangle}, y_{\mathcal{E}}^{t-1,\langle n' \rangle})$ are not strongly jointly typical with respect to $P(S_{\mathcal{V}}^m, X_{\mathcal{E}}^t, Y_{\mathcal{E}}^{t-1})$;
- $E_{t,(i,j)}^{(2)}$: for an edge $(i, j) \in \mathcal{E}$, given $x_{i,j}^{(n')}(t)$ is strongly typical, there does not exist any codeword in $\mathcal{C}_{(i,j),t}$ such that it is strongly jointly typical with $x_{i,j}^{(n')}(t)$ with respect to $P(X_{i,j}(t), Y_{i,j}(t))$;
- $E_t^{(3)}$: $(s_{\mathcal{V}}^{m,\langle n' \rangle}, x_{\mathcal{E}}^{t,\langle n' \rangle}, y_{\mathcal{E}}^{t-1,\langle n' \rangle})$ and $y_{\mathcal{E}}^{(n')}(t)$ are not strongly jointly typical with respect to $P(S_{\mathcal{V}}^m, X_{\mathcal{E}}^t, Y_{\mathcal{E}}^t)$.

Note $E_0^{(3)}$ is the event that $s_{\mathcal{V}}^{m,\langle n' \rangle}$ is not strongly jointly typical. The overall error event is given as

$$\begin{aligned} E_{n'} &= \bigcup_{t=1}^n (E_t^{(1)} \cup \bigcup_{(i,j) \in \mathcal{E}} E_{t,(i,j)}^{(2)} \cup E_t^{(3)}) \\ &= \bigcup_{t=1}^n \left(\overline{E_{t-1}^{(3)}} \cap E_t^{(1)} \right) \cup \left(\overline{E_t^{(1)}} \cap \bigcup_{(i,j) \in \mathcal{E}} E_{t,(i,j)}^{(2)} \right) \\ &\quad \cup \left(\overline{E_t^{(1)} \cup \bigcup_{(i,j) \in \mathcal{E}} E_{t,(i,j)}^{(2)}} \cap E_t^{(3)} \right), \end{aligned} \quad (25)$$

where \overline{S} is the complement of S .

By the union bound, we have

$$\begin{aligned} \Pr(E_{n'}) &\leq \sum_{t=1}^n \Pr(\overline{E_{t-1}^{(3)}} \cap E_t^{(1)}) \\ &\quad + \sum_{t=1}^n \Pr\left(\overline{E_t^{(1)}} \cap \bigcup_{(i,j) \in \mathcal{E}} E_{t,(i,j)}^{(2)}\right) \\ &\quad + \sum_{t=1}^n \Pr\left(\overline{E_t^{(1)} \cup \bigcup_{(i,j) \in \mathcal{E}} E_{t,(i,j)}^{(2)}} \cap E_t^{(3)}\right). \end{aligned} \quad (26)$$

Next we show that $\Pr(E_{n'}) \rightarrow 0$ as $n' \rightarrow \infty$. Firstly, $\Pr(E_0^{(3)}) \rightarrow 0$ by the basic properties of the strongly jointly typical sequences ([19], pp. 358-362). Since $x_{\mathcal{E}}^{(n')}(1)$ is a deterministic function of $s_{\mathcal{V}}^{m,\langle n' \rangle}$, $\Pr(\overline{E_0^{(3)}} \cap E_1^{(1)}) \rightarrow 0$, and similarly $\Pr(\overline{E_{t-1}^{(3)}} \cap E_t^{(1)}) \rightarrow 0$ for $t = 2, 3, \dots, n$. For the second summation in (26),

$$\begin{aligned} &\sum_{t=1}^n \Pr\left(\overline{E_t^{(1)}} \cap \bigcup_{(i,j) \in \mathcal{E}} E_{t,(i,j)}^{(2)}\right) \\ &\leq \sum_{t=1}^n \sum_{(i,j) \in \mathcal{E}} \Pr(\overline{E_t^{(1)}} \cap E_{t,(i,j)}^{(2)}), \end{aligned} \quad (27)$$

by the union bound. Since $\overline{E_t^{(1)}}$ implies that $x_{i,j}^{(n')}(t)$ is strongly typical, $\Pr(\overline{E_t^{(1)}} \cap E_{t,(i,j)}^{(2)}) \rightarrow 0$ for any t and $(i, j) \in \mathcal{E}$, by the properties of the strongly typical sequences ([19], Lemma 13.6.2), and the fact that the number of codewords in $\mathcal{C}_{(i,j),t}$ is $2^{n'(I(X_{i,j}(t); Y_{i,j}(t)) + \delta)}$.

To bound the third summation in (26), let us fix an arbitrary order for the edges in the set \mathcal{E} , and write it as $e_1, e_2, \dots, e_{|\mathcal{E}|}$. Define $E_{t,k}^{(3)}$ as the event that $(s_{\mathcal{V}}^{m,\langle n' \rangle}, x_{\mathcal{E}}^{t,\langle n' \rangle}, y_{\mathcal{E}}^{t-1,\langle n' \rangle})$ and $(y_{e_1}^{(n')}(t), y_{e_2}^{(n')}(t), \dots, y_{e_k}^{(n')}(t))$ are not strongly jointly typical.

We can then rewrite

$$\begin{aligned} & \overline{E_t^{(1)} \cup \bigcup_{(i,j) \in \mathcal{E}} E_{t,(i,j)}^{(2)}} \cap E_t^{(3)} \\ &= \bigcup_{k=1}^{|\mathcal{E}|} \overline{E_t^{(1)} \cup \bigcup_{(i,j) \in \mathcal{E}} E_{t,(i,j)}^{(2)} \cup E_{t,k-1}^{(3)}} \cap E_{t,k}^{(3)} \triangleq \bigcup_{k=1}^{|\mathcal{E}|} E_{t,k}^{(3)*}, \end{aligned} \quad (28)$$

where $E_{t,0}^{(3)} \triangleq \emptyset$. To bound $\Pr(E_{t,k}^{(3)*})$, observe that

$$\begin{aligned} & (S_{\mathcal{V}}^m, X_{\mathcal{E}}^{t-1}, X_{\mathcal{E} \setminus e_k}(t), Y_{\mathcal{E}}^{t-1}, Y_{e_1, e_2, \dots, e_{k-1}}(t)) \\ & \leftrightarrow X_{e_k}(t) \leftrightarrow Y_{e_k}(t) \end{aligned} \quad (29)$$

is a Markov string. Invoking Lemma 1 gives that $\Pr(E_{t,k}^{(3)*}) \rightarrow 0$, for any t and k , as $n' \rightarrow \infty$.

There are a total of n terms in the first summation of (26), $n|\mathcal{E}|$ terms in the second, and $n|\mathcal{E}|$ terms in the third. Since n and $|\mathcal{E}|$ are fixed here, and each term can be made arbitrarily small by making n' sufficiently large, we have $\Pr(E_{n'}) \rightarrow 0$ as $n' \rightarrow \infty$. This implies that the sequences $(s_{\mathcal{V}}^{m,(n')}, x_{\mathcal{E}}^{m,(n')}, y_{\mathcal{E}}^{m,(n')})$ are strongly jointly typical with respect to the original distribution $P(S_{\mathcal{V}}^m, X_{\mathcal{E}}^m, Y_{\mathcal{E}}^m)$ with probability arbitrarily close to one as $n' \rightarrow \infty$. This further implies that $s_k^{m,(n')}$ and $\hat{s}_{k,j}^{m,(n')}$ are strongly jointly typical with respect to $P(S_k^m, \hat{S}_{k,j}^m)$, and the new code induces a distortion $D_{k,j} + \epsilon + \delta'$, where $\delta' \rightarrow 0$ as $n' \rightarrow \infty$.

Rate analysis: In the chosen $(m, n, \{D_{k,j} + \epsilon, k \in \mathcal{T}_j\})$ joint source-channel code for \mathbb{P}_j , for each link $e = (i, j) \in \mathcal{E}$, the conventional channel coding theorem implies

$$I(X_e(t); Y_e(t)) \leq C_e, \quad t = 1, 2, \dots, n, \quad (30)$$

where C_e is the capacity of the channel on edge e . Thus the cardinality of above-constructed source code for \mathbb{P}_s in each session associated with any given link e is bounded as

$$2^{n'(I(X_e(t); Y_e(t)) + \delta)} \leq 2^{n'(C_e + \delta)}, \quad t = 1, 2, \dots, n. \quad (31)$$

It follows according to Definition 4 that the following rate is achievable in problem \mathbb{P}_s

$$R_e = \frac{n}{mn'} \log 2^{n'(C_e + \delta)} \leq \kappa(C_e + \delta). \quad (32)$$

Finishing the proof $\mathcal{D}_{dis} \subseteq \mathcal{D}_{dis}^*$: We have shown that by utilizing a chosen joint source-channel code $(m, n, \{D_{k,j} + \epsilon, k \in \mathcal{T}_j\})$ for \mathbb{P}_j , the constructed sequence of source codes for \mathbb{P}_s can operate at the rate-distortion-matrix tuple $(\{R_{i,j} = \kappa(C_{i,j} + \delta), (i, j) \in \mathcal{E}\}, \bar{D} + \epsilon + \delta')$, where δ and δ' can be made arbitrarily small by letting $n' \rightarrow \infty$. Since the achievable rate-distortion-matrix region for \mathbb{P}_s is a closed set, the tuple $(\{R_{i,j} = \kappa C_{i,j}, (i, j) \in \mathcal{E}\}, \bar{D} + \epsilon)$ is achievable in \mathbb{P}_s . Since the distortion matrix \bar{D} is achievable in \mathbb{P}_j , for any $\epsilon > 0$, there exists an $(m, n, \{D_{k,j} + \epsilon, k \in \mathcal{T}_j\})$ joint source-channel code, where $n \leq \kappa m$ by choosing n sufficiently large. Thus for any $\epsilon > 0$, $(\{R_{i,j} = \kappa C_{i,j}, (i, j) \in \mathcal{E}\}, \bar{D} + \epsilon)$ is achievable in \mathbb{P}_s . Again by the fact that the achievable rate-distortion-matrix region for \mathbb{P}_s is closed, the tuple $(\{R_e = \kappa C_e, e \in \mathcal{E}\}, \bar{D})$ is achievable for \mathbb{P}_s . Applying (6) now completes the proof for $\mathcal{D}_{dis} \subseteq \mathcal{D}_{dis}^*$. ■

V. OPTIMALITY OF SEPARATION FOR JOINT SOURCE-CHANNEL MULTIPLE UNICAST WITH DISTORTIONS

The following theorem formally states that source-channel separation is optimal in the JSCMUD problem. Recall \mathcal{D}_{uni} and \mathcal{D}_{uni}^* given in Definition 6 and Eqn. (12), respectively.

Theorem 2: $\mathcal{D}_{uni} = \mathcal{D}_{uni}^*$.

Proof of Theorem 2: The direction $\mathcal{D}_{uni} \supseteq \mathcal{D}_{uni}^*$ is rather obvious except one technicality. The channel coding problem given in Definitions 7 and 8 has an error probability defined as averaged over all messages. However, the codeword indices for the source codes may not have a uniform distribution, and thus the overall error probability by combing the source code and the channel code may be larger if the mapping between the source code indices and the channel code indices are chosen poorly. This however can be resolved using a standard random coding argument [19] over all possible one-to-one mappings, and the detail is thus omitted.

We next focus on the other direction $\mathcal{D}_{uni} \subseteq \mathcal{D}_{uni}^*$. For any achievable distortion vector (D_1, D_2, \dots, D_M) , and any $\epsilon > 0$, there exists an $(m, n, D_1 + \epsilon, D_2 + \epsilon, \dots, D_M + \epsilon)$ JSCMUD code, where $n \leq \kappa m$ (see Definitions 5 and 6). The sources and the above given block code induce a joint distribution

$$\prod_{i=1}^M P(S_i^m) \cdot P\left(\hat{S}_1^m, \hat{S}_2^m, \dots, \hat{S}_M^m \middle| S_1^m, S_2^m, \dots, S_M^m\right), \quad (33)$$

and the second term can be viewed as the transition probability of a block-level interference channel, which has input alphabets $S_1^m \times S_2^m \times \dots \times S_M^m$, and output alphabets $\hat{S}_1^m \times \hat{S}_2^m \times \dots \times \hat{S}_M^m$. Moreover, by the conventional rate-distortion theorem [19],

$$I(S_i^m; \hat{S}_i^m) \geq m R_i(D_i + \epsilon), \quad i = 1, 2, \dots, M, \quad (34)$$

where $R_i(\cdot)$ is the rate-distortion function for source S_i . This super interference channel operates in the same manner as a memoryless interference channel, however on a block level $(S_1^m, S_2^m, \dots, S_M^m) \rightarrow (\hat{S}_1^m, \hat{S}_2^m, \dots, \hat{S}_M^m)$, instead of on a single time instance level $(X_1, X_2, \dots, X_N) \rightarrow (Y_1, Y_2, \dots, Y_N)$.

Next we show that if a distortion vector (D_1, D_2, \dots, D_M) is achievable on the joint coding problem \mathbb{P}_j (Definitions 5 and 6), then the rate vector $(R_1(D_1), R_1(D_1), \dots, R_M(D_M))$ is achievable on the channel coding problem \mathbb{P}_c (Definitions 7 and 8). For this purpose, we construct a multiple unicast channel code for \mathbb{P}_c using the afore-mentioned $(m, n, D_1 + \epsilon, D_2 + \epsilon, \dots, D_M + \epsilon)$ joint source-channel code for \mathbb{P}_j . The coding scheme for \mathbb{P}_c can be formally described as follows.

Codebook generation: For each source S_i , $2^{mn'(R_i(D_i + \epsilon) - \delta)}$ codewords of length- (mn') are generated independently, according to the mn' -th product distribution of $P(S_i)$; denote this codebook as \mathcal{C}_i . The codebooks are revealed to all the nodes.

Encoding: To encode for \mathbb{P}_c , for a message w_i , choose the w_i -th codeword $s_i^{mn'}(w_i)$ in the \mathcal{C}_i codebook generated above. Each codeword is partitioned into n' blocks of equal length,

and denote the v -th block as $s_i^m(w_i, \langle v \rangle)$; to emphasize this partition, we also write $s_i^{mn'}(w_i)$ as $s_i^{m, \langle n' \rangle}(w_i)$. For a fixed v , the blocks $(s_1^m(w_1, \langle v \rangle), s_2^m(w_2, \langle v \rangle), \dots, s_M^m(w_M, \langle v \rangle))$ from the chosen codewords at all the nodes can be viewed as the length- m source vectors in \mathbb{P}_j , and thus the chosen $(m, n, D_1 + \epsilon, D_2 + \epsilon, \dots, D_M + \epsilon)$ JSCMUD encoding functions and decoding functions can be used on them. This results in a set of reconstruction sequences $(\hat{s}_1^m \langle v \rangle, \hat{s}_2^m \langle v \rangle, \dots, \hat{s}_M^m \langle v \rangle)$. At the end of n' blocks, we concatenate the reconstruction for each source block as $\hat{s}_i^{mn'} = \hat{s}_i^{m, \langle n' \rangle} = (\hat{s}_i^m \langle 1 \rangle, \hat{s}_i^m \langle 2 \rangle, \dots, \hat{s}_i^m \langle n' \rangle)$.

Mathematically, let the chosen joint source-channel encoding function and decoding function at node j be $\phi_j^{(t)}$ and $\psi_{k,j}$, respectively. Similarly as the notation of $s_i^m(w_i, \langle v \rangle)$, the v -th length- n block of $y_i^{n, \langle n' \rangle}$ is written as $y_i^n \langle v \rangle$, and the first t symbols of the block $y_i^n \langle v \rangle$ is written as $y_i^t \langle v \rangle$. Then the new channel code encoding function $\tilde{\phi}_j^{(t)}$ is given by

$$\begin{aligned} \tilde{\phi}_j^{(v-1)n+t} & \left(\{w_i, i \in \mathcal{S}_j\}, y_j^{(v-1)n+t-1} \right) \\ & = \phi_j^{(t)} \left(\{s_i^m(w_i, \langle v \rangle), i \in \mathcal{S}_j\}, y_j^{t-1} \langle v \rangle \right), \\ & \quad v = 1, 2, \dots, n', \quad t = 1, 2, \dots, n. \end{aligned} \quad (35)$$

The reconstructions are

$$\hat{s}_j^m \langle v \rangle = \psi_{k,j} \left(\{s_i^m(w_i, \langle v \rangle), i \in \mathcal{S}_j\}, y_j^n \langle v \rangle \right), \quad k \in \mathcal{T}_j. \quad (36)$$

Decoding: At node j , for which $k \in \mathcal{S}_j$, find a unique codeword in the codebook \mathcal{C}_k such that it is (weakly) jointly typical [19] with $\hat{s}_k^{m, \langle n' \rangle}$ according to the distribution $P(S_k^m, \hat{S}_k^m)$, *i.e.*, the marginal from (33). If there is a unique one, the corresponding message w_k^* is declared; otherwise an error is declared.

Error probability analysis: There are three kinds of errors

- $E^{(1)}$: $(s_1^{m, \langle n' \rangle}(w_1), s_2^{m, \langle n' \rangle}(w_2), \dots, s_M^{m, \langle n' \rangle}(w_M))$ are not jointly typical with respect to (33);
- $E^{(2)}$: $(s_1^{m, \langle n' \rangle}(w_1), s_2^{m, \langle n' \rangle}(w_2), \dots, s_M^{m, \langle n' \rangle}(w_M), \hat{s}_1^{m, \langle n' \rangle}, \hat{s}_2^{m, \langle n' \rangle}, \dots, \hat{s}_M^{m, \langle n' \rangle})$ are not jointly typical with respect to (33);
- $E_i^{(3)}$: for a given message w_i , there is more than one codeword in \mathcal{C}_i that is jointly typical with $\hat{s}_i^{m, \langle n' \rangle}(w_i)$, with respect to the marginal of (33).

By the union bound, the overall error probability can be bounded as

$$\begin{aligned} \Pr(E_{n'}) & \leq \Pr(E^{(1)}) + \Pr(\overline{E^{(1)}} \cap E^{(2)}) \\ & \quad + \sum_{i=1}^M \Pr(\overline{E^{(2)}} \cap E_i^{(3)}). \end{aligned} \quad (37)$$

Since all the codewords are generated according to $P(S_i)$'s independently, by the basic properties of the jointly typical sequences ([19], Theorem 14.2.1), $\Pr(E^{(1)}) \rightarrow 0$ as $n' \rightarrow \infty$. This implies that the reconstructions $\{\hat{s}_i^{m, \langle n' \rangle}, i = 1, 2, \dots, M\}$ are jointly typical with $\{s_i^{m, \langle n' \rangle}(w_i), i = 1, 2, \dots, M\}$ with probability approaching one, *i.e.*, $\Pr(\overline{E^{(1)}} \cap E^{(2)}) \rightarrow 0$ as $n' \rightarrow \infty$.

It follows that $\Pr(\overline{E^{(2)}} \cap E_i^{(3)}) \rightarrow 0$ as $n' \rightarrow \infty$, by (34) and the basic property of the jointly typical sequences ([19], Theorem 14.2.1 and Theorem 14.2.2), and the fact that the number of codewords in \mathcal{C}_i is $2^{mn'(R_i(D_i+\epsilon)-\delta)}$. Since there are a total of $M+2$ terms in (37), $\Pr(E_{n'}) \rightarrow 0$ as $n' \rightarrow \infty$.

Finishing the proof $\mathcal{D}_{uni} \subseteq \mathcal{D}_{uni}^*$: We have shown that by fixing a joint source-channel code $(m, n, D_1 + \epsilon, D_2 + \epsilon, \dots, D_M + \epsilon)$ for \mathbb{P}_j , the constructed sequence of channel codes can operate at rate tuple $(R_1(D_1 + \epsilon) - \delta, R_2(D_2 + \epsilon) - \delta, \dots, R_M(D_M + \epsilon) - \delta)$ for \mathbb{P}_c , where δ can be made arbitrarily small by letting $n' \rightarrow \infty$. Since the set \mathcal{C}_{uni} is closed, the rate tuple $(R_1(D_1 + \epsilon), R_2(D_2 + \epsilon), \dots, R_M(D_M + \epsilon)) \in \mathcal{C}_{uni}$. Since the rate-distortion functions $R_i(\cdot)$'s are continuous and the capacity region \mathcal{C}_{uni} is closed, we have $(R_1(D_1), R_2(D_2), \dots, R_M(D_M)) \in \mathcal{C}_{uni}$. It follows that $(D_1, D_2, \dots, D_M) \in \mathcal{D}_{uni}^*$ by the definition of \mathcal{D}_{uni}^* in (12), and thus $\mathcal{D}_{uni} \subseteq \mathcal{D}_{uni}^*$. This completes the proof. ■

VI. APPROXIMATE OPTIMALITY OF SEPARATION FOR JOINT SOURCE-CHANNEL MULTIPLE MULTICAST WITH DISTORTIONS

In this section the third scenario where there could be multiple receivers interested in the same source at different distortion levels is examined. We limit ourselves to a set of distortion measures referred to as the ‘‘difference’’ distortion measures, whose properties play an important role in the proof. More precisely, $\mathcal{X} = \mathcal{X}$ in this class of distortion measures, where \mathcal{X} is an Abelian group with a proper addition operation; furthermore, the distortion mapping $d(x, \hat{x})$ is a function of $x - \hat{x}$, and we shall write it as $d(x, \hat{x}) = d(x - \hat{x})$.

Some necessary definitions are quoted next from [26]. For random variables N and X in the alphabet \mathcal{X} , the capacity of the additive noise channel $X \rightarrow X + N$, under a $d(\cdot)$ distortion constraint is defined as

$$C(D, N) = \sup_{X: X \perp N, \mathbb{E}d(X) \leq D} I(X; X + N). \quad (38)$$

The addition $+$ is in the Abelian group \mathcal{X} (e.g., real addition, modulo addition or finite field addition), and \perp stands for independence. The minimax (or worst noise) capacity is defined as

$$C_{\mathcal{X}}(D) = \inf_{\mathbb{E}d(N) \leq D} C(D, N). \quad (39)$$

$C_{\mathcal{X}}(D)$ can be interpreted as the capacity at equilibrium in a mutual information jammer game, played over an additive-noise channel, in which both the expected noise and expected input are limited to within D in terms of $d(\cdot)$. The quantity $C_{\mathcal{X}}(D)$ is a function of D in general, however simplification is possible in some cases. Particularly, when the distortion is the mean squared error, $C_{\mathcal{X}}(D)$ is always 0.5 bit [26].

Our approximation result is in a genie-aided form, where additional communication links with bounded capacities are provided by a genie. We show that a separation-based approach using the original communication network together with the additional genie-provided communication links can achieve any distortion matrix \vec{D} that is achievable in the original communication network with arbitrary joint coding schemes.

It will become clear in the proof that if the reconstructions of S_i at multiple destinations in the set \mathcal{Q}_i are required to be at the same distortion level a priori, then these destinations can be viewed as a single super-destination, and the problem can be reduced; therefore, without loss of generality they are assumed to be at different distortion levels.

The decreasing sequence of distortions for the elements on the i -th row in the distortion matrix specifies an order O_i of the set \mathcal{Q}_i ; let $O_i(j)$ be the j -th element in the set of \mathcal{Q}_i according to the order O_i . We require these genie-provided links to support degraded message set broadcast from source S_i to the nodes in the set \mathcal{Q}_i for each i where $|\mathcal{Q}_i| > 1$: for such a source S_i , for each $j \in \mathcal{I}_{|\mathcal{Q}_i|}$, there is a common link of capacity $R_{i,O_i(j)}$ per source sample⁴ from S_i to all the nodes $O_i(j), O_i(j+1), \dots, O_i(|\mathcal{Q}_i|)$. These rate entries are collected and written together as the rate matrix \vec{R} . Consider adding these genie-provided links on top of the original source communication network, and denote the achievable distortion region using a separation approach of successive refinement coupled with superposition channel code on this new communication network as $\mathcal{D}_{mul}^{**}(\vec{O}, \vec{R})$.

Example: Consider the example given in Fig. 4. The sets \mathcal{Q}_i 's are

$$\mathcal{Q}_1 = \{3, 4\}, \quad \mathcal{Q}_2 = \{4\}, \quad \mathcal{Q}_3 = \{3\}. \quad (40)$$

The orders when the distortion of $\hat{S}_{1,3}$ is larger than $\hat{S}_{1,4}$ are

$$O_1 = (3, 4), \quad O_2 = (4), \quad O_3 = (3). \quad (41)$$

The rate matrix of the genie-provided links has the form

$$\vec{R} = \begin{bmatrix} \diamond & \diamond & R_{1,3} & R_{1,4} \\ \diamond & \diamond & \diamond & \diamond \\ \diamond & \diamond & \diamond & \diamond \end{bmatrix} \quad (42)$$

where \diamond at row- i and column- j means that the genie does not provide any additional communication capability from source S_i to node j , thus $R_{i,j}$ is not defined. The new network consisting of the additional genie-provided links on top of the original source communication network is given in Fig. 6.

The following theorem is our first result on general network multicast.

Theorem 3: Let \vec{D} be an achievable distortion matrix by joint source-channel coding, for which \vec{O} is the corresponding orders induced by \vec{D} . For any random variable $U_{i,j}$ in the Abelian group \mathcal{X}_i , $j = 1, 2, \dots, |\mathcal{Q}_i|$, such that

$$U_{i,O_i(|\mathcal{Q}_i|)} = V_{i,O_i(|\mathcal{Q}_i|)} \quad (43)$$

$$U_{i,O_i(j)} = U_{i,O_i(j+1)} + V_{i,O_i(j)}, \quad (44)$$

where $O_i(j)$ is the j -th node index in the set of \mathcal{Q}_i according to the order O_i , and $V_{i,j}$'s are mutually independent such that $\mathbb{E}d(U_{i,O_i(j)}) \leq D_{i,O_i(j)}$, let the genie-provided links support

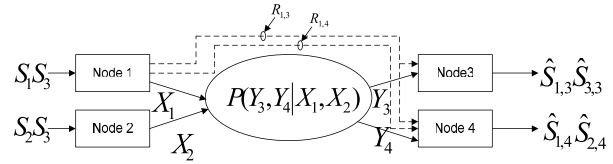


Fig. 6. The example in Fig. 4 with the additional genie-provided links, which are drawn in dashed lines. The region $\mathcal{D}_{mul}^{**}(\vec{O}, \vec{R})$ is the achievable distortion region using a separation-based scheme on this joint network.

the rate matrix $\vec{R}_{\vec{O}}$ whose elements are

$$R_{i,O_i(j)}^* = \begin{cases} C(D_{i,O_i(j)}, U_{i,O_i(j)}) & j \leq |\mathcal{Q}_i|, |\mathcal{Q}_i| > 1 \\ \diamond & \text{otherwise.} \end{cases} \quad (45)$$

Then we have $\vec{D} \in \mathcal{D}_{mul}^{**}(\vec{O}, \vec{R}^*)$.

Remark: This theorem also implies $\bigcup_{\vec{O}} \mathcal{D}_{mul}^*(\vec{O}) \subseteq \mathcal{D}_{mul} \subseteq \bigcup_{\vec{O}} \mathcal{D}_{mul}^{**}(\vec{O}, \vec{R}^*)$. It in fact provides more than one outer bound, one for each set of $V_{i,j}$ random variables, resulting in a rather powerful bounding tool. The auxiliary random variables U_k 's are used in constructing the channel code and the source code, and thus the genie-provided links are also parametrized by these random variables.

For certain distortion measures, significant simplifications can be made. The next result states that a separation-based scheme is approximately optimal, universally across all distortion values, for the quadratic distortion measure where the source alphabet and the reconstruction alphabet are reals.⁵ Note that the sources need not be Gaussian.

Theorem 4: Let \vec{D} be an achievable distortion matrix, for which \vec{O} is the corresponding orders induced by \vec{D} . Let the sources S_i 's satisfy the condition that for all letters $\hat{s}_i \in \hat{S}_i$, $\mathbb{E}(S_i - \hat{s}_i)^2 < \infty$. Let the genie-provided links support the rate matrix $\vec{R}_{\vec{O}}$ whose elements are

$$R_{i,O_i(j)}^* = \begin{cases} 1/2 \text{ bit} & j \leq |\mathcal{Q}_i|, |\mathcal{Q}_i| > 1 \\ \diamond & \text{otherwise} \end{cases}. \quad (46)$$

We have $\vec{D} \in \mathcal{D}_{mul}^{**}(\vec{O}, \vec{R}^*)$ under the mean squared error distortion measure.

In the simplest case where a single node broadcasts a Gaussian source to a set of receivers, this result essentially reduces to Corollary 1 given in [28]. The intuitive translation of the above result is that when a genie helps the separation-based scheme by providing half a bit information for each receiver, and at the same time, all the receivers with better quality reconstructions receive this information for free, then the genie-aided separation-based scheme is as good as the optimal ones. For any fixed network, the approximation in Theorem 4 holds regardless of the quality of the channel. As such, this result is more useful in the high resolution regime for large networks, when the genie-provided links become negligible compared to the original communication network.

⁴If S_i is present at more than one node, i.e., $|\{k : i \in \mathcal{S}_k\}| > 1$, then $R_{i,O_i(j)}$ should be the sum rate per source sample of such common links from each of the node in $\{k : i \in \mathcal{S}_k\}$ to all the nodes $O_i(j), O_i(j+1), \dots, O_i(|\mathcal{Q}_i|)$.

⁵Our proofs for the JSCMUD and JSCMMD problems rely only on weak typicality instead of strong typicality, thus the result can be extended to the continuous sources and channels with continuous alphabets and unbounded distortion measures under the technical condition that for each source S_i , for all letters $\hat{s}_i \in \hat{S}_i$, $\mathbb{E}d(S_i, \hat{s}_i) < \infty$. This "bounded expected distortion" condition [27] assures that the asymptotically small decoding error probability does not cause significant change in the distortion behavior.

In the remainder of the section we focus on the proof of Theorem 3, since Theorem 4 can be directly obtained by using Gaussian auxiliary random variables V 's in Theorem 3.

Proof of Theorem 3: To simplify the notation, let us first consider a single source S ; assume for the time-being that the joint source-channel encoding procedure is still performed on other sources. Without loss of generality, assume the destination nodes of source S are $1, 2, \dots, K$; moreover, the distortions, which are achieved by this given source-channel joint code, are ordered as $D_1 \geq D_2 \geq \dots \geq D_K$.

A set of auxiliary random variables are chosen in the alphabet \mathcal{S} such that,

$$U_K = V_K, \quad U_k = U_{k+1} + V_k, \quad k = 1, 2, \dots, K-1, \quad (47)$$

where V_k 's are random variables in the alphabet \mathcal{S} , *independent of everything else*; furthermore, they have to satisfy $\mathbb{E}d(U_k) \leq D_k$.

Consider a joint source-channel code which induces the distortion vector (D_1, D_2, \dots, D_K) for source S , whose reconstructions are $\hat{S}_1^m, \hat{S}_2^m, \dots, \hat{S}_K^m$. The transition probability $P(\hat{S}_1^m, \hat{S}_2^m, \dots, \hat{S}_K^m | S^m)$ can be viewed as a broadcast channel, denoted as P_{bc} . We need the following lemma, whose proof will be given shortly. The asymptotically small quantities δ, ϵ are omitted in the sequel, which are inconsequential.

Lemma 2: The following degraded message set broadcast rates can be (asymptotically) supported on P_{bc}

$$\begin{aligned} R_1^c &= I(S^m + U_1^m; S^m) - mC(D_1, U_1) \\ R_k^c &= I(S^m + U_k^m; S^m | S^m + U_{k-1}^m) - mC(D_k, U_k), \\ &\quad k = 2, 3, \dots, K. \end{aligned} \quad (48)$$

Moreover, these rates can be achieved by a random superposition code based on the joint distribution $P(S^m + U_1^m, S^m + U_2^m, \dots, S^m + U_{K-1}^m, S^m)$.

Though this lemma is regarding the channel P_{bc} , in a manner similar to the proof for general network unicast, we can conclude that on the original network, when all the other encoders still perform the original joint source-channel encoding, the communication channel from source S to its destinations can support these rates per m source samples. This is because the broadcast channel P_{bc} is simply the original communication channel with certain additional operations on the block level. Thus together with the genie-provided links, we can send messages from source S_i to its destinations at rates

$$\begin{aligned} R_1^{(m)} &= I(S^m + U_1^m; S^m) \\ R_k^{(m)} &= I(S^m + U_k^m; S^m | S^m + U_{k-1}^m), \quad k = 2, 3, \dots, K. \end{aligned} \quad (49)$$

The rates $(R_1^{(m)}, R_2^{(m)}, \dots, R_K^{(m)})$ are exactly the (asymptotic) source coding rates per m -samples in a successive refinement random code [20] constructed using the distribution $P(S^m + U_1^m, S^m + U_2^m, \dots, S^m + U_K^m)$. Thus the distortion $\mathbb{E}d(S + U_k - S) = \mathbb{E}d(U_k) \leq D_k$ is achievable using the separation approach in this genie-aided network, if this successive refinement source code is used.

It remains to argue that if all the users simultaneously replace the original joint source-channel codes with the newly

constructed channel codes, the rates that can be supported are still the same. This is indeed true, because in Lemma 2, we only rely on the joint typicality on the block level when the channel input is of distribution $P(S^m)$. This however does not change if all the users replace the joint source-channel codes with their newly constructed channel codes, since these superposition channel codes preserve the joint typicality according to $P(S_1^m, S_2^m, \dots, S_K^m)$. This completes the proof, except for Lemma 2.

To prove Lemma 2, we first give an auxiliary lemma.

Lemma 3: Let S^m, U_i^m and \hat{S}_i^m be specified as earlier, then we have for $i = 1, 2, \dots, K$

$$\begin{aligned} I(S^m + U_1^m; S^m) - I(S^m + U_1^m; \hat{S}_1^m) \\ \leq mC(D_i, U_1), \end{aligned} \quad (50)$$

$$\begin{aligned} I(S^m + U_k^m; S^m | S^m + U_{k-1}^m) - I(S^m + U_k^m; \hat{S}_i^m | S^m + U_{k-1}^m) \\ \leq mC(D_i, U_k), \quad k = 2, \dots, K-1, \end{aligned} \quad (51)$$

$$\begin{aligned} I(S^m + U_K^m; S^m | S^m + U_{K-1}^m) - I(S^m; \hat{S}_i^m | S^m + U_{K-1}^m) \\ \leq mC(D_i, U_K). \end{aligned} \quad (52)$$

Proof of Lemma 3:

We can write $I(S^m + U_1^m; \hat{S}_i^m, S^m)$ in two ways

$$\begin{aligned} I(S^m + U_1^m; \hat{S}_i^m, S^m) \\ = I(S^m + U_1^m; \hat{S}_i^m) + I(S^m + U_1^m; S^m | \hat{S}_i^m), \end{aligned} \quad (53)$$

$$\begin{aligned} I(S_i^m + U_1^m; \hat{S}_i^m, S^m) \\ = I(S^m + U_1^m; S^m) + I(S^m + U_1^m; \hat{S}_i^m | S^m) \\ = I(S^m + U_1^m; S^m), \end{aligned} \quad (54)$$

where $I(S^m + U_1^m; \hat{S}_i^m | S^m) = I(U_1^m; \hat{S}_i^m | S^m) = 0$, because the construction of the auxiliary random variable U_1 ensures that U_1^m is independent of (\hat{S}_i^m, S^m) , as seen in (47). Thus we have

$$\begin{aligned} I(S^m + U_1^m; S^m) - I(S^m + U_1^m; \hat{S}_i^m) \\ = I(S^m U_1^m; S^m | \hat{S}_i^m) \\ \stackrel{(a)}{=} H(S^m + U_1^m | \hat{S}_i^m) - H(U_1^m) \\ \leq H(S^m - \hat{S}_i^m + U_1^m) - H(U_1^m) \\ \leq \sum_{j=1}^m H(S(j) - \hat{S}_i(j) + U_1(j)) - H(U_1(j)) \\ = \sum_{j=1}^m I(S(j) - \hat{S}_i(j); S(j) - \hat{S}_i(j) + U_1(j)) \\ \leq mC(D_i, U_1), \end{aligned} \quad (55)$$

where (a) follows again since U_1^m is independent of (\hat{S}_i^m, S^m) , and the last step follows the concavity of $I(X; Y)$ as a function of the marginal distribution. This proves (50).

Note further that for $k = 2, 3, \dots, K$, we have

$$\begin{aligned} I(S^m + U_k^m; S^m, \hat{S}_i^m | S^m + U_{k-1}^m) \\ = I(S^m + U_k^m; S^m | S^m + U_{k-1}^m) \\ \quad + I(S^m + U_k^m; \hat{S}_i^m | S^m, S^m + U_{k-1}^m) \\ = I(S^m + U_k^m; S^m | S^m + U_{k-1}^m), \end{aligned}$$

as well as

$$\begin{aligned} & I(S^m + U_k^m; S^m, \hat{S}_i^m | S^m + U_{k-1}^m) \\ &= I(S^m + U_k^m; \hat{S}_i^m | S^m + U_{k-1}^m) \\ & \quad + I(S^m + U_k^m; S^m | \hat{S}_i^m, S^m + U_{k-1}^m). \end{aligned}$$

It follows that

$$\begin{aligned} & I(S^m + U_k^m; S^m | S^m + U_{k-1}^m) \\ & \quad - I(S^m + U_k^m; \hat{S}_i^m | S^m + U_{k-1}^m) \\ &= I(S^m + U_k^m; S^m | \hat{S}_i^m, S^m + U_{k-1}^m). \end{aligned} \quad (56)$$

Thus we have

$$\begin{aligned} & I(S^m + U_k^m; S^m | \hat{S}_i^m, S^m + U_{k-1}^m) \\ & \stackrel{(b)}{=} H(S^m | \hat{S}_i^m, S^m + U_{k-1}^m) - H(S^m | \hat{S}_i^m, S^m + U_k^m) \\ & \leq H(S^m | \hat{S}_i^m) - H(S^m | \hat{S}_i^m, S^m + U_k^m) \\ & = I(S^m; S^m + U_k^m | \hat{S}_i^m) \\ & = H(S^m + U_k^m | \hat{S}_i^m) - H(U_k^m) \\ & \leq H(S^m - \hat{S}_i^m + U_k^m) - H(U_k^m) \\ & \leq mC(D_i, U_k), \end{aligned} \quad (57)$$

where (b) is due to the Markov string $S^m + U_{k-1}^m \leftrightarrow S^m + U_k^m \leftrightarrow S^m \leftrightarrow \hat{S}_i^m$. This proves (51).

Because of the Markov string $\hat{S}_i^m \leftrightarrow S^m \leftrightarrow S^m + U_K^m \leftrightarrow S^m + U_{K-1}^m$, we have

$$\begin{aligned} & I(S^m; \hat{S}_i^m | S^m + U_{K-1}^m) - I(S^m + U_K^m; \hat{S}_i^m | S^m + U_{K-1}^m) \\ &= H(\hat{S}_i^m | S^m + U_{K-1}^m, S^m + U_K^m) \\ & \quad - H(\hat{S}_i^m | S^m + U_{K-1}^m, S^m) \\ &= H(\hat{S}_i^m | S^m + U_K^m) - H(\hat{S}_i^m | S^m) \\ &= H(\hat{S}_i^m | S^m + U_K^m) - H(\hat{S}_i^m | S^m, S^m + U_K^m) \\ &= I(\hat{S}_i^m; S^m | S^m + U_K^m) \geq 0, \end{aligned} \quad (58)$$

and it follows that

$$\begin{aligned} & I(S^m + U_K^m; S^m | S^m + U_{K-1}^m) - I(S^m; \hat{S}_i^m | S^m + U_{K-1}^m) \\ & \leq I(S^m + U_K^m; S^m | S^m + U_{K-1}^m) \\ & \quad - I(S^m + U_K^m; \hat{S}_i^m | S^m + U_{K-1}^m) \\ & \leq mC(D_i, U_K). \end{aligned} \quad (59)$$

This proves (52). \blacksquare

Proof of Lemma 2:

We shall use the distribution $P(S^m + U_1^m, S^m + U_2^m, \dots, S^m + U_{K-1}^m, S^m)$ to construct superposition broadcast channel code on the broadcast channel P_{bc} for a degraded message set. The rates (per length- m block) for these messages within the degraded message set are (asymptotically)

$$R_1^c = I(S^m + U_1^m; S^m) - mC(D_1, U_1), \quad (60)$$

$$\begin{aligned} R_i^c &= I(S^m + U_i^m; S^m | S^m + U_{i-1}^m) - mC(D_i, U_i), \\ & \quad i = 2, 3, \dots, K, \end{aligned} \quad (61)$$

which need to be shown to be indeed achievable on P_{bc} .

Since this channel itself is not degraded, we have to show that the superposition coding scheme succeeds for all the

receivers. To see this, observe that for the i -th receiver, we have

$$I(S^m + U_1^m; S^m) - I(S^m + U_1^m; \hat{S}_i^m) \leq mC(D_i, U_1), \quad (62)$$

by Lemma 3. It follows that

$$\begin{aligned} & I(S^m + U_1^m; \hat{S}_i^m) - R_1^c \\ &= I(S^m + U_1^m; \hat{S}_i^m) - I(S^m + U_1^m; S^m) + mC(D_1, U_1) \\ & \geq mC(D_1, U_1) - mC(D_i, U_1) \geq 0, \end{aligned} \quad (63)$$

where the last inequality is straightforward by noticing

$$C(D, N) \geq C(D', N), \quad (64)$$

when $D \geq D'$. Thus the i -th receiver, $i \geq 1$, can indeed decode the first message.

Similarly, we have for $i \geq k$

$$\begin{aligned} & I(S^m + U_k^m; \hat{S}_i^m | S^m + U_{k-1}^m) - R_k^c \\ &= I(S^m + U_k^m; \hat{S}_i^m | S^m + U_{k-1}^m) \\ & \quad - I(S^m + U_k^m; S^m | S^m + U_{k-1}^m) + mC(D_k, U_k) \\ & \geq mC(D_k, U_k) - mC(D_i, U_k) \geq 0, \end{aligned} \quad (65)$$

and thus we conclude the i -th receiver can decode the messages $1, 2, \dots, i$. The K -th receiver does not pose any additional difficulty. Thus indeed the rates specified in (60)-(61) can be supported on P_{bc} , and the proof is complete. \blacksquare

VII. CONCLUSION

We considered the optimality of the source-channel separation architecture in networks, and showed that the separation approach is optimal for the problems of distributed network joint source-channel coding and joint source-channel multiple unicast with distortions. Moreover, the separation approach is also approximately optimal for the problem of joint source-channel multiple multicast with distortions under certain distortion measures. The results in this work are obtained without explicit characterizations of the underlying regions. The source coding problem extracted from the distributed network source coding scenario implies that the interactive coding aspect needs to be carefully incorporated, which suggests a distinct line of research direction into network source coding.

For notational and conceptual simplicity, we made many assumptions which are not strictly necessary. We believe the results can be extended to more general cases with some minimal efforts.

• Distributed network joint source-channel coding:

The synchronization requirement among sources can be removed, *i.e.*, the source bandwidths do not have to be the same throughout the whole network. The reconstructions of a source S_i can be under different distortion measures; in fact the distortion measures can be defined on multiple sources, such as to reconstruct $(S_1 - S_2)$. The restriction on the sources and the channels being finite-alphabet may be relaxed using the techniques in [29].

• Joint source-channel multiple unicast with distortions:

The synchronization requirement among sources and channels can be removed and the memoryless requirement on the channel can be relaxed to channels with

finite memory (see [15] for an outline). As mentioned, the restriction on the finite alphabets can be relaxed. The condition that each source is to be reconstructed at one destination can be relaxed to some extent: when each source is to be reconstructed at multiple destinations but at the exact same distortion, then the source-channel separation architecture is still optimal.

- **Joint source-channel multiple multicast with distortions:** Similar to the JSCMUD case, the synchronization, the memoryless channel, and the finite-alphabet requirement can be relaxed. The condition that each source is to be reconstructed under the same distortion measure can be relaxed to different distortion measures. If some of the reconstructions of a source S_i are specified to have the same distortion a priori, then the approximation upper bound can be improved.

In the point-to-point setting, the source and the channel are specified by their statistical behaviors alone; however in the network setting, the new components of the connectivity structure among nodes and the source-demand coding requirements are introduced. Our result in DNJSCC treats the source statistics and these network components as a whole, and the channel statistics as the other, resulting in the separation between a complex network source coding problem and multiple conventional point-to-point channel coding problems. In contrast, the result in JSCMUD treats the channel statistics and the network components as a whole, and the source statistics as the other, resulting in the separation between a complex network channel coding problem and multiple conventional point-to-point source coding problems. These separations are not the only possibilities, and one can choose to separate in a different manner. In this work we have not considered transmitting generally correlated sources over a general channel network, and it is unclear whether there exist scenarios for which a separation architecture is optimal or approximately optimal. Thus the problem of source-channel separation is by no means solved, and it in fact calls for further investigation.

ACKNOWLEDGMENT

The authors wish to thank Associate Editor Young-Han Kim and the anonymous reviewers for their constructive comments.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell. Syst. Tech. J.*, vol. 27, pp. 379–423, pp. 623–656, Oct. 1948.
- [2] T. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.
- [3] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. 11, no. 4, pp. 558–567, Oct. 1965.
- [4] J.-J. Xiao and Z.-Q. Luo, "Multiterminal source-channel communication over an orthogonal multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3255–3264, Sep. 2007.
- [5] T. S. Han, "Multicasting correlated multi-source to multi-sink over a network," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 4–13, Jan. 2011.
- [6] T. S. Han, "Slepian-Wolf-cover theorem for a network of channels," *Inform. Control*, vol. 47, no. 1, pp. 67–83, 1980.
- [7] R. Yeung, "Separation principle in multi-user communication systems," private communication, Apr. 2010.

- [8] Y. Steinberg and N. Merhav, "On hierarchical joint source-channel coding with degraded side information," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 886–903, Mar. 2006.
- [9] A. Maor and N. Merhav, "Two-way successively refined joint source-channel coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1483–1494, Apr. 2006.
- [10] R. Koetter, M. Effros, and M. Medard, "A theory of network equivalence—Part I: Point-to-point channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, Feb. 2011.
- [11] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
- [12] S. Jalali and M. Effros, "On the separation of lossy source-network coding and channel coding in wireline networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 500–504.
- [13] C. Tian, J. Chen, S. Diggavi, and S. Shamai (Shitz), "Optimality and approximate optimality of source-channel separation in networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 495–499.
- [14] C. Tian, J. Chen, S. Diggavi, and S. Shamai (Shitz), "On source-channel separation in networks," in *Proc. IEEE Int. Conf. Signal Process. Commun.*, Jul. 2010, pp. 1–5.
- [15] C. Tian, J. Chen, S. Diggavi, and S. Shamai (Shitz), (2010, Apr.). *Optimality and Approximate Optimality of Source-Channel Separation in Networks* [Online]. Available: <http://arxiv.org/abs/1004.2648v1>
- [16] M. Agarwal, A. Sahai, and S. Mitter, "Coding into a source: A direct inverse rate-distortion theorem," in *Proc. 44th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2006, pp. 569–578.
- [17] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 14690–1482, Apr. 2006.
- [18] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York, NY, USA: Springer-Verlag, 1978, pp. 171–231.
- [19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.
- [20] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.
- [21] J. Korner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, Mar. 1977.
- [22] T. Berger, *Rate Distortion Theory: Mathematical Basis for Data Compression*. Upper Saddle River, NJ, USA: Prentice-Hall, 1971.
- [23] M. U. Chang, "Rate-distortion with a fully informed decoder and a partially informed encoder," Ph.D. dissertation, School Electr. Eng., Cornell University, Ithaca, NY, USA, 1978.
- [24] A. Kaspi, "Rate distortion for correlated sources with partially separated encoders," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 828–840, Nov. 1982.
- [25] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School Electr. Eng., Cornell Univ., Ithaca, NY, USA, 1978.
- [26] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.
- [27] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder II: General sources," *Inform. control*, vol. 38, no. 1, pp. 60–80, 1978.
- [28] C. Tian, S. N. Diggavi, and S. Shamai (Shitz), "Approximate characterizations for the Gaussian broadcasting distortion region," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 124–136, Jan. 2011.
- [29] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Press, 2011.

Chao Tian (S'00–M'05–SM'12) received the B.E. degree in Electronic Engineering from Tsinghua University, Beijing, China, in 2000 and the M.S. and Ph. D. degrees in Electrical and Computer Engineering from Cornell University, Ithaca, NY in 2003 and 2005, respectively.

Dr. Tian was a postdoctoral researcher at Ecole Polytechnique Federale de Lausanne (EPFL) from 2005 to 2007. He joined AT&T Labs-Research, Florham Park, New Jersey in 2007, where he is now a Senior Member of Technical Staff. He is currently also an Adjunct Associate Professor at Columbia University and an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. His research interests include multi-user information theory, joint source-channel coding, quantization design and analysis, as well as image/video coding and processing.

Dr. Tian received the Liu Memorial Award at Cornell University in 2004, and the AT&T Key Contributor Award in 2010 and 2011.

Jun Chen (S'03–M'06) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from 2005 to 2006, and a Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from 2006 to 2007. He is currently an Associate Professor of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. His research interests include information theory, wireless communications, and signal processing.

He received several awards for his research, including the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, and the IBM Faculty Award in 2010.

Suhas N. Diggavi (S'93–M'99–F'13) received the B. Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1998.

After completing his Ph.D., he was a Principal Member Technical Staff in the Information Sciences Center, AT&T Shannon Laboratories, Florham Park, NJ. Since then he had been in the faculty of the School of Computer and Communication Sciences, EPFL, where he directed the Laboratory for Information and Communication Systems (LICOS). He is currently a Professor, in the Department of Electrical Engineering, at the University of California, Los Angeles. His research interests include wireless communications networks, information theory, network data compression and network algorithms.

He is a co-recipient of the 2013 IEEE Information Theory Society & Communications Society Joint Paper Award, the 2013 ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) best paper award, the 2006 IEEE Donald Fink prize paper award, 2005 IEEE Vehicular Technology Conference best paper award, the Okawa foundation research award and is a Fellow of the IEEE. He has served on the editorial board for the IEEE TRANSACTIONS ON INFORMATION THEORY, the ACM/IEEE TRANSACTIONS ON NETWORKING, and the IEEE COMMUNICATION LETTERS, a guest editor for the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING and was the Technical Program Co-Chair for 2012 IEEE Information Theory Workshop. He has 8 issued patents.

Shlomo Shamai (Shitz) (S'80–M'82–SM'89–F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion-Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975–1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai is an IEEE Fellow, a member of the Israeli Academy of Sciences and Humanities and a Foreign Associate of the US National Academy of Engineering. He is the recipient of the 2011 Claude E. Shannon Award and the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering. He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission FP7, Network of Excellence in Wireless Communications (NEWCOM++) Best Paper Award, and the 2010 Thomson Reuters Award for International Excellence in Scientific Research. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He is a member of the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY.