

Hybrid Coding: An Interface for Joint Source–Channel Coding and Network Communication

Paolo Minero, Sung Hoon Lim, and Young-Han Kim

May 7, 2014

Abstract

A new approach to joint source–channel coding is presented in the context of communicating correlated sources over multiple access channels. Similar to the separation architecture, the joint source–channel coding system architecture in this approach is modular, whereby the source encoding and channel decoding operations are decoupled. However, unlike the separation architecture, the same codeword is used for both source coding and channel coding, which allows the resulting hybrid coding scheme to achieve the performance of the best known joint source–channel coding schemes. Applications of the proposed architecture to relay communication are also discussed.

Index Terms

Analog/digital coding, hybrid coding, joint source–channel coding, network information theory, relay networks.

Paolo Minero is with the Department of Electrical Engineering of the University of Notre Dame, Notre Dame, IN, 46556, USA (Email: pminero@nd.edu). Sung Hoon Lim was with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea, and is now with the Samsung Advanced Institute of Technology, Yongin-si, Gyeonggi-do, Korea (Email: sunghlim@kaist.ac.kr). Young-Han Kim is with the Department of Electrical and Computer Engineering, University of California, San Diego CA, 92093, USA (Email: yhk@ucsd.edu).

This work was supported in part by the National Science Foundation Grant CCF-1117728.

This work was presented in part at the 2010 Allerton Conference on Communication, Control, and Computing and at the 2011 International Symposium on Information Theory (ISIT 2011).

I. INTRODUCTION

The fundamental architecture of most of today’s communication systems is inspired by Shannon’s source–channel separation theorem [1], [2]. This fundamental theorem states that a source can be optimally communicated over a point-to-point channel by concatenating an optimal source coder that compresses the source into “bits” at the rate of its entropy (or rate–distortion function) with an optimal channel coder that communicates those “bits” reliably over the channel at the rate of its capacity. The appeal of Shannon’s separation theorem is twofold. First, it suggests a simple system architecture in which source coding and channel coding are separated by a universal digital interface. Second, it guarantees that this separation architecture does not incur any asymptotic performance loss.

The optimality of the source–channel separation architecture, however, does not extend to communication systems with multiple users. Except for a few special network models in which sources and channels are suitably “matched” [3]–[8], the problem of lossy communication over a general multiuser network requires the joint optimization of the source coding and channel coding operations. Consequently, there is a vast body of literature on joint source–channel coding schemes for multiple access channels [9]–[16], broadcast channels [17]–[27], interference channels [28], [29], and other multiuser channels [30]–[32].

This paper takes a new approach to studying the problem of lossy communication of correlated sources over networks. We start by revisiting the problem of transmitting a source over a point-to-point channel, for which we propose a hybrid analog/digital scheme for joint source–channel coding that generalizes both the digital, separate source and channel coding scheme and the analog, uncoded transmission scheme. The proposed hybrid coding scheme employs the architecture depicted in Fig. 1 that has the following features:

- 1) A single code performs both source coding and channel coding.
- 2) An encoder generates a (digital) codeword from the (analog) source and selects the channel input as a symbol-by-symbol function of the codeword and the source.
- 3) A decoder recovers the (digital) codeword from the (analog) channel output and selects the source estimate as a symbol-by-symbol function of the codeword and the channel output.

The basic components in this architecture are not new. The idea of using a single code for performing both source coding and channel coding appears, for instance, in the celebrated

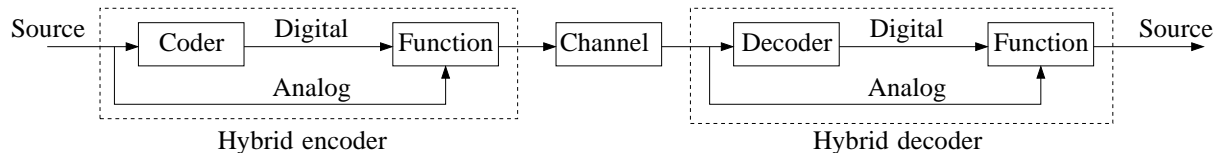


Fig. 1. A joint source–channel coding system architecture based on hybrid coding.

coding scheme by Gelfand and Pinsker [33] for channels with state. The use of symbol-by-symbol functions for the channel input and the source estimate is reminiscent of the Shannon strategy [34] for channels with states and of the Wyner–Ziv coding scheme [35] for lossy source coding with side information. Finally, several hybrid analog–digital communication schemes have been proposed for joint source–channel coding over Gaussian channels; see e.g., [36], where the channel input is formed by a combination of digital and analog information.

The main contribution of this paper lies in combining all these known techniques into a unifying framework that can be used to construct coding schemes for various network communication scenarios. One of the most appealing features of the resulting hybrid coding schemes is that the first-order performance analysis can be done by separately studying the conditions for source coding and for channel coding, exactly as in Shannon’s separation theorem. Furthermore, despite its simplicity, hybrid coding yields the best known performance for a general lossy network communication problem (except sui generis examples such as [37]). We illustrate the advantages of hybrid coding by focusing on two specific problems:

- 1) *Joint source–channel coding over multiple access channels.* We construct a joint source–channel coding scheme for lossy communications over multiple access channels whereby each encoder/decoder in the network operates according to the hybrid coding architecture in Fig. 1. We establish a sufficient condition for lossy communications over multiple access channels that recover and generalize several existing results on joint source–channel coding over this channel model. We also discuss applications of hybrid coding to other channel models such as broadcast channels, interference channels, and channels with feedback.
- 2) *Relay networks.* We apply hybrid coding beyond joint source–channel coding and propose a new (channel) coding scheme for noisy relay networks. This coding scheme operates in a similar manner to the noisy network coding scheme proposed in [38], except that each relay node uses the hybrid coding interface to transmit a symbol-by-symbol function

of the received sequence and its compressed version. This coding scheme unifies both amplify–forward [39] and compress–forward [40], and can strictly outperform both. The potential of the hybrid coding interface for relaying is demonstrated through two specific examples—communication over a two–way relay channel [41] and over a diamond relay network [39].

The remaining of the paper is organized as follows. Section II is devoted to the problem of communicating a source over a point-to-point channel. In Section III, we consider the problem of joint source–channel coding over multiple access channels. In Section IV, we apply hybrid coding to communication over noisy relay networks. Section V concludes the paper.

Throughout we closely follow the notation in [42]. In particular, for a discrete random variable $X \sim p(x)$ on an alphabet \mathcal{X} and $\epsilon \in (0, 1)$, we define the set of ϵ -typical n -sequences x^n (or the typical set in short) [43] as $\mathcal{T}_\epsilon^{(n)}(X) = \{x^n : |\#\{i : x_i = x\}/n - p(x)| \leq \epsilon p(x) \text{ for all } x \in \mathcal{X}\}$. We use $\delta(\epsilon) > 0$ to denote a generic function of $\epsilon > 0$ that tends to zero as $\epsilon \rightarrow 0$. Similarly, we use $\epsilon_n \geq 0$ to denote a generic sequence in n that tends to zero as $n \rightarrow \infty$.

II. POINT-TO-POINT CHANNELS

Consider the point-to-point communication system depicted in Fig. 2, where a sender wishes to communicate n symbols of a discrete memoryless source (DMS) $S \sim p(s)$ over the discrete memoryless channel (DMC) $p(y|x)$ in n transmissions so that the receiver can reconstruct the source symbols with a prescribed distortion D .

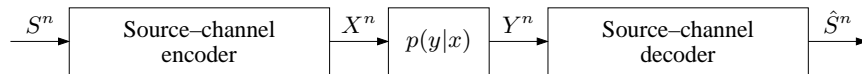


Fig. 2. Point-to-point communication system.

An $(|\mathcal{S}|^n, n)$ joint source–channel code consists of

- an encoder that assigns a sequence $x^n(s^n) \in \mathcal{X}^n$ to each sequence $s^n \in \mathcal{S}^n$ and
- a decoder that assigns an estimate $\hat{s}^n \in \hat{\mathcal{S}}^n$ to each sequence $y^n \in \mathcal{Y}^n$.

Let $d(s, \hat{s})$ be a nonnegative distortion measure that quantifies the cost of representing a symbol s by a symbol \hat{s} . A distortion D is said to be achievable for communication of the DMS S over

the DMC $p(y|x)$ if there exists a sequence of $(|\mathcal{S}|^n, n)$ joint source–channel codes such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}(d(S_i, \hat{S}_i)) \leq D.$$

Shannon’s source–channel separation theorem [1], [2] shows that D is achievable if

$$R(D) < C \tag{1}$$

and only if $R(D) \leq C$, where

$$R(D) = \min_{p(\hat{s}|s): \mathbb{E}(d(S, \hat{S})) \leq D} I(S; \hat{S})$$

is the rate–distortion function for the source S and the distortion measure $d(s, \hat{s})$, and

$$C = \max_{p(x)} I(X; Y)$$

is the capacity of the channel $p(y|x)$. The proof of achievability is based on separate lossy source coding and channel coding.

In this section, we establish the following alternative characterization of the set of achievable distortions.

Theorem 1 (Shannon [1], [2]): A distortion D is achievable for communication of the DMS S over the DMC $p(y|x)$ if

$$I(S; U) < I(U; Y) \tag{2}$$

for some conditional pmf $p(u|s)$, channel encoding function $x(u, s)$, and source decoding function $\hat{s}(u, y)$ such that $\mathbb{E}(d(S, \hat{S})) \leq D$.

In the rest of this section, we first describe a joint source–channel coding scheme that is based on the hybrid-coding architecture in Fig. 1 and then provide a formal proof of Theorem 1 using the proposed coding scheme.

A. Hybrid Coding Architecture

The proposed joint source–channel coding scheme can be described by the block diagram depicted in Fig. 3. A source encoder maps the source sequence S^n into a sequence $U^n(M)$ from a randomly generated codebook $\mathcal{C} = \{U^n(m) : m \in [1 : 2^{nR}]\}$ of independent and identically distributed codewords. The selected sequence and the source S^n are then mapped *symbol-by-symbol* through an encoding function $x(s, u)$ to a sequence X^n that is transmitted over the

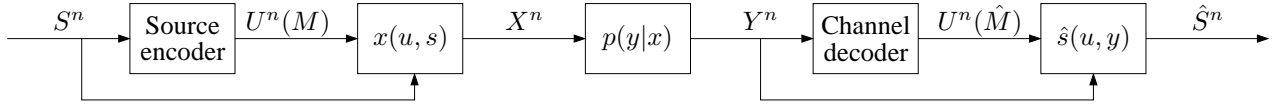


Fig. 3. The hybrid coding architecture for point-to-point communication.

channel. Upon receiving the channel output sequence Y^n , the channel decoder finds an estimate $U^n(\hat{M})$ of $U^n(M)$ and reconstructs \hat{S}^n from $U^n(\hat{M})$ and Y^n again by a *symbol-by-symbol* mapping $\hat{s}(u, y)$.

Comparing the general architecture in Fig. 1 with the one in Fig. 3, we notice that the codeword $U^n(M)$ encodes the (digital) compression index $M \in [1 : 2^{nR}]$ that has to be reliably transmitted over the channel. At the same time, $U^n(M)$ is the input sequence transmitted over the channel $p(y|u) = \sum_s p(y|x(u, s))p(s|u)$. Hence, $U^n(M)$ plays the roles of both the source codeword that compresses the source sequence within the desired distortion and the channel codeword that encodes the compression index sent across the channel.

The proposed architecture generalizes both Shannon's source–channel separation architecture and the uncoded analog transmission architecture, as shown next.

a) *Shannon's source–channel separation* [1], [2]: Under this architecture the source sequence is mapped into a compression index $M \in [1 : 2^{nR}]$, which is then mapped into a channel codeword X^n to be transmitted over the channel. Upon receiving Y^n , the decoder finds an estimate \hat{M} of the message M and reconstructs $\hat{S}^n(\hat{M})$ from \hat{M} .

Suppose that in Theorem 1 we set $U = (X, \hat{S})$, where $\hat{S} \sim p(\hat{s}|s)$ and $X \sim p(x)$ is independent of S and \hat{S} , $x(u, s) = x$, and $\hat{s}(u, y) = \hat{s}$, so the codeword $U^n(M)$ consists of the source codeword $\hat{S}^n(M)$ as well as the channel codeword $X^n(M)$, the source encoding functions sets the channel input equal to $X^n(M)$, and the source decoding function recovers $\hat{S}^n(\hat{M})$ from the estimate \hat{M} for the transmitted index M . In this case the proposed hybrid coding architecture reduces to the source–channel separation architecture depicted in Fig. 4.

It can be easily checked that (2) simplifies to $R(D) < C$, that is, Theorem 1 recovers (1).

b) *Uncoded transmission*: Under this architecture the source sequence S^n is mapped through a symbol-by-symbol encoding function $x(s)$ into a channel codeword X^n that is transmitted over the channel. Upon receiving Y^n , the decoder forms an estimate \hat{S}^n of the trans-

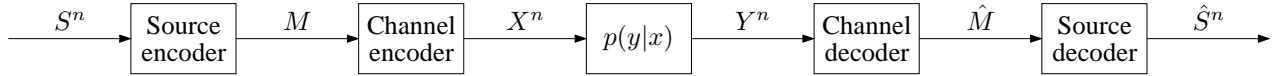


Fig. 4. Separate source and channel coding system architecture.

mitted source again through a symbol-by-symbol source decoding function $\hat{s}(y)$. Despite its simplicity, uncoded transmission can be sometimes optimal [3], for example, when communicating a Gaussian source over a Gaussian channel under the quadratic distortion measure [44] or communicating a binary source over a binary symmetric channel under the Hamming distortion measure. In both cases, the desired distortion D can be achieved if $C \geq R(D)$. (Note the nonstrict inequality, unlike the strict inequality in Shannon's sufficient condition (1).)

Suppose that in Theorem 1 we set $U = \emptyset$, $x(u, s) = x(s)$, and $\hat{s}(u, y) = \hat{s}(y)$, so the channel input is a symbol-by-symbol function of the source and the source estimate is a symbol-by-symbol function of the channel output. In this case the proposed hybrid coding architecture reduces to the uncoded transmission architecture depicted in Fig. 5, and a distortion D is achievable if there exists $\hat{x}(s)$ and $\hat{s}(y)$ such that $E(d(S, \hat{S})) \leq D$.

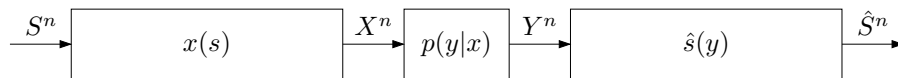


Fig. 5. Uncoded transmission architecture.

The conditions under which a distortion D is achievable can be obtained by studying the conditions for source coding and channel coding separately. Roughly speaking, by the lossy source coding theorem, the condition $R > I(U; S)$, where R is the rate of the codebook \mathcal{C} , guarantees a reliable source encoding operation, and by the channel coding theorem, the condition $R < I(U; Y)$ guarantees a reliable channel decoding operation. Combining the two conditions leads to (2).

The precise performance analysis, however, involves a technical subtlety and requires a careful treatment of the error probability. In particular, because $U^n(M)$ is used as a source codeword, the index M depends on the entire codebook \mathcal{C} . But the conventional random coding proof

technique for a channel codeword $U^n(M)$ is developed for situations in which the index M and the (random) codebook \mathcal{C} are independent of each other. This dependency issue has been well noted by Lapidoth and Tinguely [13, Proof of Proposition D.1], who developed a geometric approach for sending a bivariate Gaussian source over a Gaussian multiple access channel. Here we develop a recipe for the general case.

B. Proof of Theorem 1

We provide a formal proof of the sufficient condition (2) along with a new analysis technique that handles the dependency between the transmitted index and the codebook. The standard proof steps are omitted for brevity and can be found in [42].

Codebook generation: Let $\epsilon > \epsilon' > 0$. Fix a conditional pmf $p(u|s)$, an encoding function $x(u, s)$, and a source reconstruction function $\hat{s}(u, y)$ such that $E(d(S, \hat{S})) \leq D/(1+\epsilon)$. Randomly and independently generate 2^{nR} sequences $u^n(m)$, $m \in [1 : 2^{nR}]$, each according to $\prod_{i=1}^n p_U(u_i)$. The codebook $\mathcal{C} = \{u^n(m) : m \in [1 : 2^{nR}]\}$ is revealed to both the encoder and the decoder.

Encoding: We use joint typicality encoding. Upon observing a sequence s^n , the encoder finds an index m such that $(u^n(m), s^n) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there is more than one such index, it chooses one of them at random. If there is no such index, it chooses an arbitrary index at random from $[1 : 2^{nR}]$. The encoder then transmits $x_i = x(u_i(m), s_i)$ for $i \in [1 : n]$.

Decoding: We use joint typicality decoding. Upon receiving y^n , the decoder finds the unique index \hat{m} such that $(u^n(\hat{m}), y^n) \in \mathcal{T}_{\epsilon}^{(n)}$. If there is none or more than one, it chooses an arbitrary index, say, $\hat{m} = 1$. The decoder then sets the reproduction sequence as $\hat{s}_i = \hat{s}(u_i(\hat{m}), y_i)$ for $i \in [1 : n]$.

Analysis of the expected distortion: We bound the distortion averaged over S^n , the random choice of the codebook \mathcal{C} , and the random index assignment in the encoding procedure. Let M be the random variable denoting the chosen index at the encoder. Define the “error” event

$$\mathcal{E} = \{(S^n, U^n(\hat{M}), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}$$

and partition it into

$$\mathcal{E}_1 = \{(U^n(m), S^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } m\},$$

$$\mathcal{E}_2 = \{(S^n, U^n(M), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\},$$

$$\mathcal{E}_3 = \{(U^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \neq M\}.$$

Then by the union of events bound,

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c) + \mathbb{P}(\mathcal{E}_3). \quad (3)$$

We show that all three terms tend to zero as $n \rightarrow \infty$ under suitable conditions on the codebook rate R . This implies that the probability of “error” tends to zero as $n \rightarrow \infty$, which, in turn, implies that, by the law of total expectation and the typical average lemma [42, Section 2.4],

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{E}(d(S^n, \hat{S}^n)) &\leq \limsup_{n \rightarrow \infty} (\mathbb{P}(\mathcal{E}) \mathbb{E}(d(S^n, \hat{S}^n) | \mathcal{E}) + \mathbb{P}(\mathcal{E}^c) \mathbb{E}(d(S^n, \hat{S}^n) | \mathcal{E}^c)) \\ &\leq (1 + \epsilon) \mathbb{E}(d(S, \hat{S})), \end{aligned}$$

and hence the desired distortion is achieved.

By the covering lemma and the conditional typicality lemma [42, Sections 2.4 and 3.7], it can be easily shown that the first two terms in (3) tend to zero as $n \rightarrow \infty$ if $R > I(U; S) + \delta(\epsilon')$. The third term requires some special attention. By the symmetry of the codebook generation and encoding, we analyze the probability conditioned on the event $M = 1$. By the union of events bound, for n sufficiently large,

$$\begin{aligned} &\mathbb{P}\{(U^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \neq 1 | M = 1\} \\ &\leq \sum_{m=2}^{2^{nR}} \mathbb{P}\{(U^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} | M = 1\} \\ &= \sum_{m=2}^{2^{nR}} \sum_{(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{U^n(m) = u^n, Y^n = y^n | M = 1\} \\ &= \sum_{m=2}^{2^{nR}} \sum_{(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{U^n(m) = u^n | Y^n = y^n, M = 1\} \mathbb{P}\{Y^n = y^n | M = 1\} \\ &= \sum_{m=2}^{2^{nR}} \sum_{(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \sum_{\tilde{u}^n, s^n} \mathbb{P}\{U^n(m) = u^n | U^n(1) = \tilde{u}^n, S^n = s^n, Y^n = y^n, M = 1\} \\ &\quad \cdot \mathbb{P}\{U^n(1) = \tilde{u}^n, S^n = s^n | Y^n = y^n, M = 1\} \mathbb{P}\{Y^n = y^n | M = 1\} \\ &\stackrel{(a)}{=} \sum_{m=2}^{2^{nR}} \sum_{(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \sum_{\tilde{u}^n, s^n} \mathbb{P}\{U^n(m) = u^n | U^n(1) = \tilde{u}^n, S^n = s^n, M = 1\} \\ &\quad \cdot \mathbb{P}\{U^n(1) = \tilde{u}^n, S^n = s^n | Y^n = y^n, M = 1\} \mathbb{P}\{Y^n = y^n | M = 1\} \end{aligned}$$

$$\begin{aligned}
&\stackrel{(b)}{\leq} (1 + \epsilon) \sum_{m=2}^{2^{nR}} \sum_{(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \prod_{i=1}^n p_U(u_i) \mathbb{P}\{Y^n = y^n | M = 1\} \\
&\leq (1 + \epsilon) 2^{n(R - I(U; Y) + \delta(\epsilon))},
\end{aligned}$$

which tends to zero as $n \rightarrow \infty$, if $R < I(U; Y) - \delta(\epsilon)$. Here step (a) follows from the fact that given $M = 1$, $U^n(m) \rightarrow (U^n(1), S^n) \rightarrow Y^n$ form a Markov chain for all $m \neq 1$. To justify step (b), we make use of the following lemma, the proof of which is delegated to Appendix A.

Lemma 1: Let $(U, S) \sim p(u, s)$ and $\epsilon, \epsilon' > 0$. Let $S^n \sim \prod_{i=1}^n p_S(s_i)$ and $U^n(m)$, $m \in [1 : 2^{nR}]$, be independently generated sequences, each drawn according to $\prod_{i=1}^n p_U(u_i)$, independent of S^n . Let $\mathcal{I} = \{m \in [1 : 2^{nR}] : (U^n(m), S^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U, S)\}$ be a set of random indices and let $M \sim \text{Unif}(\mathcal{I})$, if $|\mathcal{I}| > 0$, and $M \sim \text{Unif}([1 : 2^{nR}])$, otherwise. Then, for every (u^n, \tilde{u}^n, s^n) ,

$$\mathbb{P}\{U^n(2) = u^n | U^n(1) = \tilde{u}^n, S^n = s^n, M = 1\} \leq (1 + \epsilon) \cdot \prod_{i=1}^n p_U(u_i)$$

for n sufficiently large.

Step (b) above now follows from the inequality in Lemma 1, which by symmetry holds for all $\hat{m} \neq 1$. Therefore, if $I(U; S) < I(U; Y) - \delta(\epsilon) - \delta(\epsilon')$ the probability of “error” tends to zero as $n \rightarrow \infty$ and the average distortion over the random codebook is bounded as desired. Thus, there exists at least one sequence of codes achieving the desired distortion. By letting $\epsilon \rightarrow 0$, the sufficient condition (2) for lossy communication via hybrid coding is established.

C. Discussion

Similar to the source–channel separation architecture, the proposed hybrid coding architecture is modular, whereby the source encoding and channel decoding operations are decoupled and can be analyzed separately. However, unlike the separation architecture, the same codeword is used for both source coding and channel coding, which allows the resulting scheme to perform joint source–channel coding.

The proposed coding scheme can be readily extended to the case of a source transmitted over a DMC with state or over a compound DMC [45], for which hybrid coding achieves the best known performance, recovering and generalizing several existing results in the literature [19], [26], [27], [33], [36], [46]–[48]. The proposed architecture can also be extended to the case of source–channel bandwidth mismatch, whereby k samples of a DMS are transmitted through n

uses of a DMC. This can be accomplished by replacing the source and channel symbols in Fig. 3 by supersymbols of lengths k and n , respectively.

III. JOINT SOURCE–CHANNEL CODING OVER MULTIPLE ACCESS CHANNELS

In this section, we illustrate how the hybrid coding system architecture described in Section II can be generalized to construct joint source–channel coding schemes for lossy communication over multiuser channels. To illustrate the main ideas, we focus on the specific problem of communicating a pair of correlated discrete memoryless sources (2-DMS) $(S_1, S_2) \sim p(s_1, s_2)$ over a discrete memoryless multiple access channel (DM-MAC) $p(y|x_1, x_2)$, as depicted in Fig. 6. Here each sender $j = 1, 2$ wishes to communicate in n transmissions its source S_j to a common receiver so the sources can be reconstructed within desired distortions.

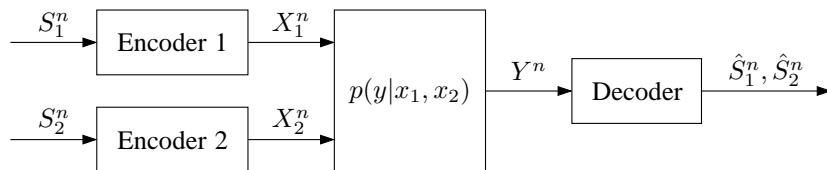


Fig. 6. Communication of a 2-DMS over a DM-MAC.

An $(|\mathcal{S}_1|^n, |\mathcal{S}_2|^n, n)$ joint source–channel code consists of

- two encoders, where encoder $j = 1, 2$ assigns a sequence $x_j^n(s_j^n) \in \mathcal{X}_j^n$ to each sequence $s_j^n \in \mathcal{S}_j^n$, and
- a decoder that assigns an estimate $(\hat{s}_1^n, \hat{s}_2^n) \in \hat{\mathcal{S}}_1^n \times \hat{\mathcal{S}}_2^n$ to each sequence $y^n \in \mathcal{Y}^n$.

Let $d_1(s_1, \hat{s}_1)$ and $d_2(s_2, \hat{s}_2)$ be two nonnegative distortions measures. A distortion pair (D_1, D_2) is said to be achievable for communication of the 2-DMS (S_1, S_2) over the DM-MAC $p(y|x_1, x_2)$ if there exists a sequence of $(|\mathcal{S}_1|^n, |\mathcal{S}_2|^n, n)$ joint source–channel codes such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}(d_j(S_{ji}, \hat{S}_{ji}) \leq D_j, \quad j = 1, 2.$$

The optimal distortion region is the closure of the set of all achievable distortion pairs (D_1, D_2) . A computable characterization of the optimal distortion region is not known in general. Hybrid coding yields the following inner bound on the optimal distortion region.

Theorem 2: A distortion pair (D_1, D_2) is achievable for communication of the 2-DMS (S_1, S_2) over the DM-MAC $p(y|x_1, x_2)$ if

$$\begin{aligned} I(U_1; S_1|U_2, Q) &< I(U_1; Y|U_2, Q), \\ I(U_2; S_2|U_1, Q) &< I(U_2; Y|U_1, Q), \\ I(U_1, U_2; S_1, S_2|Q) &< I(U_1, U_2; Y|Q) \end{aligned}$$

for some pmf $p(q)p(u_1|s_1, q)p(u_2|s_2, q)$ and functions $x_1(q, u_1, s_1)$, $x_2(q, u_2, s_2)$, $\hat{s}_1(q, u_1, u_2, y)$, and $\hat{s}_2(q, u_1, u_2, y)$ such that $E(d_j(S_j, \hat{S}_j)) \leq D_j$, $j = 1, 2$.

The proof of the theorem is given in Appendix B.

Application of Theorem 2 yields the following results as special cases:

- a) *Lossless communication:* When specialized to the case in which d_1 and d_2 are Hamming distortion measures and $D_1 = D_2 = 0$, Theorem 2 recovers the following sufficient condition for lossless communication of a 2-DMS over a DM-MAC.

Corollary 1 (Cover, El Gamal, and Salehi [9]): A 2-DMS (S_1, S_2) can be communicated losslessly over the DM-MAC $p(y|x_1, x_2)$ if

$$\begin{aligned} H(S_1|S_2) &< I(X_1; Y|X_2, S_2, Q), \\ H(S_2|S_1) &< I(X_2; Y|X_1, S_1, Q), \\ H(S_1, S_2) &< I(X_1, X_2; Y|Q) \end{aligned}$$

for some pmf $p(q)p(x_1|s_1, q)p(x_2|s_2, q)$.

Proof: It suffices to choose in Theorem 2 $U_j = (X_j, S_j)$, $x_j(q, u_j, s_j) = x_j$, and $\hat{s}_j(q, u_1, u_2, y) = s_j$, $j = 1, 2$, under a pmf of the form $p(q)p(x_1|s_1, q)p(x_2|s_2, q)$. ■

- b) *Distributed lossy source coding:* When specialized to the case of a noiseless DM-MAC $Y = (X_1, X_2)$ with $\log |\mathcal{X}_1| = R_1$ and $\log |\mathcal{X}_2| = R_2$ and (X_1, X_2) independent of the sources, Theorem 2 recovers the Berger–Tung inner bound on the rate–distortion region for distributed lossy source coding.

Corollary 2 (Berger [49] and Tung [50]): A distortion pair (D_1, D_2) with rate pair (R_1, R_2) is achievable for distributed lossy source coding of a 2-DMS (S_1, S_2) if

$$R_1 > I(S_1; U_1|U_2, Q),$$

$$R_2 > I(S_2; U_2 | U_1, Q),$$

$$R_1 + R_2 > I(S_1, S_2; U_1, U_2 | Q)$$

for some pmf $p(q)p(u_1|s_1, q)p(u_2|s_2, q)$ and functions $\hat{s}_1(q, u_1, u_2)$ and $\hat{s}_2(q, u_1, u_2)$ such that $E(d_j(S_j, \hat{S}_j)) \leq D_j$, $j = 1, 2$.

Proof: It suffices to choose in Theorem 2 $U_j = (X_j, \tilde{U}_j)$, $x_j(q, u_j, s_j) = x_j$, and $\hat{s}_j(q, u_1, u_2, y) = \hat{s}_j(q, \tilde{u}_1, \tilde{u}_2)$, $j = 1, 2$, under a pmf of the form $p(q)p(\tilde{u}_1|s_1, q)p(\tilde{u}_2|s_2, q)$, and to relabel the tilded random variables. ■

- c) *Bivariate Gaussian source over a Gaussian MAC:* Suppose that the source is a bivariate Gaussian pair with equal variance σ^2 and that each source component has to be reconstructed by the decoder under quadratic distortion measures $d_j(s_j, \hat{s}_j) = (s_j - \hat{s}_j)^2$, $j = 1, 2$. In addition, assume that the channel is the Gaussian MAC $Y = X_1 + X_2 + Z$, where Z is AWGN and the channel inputs X_1 and X_2 are subject to average power constraints. Theorem 2 can be adapted to this case via the standard discretization method [42, Sections 3.4 and 3.8]. Suppose that in Theorem 2 we choose (U_1, U_2) as jointly Gaussian random variables conditionally independent given (S_1, S_2) , the encoding function $x_j(u_j, s_j)$, $j = 1, 2$, as a linear function of u_j and s_j , and the decoding function $\hat{s}_j(u_1, u_2, y)$ as the minimum mean-square error (MMSE) estimate of S_j given U_1, U_2 , and Y . Then, Theorem 2 recovers to the sufficient condition for lossy communication derived in Lapidoth and Tinguely [13, Theorem IV.6] via a hybrid analog/digital scheme that combines uncoded transmission and vector quantization.

A. Hybrid Coding Architecture

The joint source–channel coding scheme used in the proof of achievability of Theorem 2 is based on the hybrid coding system architecture depicted in Fig. 7. Here the source sequence S_j^n is mapped by source encoder $j = 1, 2$ into a sequence $U_j^n(M_j)$ from a randomly generated codebook $\mathcal{C}_j = \{U_j^n(m_j) : m_j \in [1 : 2^{nR_j}]\}$ of independently distributed codewords. The selected sequence and the source S_j^n are then mapped symbol-by-symbol through an encoding function $x_j(s_j, u_j)$ to a sequence X_j^n , which is transmitted over the MAC. Upon receiving Y^n , the decoder finds the estimates $U_1^n(\hat{M}_1)$ and $U_2^n(\hat{M}_2)$ of $U_1^n(M_1)$ and $U_2^n(M_2)$, respectively, and reconstructs \hat{S}_1^n and \hat{S}_2^n from $U_1^n(\hat{M}_1)$, $U_2^n(\hat{M}_2)$, and Y^n by symbol-by-symbol mappings

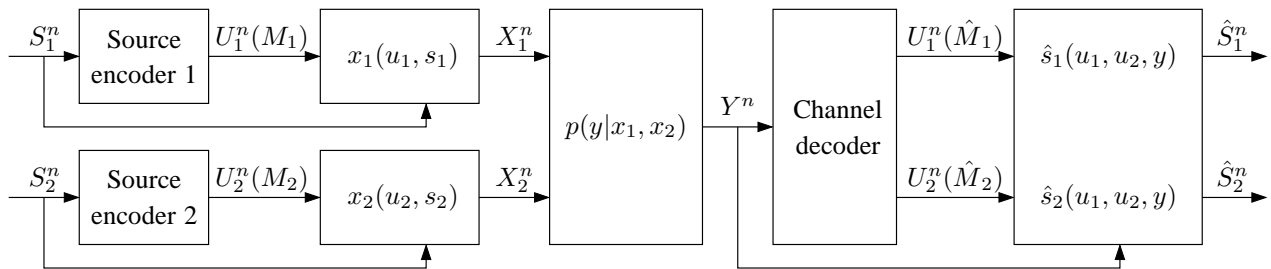


Fig. 7. Joint source–channel coding system architecture for communicating a 2-DMS over a DM-MAC.

$\hat{s}_j(u_1, u_2, y)$, $j = 1, 2$. Comparing Fig. 7 with Fig. 3, we notice that the two separate encoders in Fig. 7 operate exactly as the hybrid encoder in Fig. 3, while the channel decoder at the output of the DM-MAC differs from the one used in the point-to-point case because it *jointly* decodes $U_1^n(\hat{M}_1)$ and $U_2^n(\hat{M}_2)$ and forms the source estimates from the pair $(U_1^n(\hat{M}_1), U_2^n(\hat{M}_2))$ as well as Y^n .

As in the case of point-to-point communication, the conditions under which a distortion pair (D_1, D_2) is achievable can be obtained by studying the conditions for source coding and channel coding separately. By the covering lemma [42, Section 3.7], the source encoding operation is successful if

$$R_1 > I(U_1; S_1),$$

$$R_2 > I(U_2; S_2),$$

while by the packing lemma [42, Section 3.2], suitably modified to account for the dependence between the indices and the codebook that we have mentioned in Section II, the channel decoding operation is successful if

$$R_1 < I(U_1; Y, U_2),$$

$$R_2 < I(U_2; Y, U_1),$$

$$R_1 + R_2 < I(U_1, U_2; Y) + I(U_1; U_2).$$

Then, the sufficient condition in Theorem 2 (with $Q = \emptyset$) is established by combining the above inequalities and eliminating the intermediate rate pair (R_1, R_2) . The sufficient condition with a

general Q can be proved by introducing a time sharing random variable Q and using coded time sharing [42, Section 4.5.3].

B. Remarks

The proposed joint source–channel coding scheme is conceptually similar to separate source and channel coding and, loosely speaking, is obtained by concatenating the source coding scheme in [49], [50] for distributed lossy source coding with a channel code for multiple access communication, except that the same codeword is used by both the source encoder and the channel encoder. Similarly to the coding scheme by Cover, El Gamal, and Salehi in [9] for lossless communication over a DM-MAC, the hybrid coding in Theorem 2 enables coherent communication over the MAC by preserving the correlation between the sources at the channel inputs chosen by the two senders.

The achievable distortion region in Theorem 2 can be increased when the 2-DMS (S_1, S_2) has a nontrivial common part in the sense of Gács–Körner [51] and Witsenhausen [52]. In this case, the encoders can jointly compress the common part and use it to establish coherent communication over the MAC. This extension will be considered elsewhere [45], where a hybrid coding scheme is proposed by combining the distributed lossy source coding scheme in [53] for sources with a nonempty common part and the channel coding scheme in [54] for multiple access communication with a common message shared by the two encoders. The result in Theorem 2 can also be generalized to the setting in which the source consists of a random triple (S, S_1, S_2) , the distortion measures are $d_j : \mathcal{S} \times \mathcal{S}_1 \times \mathcal{S}_2 \rightarrow \hat{\mathcal{S}}_j$, $j = 1, 2$, but encoder j can only observe the source component S_j , $j = 1, 2$. This setting includes as special cases the CEO problem [55]–[58] and the Gaussian sensor network [59].

The modular approach presented here for lossy communications over multiple access channels can be adapted to construct joint source–channel coding schemes for other channel models. In [45], extensions to several canonical channel models studied in the literature will be presented—the broadcast channel, the interference channel, as well as channels with state or noiseless output feedback. In all these examples, we establish sufficient conditions for lossy communication based on hybrid coding. The basic design principle consists in combining a source coding scheme with a suitably “matched” channel coding scheme by the means of the hybrid coding architecture described in Section II. For instance, in the case of lossy communication

over broadcast channels, a hybrid coding scheme can be constructed by concatenating the Gray–Wyner lossy source coding scheme [35] with the Marton coding scheme [60] for the general broadcast channel with a common message.

IV. RELAY NETWORKS

In this section we explore applications of hybrid coding in the context of relay networks, wherein a source node wishes to send a message to a destination node with the help of intermediate relay nodes. Over the past decades, three dominant paradigms have been proposed for relay communication: decode–forward, compress–forward, and amplify–forward.

- In decode–forward, each relay recovers the transmitted message by the source either fully or partially and forwards it to the receiver (digital-to-digital interface) while coherently cooperating with the source node. Decode–forward was originally proposed in [40] for the relay channel and has been generalized to multiple relay networks, for example, in [61], [62] and further improved by combining it with structured coding [63], [64].
- In amplify–forward, each relay sends a scaled version of its received sequence and forwards it to the receiver (analog-to-analog interface). Amplify–forward was proposed in [39] for the Gaussian two–relay diamond network and subsequently studied for the Gaussian relay channel in [65]. Generalizations of amplify–forward to general nonlinear analog mappings for relay communication have been proposed in [66].
- In compress–forward, each relay vector-quantizes its received sequence and forwards it to the receiver (analog-to-digital interface). Compress–forward was proposed in [40] for the relay channel and has been generalized to arbitrary noisy networks in [38] as noisy network coding.

In this section we propose a new coding scheme for relay networks that uses hybrid analog/digital coding at the relay nodes. The proposed scheme naturally extends both amplify–forward and compress–forward since each relay node uses the hybrid coding architecture introduced in Section II to transmit a symbol-by-symbol function of the received sequence and its quantized version (analog-to-analog/digital interface). More important than this conceptual unification is the performance improvement of hybrid coding. We demonstrate through two specific examples, the two–way relay channel (Section IV-A) and the two–relay diamond network

(Section IV-B), that hybrid coding can strictly outperform the existing coding schemes, not only amplify–forward and compress–forward, but also decode–forward.

A. Two–Way Relay Channel

Consider the relay network depicted in Fig. 8, where two source/destination nodes communicate with each other with the help of one relay. Node 1 wishes to send the message $M_1 \in [1 : 2^{nR_1}]$ to node 2 and node 2 wishes to send the message $M_2 \in [1 : 2^{nR_2}]$ to node 1 with the help of the relay node 3. Nodes 1 and 2 are connected to the relay through the MAC $p(y_3|x_1, x_2)$, while the relay is connected to nodes 1 and 2 via the broadcast channel $p(y_1, y_2|x_3)$.

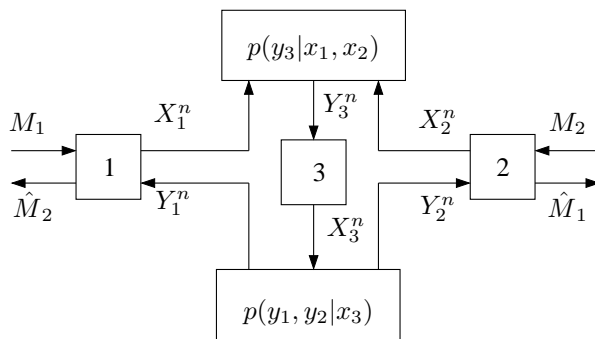


Fig. 8. The two–way relay channel.

This network is modeled by a 3-node discrete memoryless two–way relay channel (DM-TWRC) $p(y_1, y_2|x_3)p(y_3|x_1, x_2)$. A $(2^{nR_1}, 2^{nR_2}, n)$ code for the DM-TWRC consists of

- two message sets $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$,
- two encoders, where at time $i \in [1 : n]$ encoder $j = 1, 2$ assigns a symbol $x_{j,i}(m_j, y_j^{i-1}) \in \mathcal{X}_j$ to each message $m_j \in [1 : 2^{nR_j}]$ and past received output sequence $y_j^{i-1} \in \mathcal{Y}_j^{i-1}$,
- a relay encoder that assigns a symbol $x_{3,i}(y_3^{i-1})$ to each past received output sequence $y_3^{i-1} \in \mathcal{Y}_3^{i-1}$, and
- two decoders, where decoder 1 assigns an estimate \hat{m}_2 or an error message to each received sequence $y_1^n \in \mathcal{Y}_1^n$ and decoder 2 assigns an estimate \hat{m}_1 or an error message to each received sequence $y_2^n \in \mathcal{Y}_2^n$.

We assume that the message pair (M_1, M_2) is uniformly distributed over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. The average probability of error is defined as $P_e^{(n)} = \mathbb{P}\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\}$. A rate pair (R_1, R_2) is said to be achievable for the DM-TWRC if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n \rightarrow \infty} P_e^{(n)} = 0$. The capacity region of the DM-TWRC is the closure of the set of achievable rate pairs (R_1, R_2) and the sum-capacity is the supremum of the achievable sum rates $R_1 + R_2$.

The capacity region of the DM-TWRC $p(y_1, y_2|x_3)p(y_3|x_1, x_2)$ is not known in general. Rankov and Wittneben [41] characterized inner bounds on the capacity region based on decode–forward, compress–forward, and amplify–forward. Another inner bound based on noisy network coding is given in [38]. In the special case of a Gaussian TWRC, Nam, Chung, and Lee [64] proposed a coding scheme based on nested lattice codes and structured binning that achieves within 1/2 bit per dimension from the capacity region for all underlying channel parameters.

Hybrid coding yields the following inner bound on the capacity region, the proof of which is given in Appendix C.

Theorem 3: A rate pair (R_1, R_2) is achievable for the DM-TWRC $p(y_1, y_2|x_3)p(y_3|x_1, x_2)$ if

$$\begin{aligned} R_1 &< \min(I(X_1; Y_2, U_3|X_2), I(X_1, U_3; X_2, Y_2) - I(Y_3; U_3|X_1)), \\ R_2 &< \min(I(X_2; Y_1, U_3|X_1), I(X_2, U_3; X_1, Y_1) - I(Y_3; U_3|X_1)), \end{aligned} \quad (4)$$

for some pmf $p(x_1)p(x_2)p(u_3|y_3)$ and function $x_3(u_3, y_3)$.

Remark 1: Theorem 3 includes both the noisy network coding inner bound, which is recovered by letting $U_3 = (\hat{Y}_3, X_3)$ under a pmf $p(\hat{y}_3, x_3|y_3) = p(\hat{y}_3|y_3)p(x_3)$, and the amplify–forward inner bound, which is obtained by setting $U_3 = \emptyset$, and the inclusion can be strict in general.

1) *Gaussian Two-Way Relay Channel:* As an application of Theorem 3, consider the special case of the Gaussian TWRC, where the channel outputs corresponding to the inputs X_1, X_2 , and X_3 are

$$\begin{aligned} Y_1 &= g_{13}X_3 + Z_1, \\ Y_2 &= g_{23}X_3 + Z_2, \\ Y_3 &= g_{31}X_1 + g_{32}X_2 + Z_3, \end{aligned}$$

and the noise components Z_k , $k = 1, 2, 3$, are i.i.d. $\mathcal{N}(0, 1)$. The channel gains g_{kj} from node j to node k are assumed to be real, constant over time, and known throughout the network. We

assume expected power constraint P at each sender. Denote the received SNR $S_{jk} = g_{jk}^2 P$.

Theorem 3 yields the following inner bound on the capacity region.

Corollary 3: A rate pair (R_1, R_2) is achievable for the Gaussian TWRC if

$$\begin{aligned}
R_1 &< \frac{1}{2} \log \frac{\left(\frac{\alpha S_{23}(S_{31}+1)}{S_{31}+S_{32}+1} + \beta S_{23} + 1 \right) (S_{31} + 1 + \sigma^2) - S_{23} \left(\sqrt{\frac{\alpha(S_{31}+1)}{S_{31}+S_{32}+1}} + \sqrt{\beta\sigma^2} \right)^2}{\left(\frac{\alpha S_{23}}{S_{31}+S_{32}+1} + \beta S_{23} + 1 \right) (1 + \sigma^2) - S_{23} \left(\sqrt{\frac{\alpha}{S_{31}+S_{32}+1}} + \sqrt{\beta\sigma^2} \right)^2}, \\
R_1 &< \frac{1}{2} \log \frac{\left(\frac{\alpha S_{23}(S_{31}+1)}{S_{31}+S_{32}+1} + (1 - \alpha)S_{23} + 1 \right) (1 + \sigma^2)}{\left(\frac{\alpha S_{23}}{S_{31}+S_{32}+1} + \beta S_{23} + 1 \right) (1 + \sigma^2) - S_{23} \left(\sqrt{\frac{\alpha}{S_{31}+S_{32}+1}} + \sqrt{\beta\sigma^2} \right)^2} - C(1/\sigma^2), \\
R_2 &< \frac{1}{2} \log \frac{\left(\frac{\alpha S_{13}(S_{32}+1)}{S_{31}+S_{32}+1} + \beta S_{13} + 1 \right) (S_{32} + 1 + \sigma^2) - S_{13} \left(\sqrt{\frac{\alpha(S_{32}+1)}{S_{31}+S_{32}+1}} + \sqrt{\beta\sigma^2} \right)^2}{\left(\frac{\alpha S_{13}}{S_{31}+S_{32}+1} + \beta S_{13} + 1 \right) (1 + \sigma^2) - S_{13} \left(\sqrt{\frac{\alpha}{S_{31}+S_{32}+1}} + \sqrt{\beta\sigma^2} \right)^2}, \\
R_2 &< \frac{1}{2} \log \frac{\left(\frac{\alpha S_{13}(S_{32}+1)}{S_{31}+S_{32}+1} + (1 - \alpha)S_{13} + 1 \right) (1 + \sigma^2)}{\left(\frac{\alpha S_{13}}{S_{31}+S_{32}+1} + \beta S_{13} + 1 \right) (1 + \sigma^2) - S_{13} \left(\sqrt{\frac{\alpha}{S_{31}+S_{32}+1}} + \sqrt{\beta\sigma^2} \right)^2} - C(1/\sigma^2)
\end{aligned}$$

for some $\alpha, \beta \in [0, 1]$ such that $\alpha + \beta \leq 1$ and $\sigma^2 > 0$.

Proof: It suffices to set in Theorem 3 X_1 and X_2 as i.i.d. $\sim N(0, P)$, $U_3 = (V_3, \hat{Y}_3)$, where $\hat{Y}_3 = Y_3 + \hat{Z}_3$, \hat{Z}_3 and V_3 are i.i.d. zero-mean Gaussian independent of (X_1, X_2, Y_3) with variance σ^2 and 1, respectively, and

$$x_3(u_3, y_3) = \sqrt{\frac{\alpha P}{S_{31} + S_{32} + 1}} y_3 + \sqrt{\frac{\beta P}{\sigma^2}} (y_3 - \hat{y}_3) + \sqrt{(1 - \alpha - \beta)P} v_3, \quad (5)$$

for some $\alpha, \beta \in [0, 1]$ such that $\alpha + \beta \leq 1$ and $E(X_3^2) = P$. \blacksquare

Note from (5) that the channel input sequence produced by the relay node is a linear combination of the (analog) sequence Y_3 , the (digital) quantized sequence $\hat{Y}_3 = Y_3 + \hat{Z}_3$, whose resolution is determined by σ_j^2 , and the (digital) sequence V_3 . Hence by varying α and β , we can vary the amount of power allocated to the digital and analog parts in order to optimize the achievable rate region. In particular, by letting $\alpha = \beta = 0$, then the hybrid coding inner bound in Corollary 3 reduces to the noisy network coding inner bound [38] that consists of all rate pairs (R_1, R_2) such that

$$\begin{aligned}
R_1 &< \min \left(C \left(\frac{S_{31}}{1+\sigma^2} \right), C(S_{23}) - C(1/\sigma^2) \right), \\
R_2 &< \min \left(C \left(\frac{S_{32}}{1+\sigma^2} \right), C(S_{13}) - C(1/\sigma^2) \right)
\end{aligned} \quad (6)$$

for some $\sigma^2 > 0$. If instead we let $\alpha = 1$, $\beta = 0$, and $\sigma^2 \rightarrow \infty$, then the hybrid coding inner bound reduces to the amplify–forward inner bound [41] that consists of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< C\left(\frac{S_{23}S_{31}}{1 + S_{23} + S_{31} + S_{32}}\right), \\ R_2 &< C\left(\frac{S_{13}S_{32}}{1 + S_{13} + S_{31} + S_{32}}\right). \end{aligned} \quad (7)$$

Similarly, by letting $\alpha = 0$ and $\beta = 1$, then the hybrid coding inner bound in Corollary 3 reduces to the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< \min\left(C\left(\frac{S_{31}(1 + S_{23})}{1 + \sigma^2 + S_{23}}\right), C\left(\frac{S_{23}\sigma^2}{1 + \sigma^2 + S_{23}}\right) - C(1/\sigma^2)\right), \\ R_2 &< \min\left(C\left(\frac{S_{32}(1 + S_{13})}{1 + \sigma^2 + S_{13}}\right), C\left(\frac{S_{13}\sigma^2}{1 + \sigma^2 + S_{13}}\right) - C(1/\sigma^2)\right) \end{aligned} \quad (8)$$

for some $\sigma^2 > 0$. Finally, by setting $\alpha + \beta = 1$, Corollary 3 includes as a special case a hybrid coding scheme recently proposed in [67].

Fig. 9 compares the cutset bound [42] on the sum-capacity with the inner bound achieved by decode–forward [41], noisy network coding (6), amplify–forward (7), and hybrid coding (8). The plots in the figure assume that nodes 1 and 2 are unit distance apart and node 3 is at distance $r \in [0, 1]$ from node 1 along the line between nodes 1 and 2; the channel gains are of the form $g_{jk} = r_{jk}^{-3/2}$, where r_{jk} is the distance between nodes j and k , hence $g_{13} = g_{31} = r^{-3/2}$, $g_{23} = g_{32} = (1 - r)^{-3/2}$, and the power $P = 10$. Note that the hybrid coding bound in (8) strictly outperforms amplify–forward and noisy network coding for every $r \in (0, 1/2)$.

2) *Hybrid Coding Architecture*: The proposed relay coding scheme can be described as follows. A channel encoder at source node $j = 1, 2$ maps the message M_j into one of 2^{nR_j} sequences $X_j^n(M_j)$ generated i.i.d. according to $\prod_{i=1}^n p_{X_j}(x_{ji})$. The relay node uses the hybrid coding architecture introduced in Section II for the problem of lossy communication over a point-to-point channel. Specifically, at the relay node, the “source” sequence Y_3^n is mapped via hybrid coding to one of 2^{nR_3} independently generated sequences $U_3^n(L_3)$ via joint typicality encoding and then the pair $(Y_3^n, U_3^n(L_3))$ is mapped to X_3^n via the symbol-by-symbol map $x_3(u_3, y_3)$. Decoding at node 1 is performed by searching for the unique message $\hat{M}_2 \in [1 : 2^{nR_2}]$ such that the tuple $(X_1^n(M_1), U_3^n(L_3), X_2^n(\hat{M}_2), Y_4^n)$ is jointly typical for some $L_3 \in [1 : 2^{nR_3}]$. In other words, node 1 nonuniquely decodes the sequence $U_3^n(L_3)$ selected by the relay node.

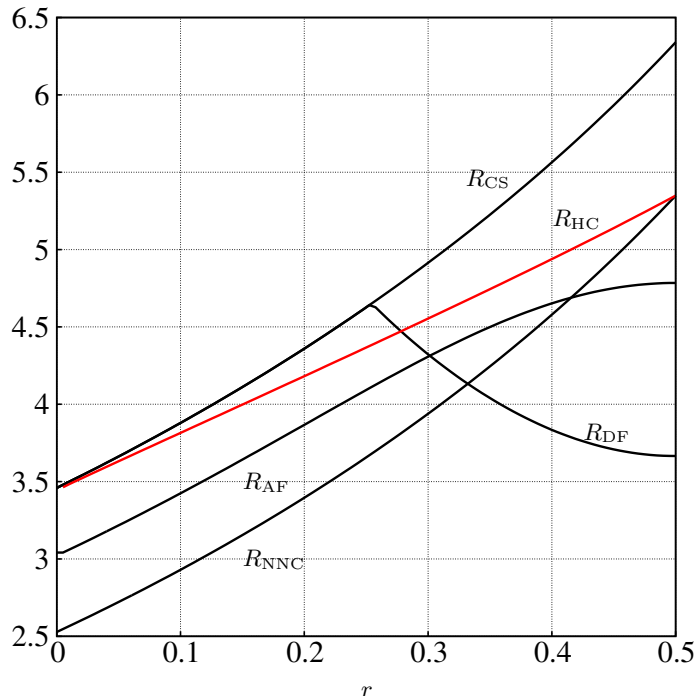


Fig. 9. Comparison of the cutset bound R_{CS} , decode–forward lower bound R_{DF} , amplify–forward lower bound R_{AF} , noisy network coding lower bound R_{NNC} , and hybrid coding lower bound R_{HC} on the sum-capacity for the Gaussian TWRC as a function of the distance r between nodes 1 and 3.

The conditions under which a rate pair (R_1, R_2) is achievable can be obtained by studying the conditions for channel decoding at the destinations and for hybrid encoding at the relay separately. By the covering lemma, the encoding operation at the relay node is successful if

$$R_3 > I(Y_3; U_3).$$

On the other hand, by the packing lemma, suitably modified to account for the dependence between the index and the codebook at the relay node, the channel decoding operation at node 1 is successful if

$$R_2 < I(X_2; Y_1, U_3 | X_1),$$

$$R_2 + R_3 < I(X_2, U_3; X_1, Y_1) + I(X_2; U_3).$$

Similar conditions hold for the case of decoder 2. The lower bound (4) is then established by combining the above inequalities and eliminating the intermediate rate R_3 .

B. Diamond Relay Network

A canonical channel model used to feature the benefits of node cooperation in relay networks is the diamond channel introduced in [39]; see Fig. 10. This two-hop network consists

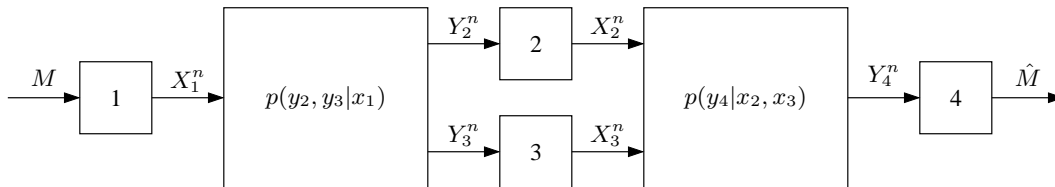


Fig. 10. Diamond channel.

of a source node (node 1) that wishes to send a message $M \in [1 : 2^{nR}]$ to a destination (node 4) with the help of two relay nodes (nodes 2 and 3). The source node is connected through the broadcast channel $p(y_2, y_3 | x_1)$ to the two relay nodes that are in turn connected to the destination node through the multiple-access channel $p(y_4 | x_2, x_3)$. A diamond channel $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3, p(y_2, y_3 | x_1)p(y_4 | x_2, x_3), \mathcal{Y}_2 \times \mathcal{Y}_3 \times \mathcal{Y}_4)$ consists of six alphabet sets and a collection of conditional pmfs on $\mathcal{Y}_2 \times \mathcal{Y}_3 \times \mathcal{Y}_4$. A $(2^{nR}, n)$ code for the diamond channel consists of

- a message set $[1 : 2^{nR}]$,
- an encoder that assigns a codeword $x_1^n(m)$ to each message $m \in [1 : 2^{nR}]$,
- two relay encoders, where relay encoder $j = 2, 3$ assigns a symbol $x_{j,i}(y_j^{i-1})$ to each past received output sequence $y_j^{i-1} \in \mathcal{Y}_j^{i-1}$, and
- a decoder that assigns an estimate \hat{m} or an error message to each received sequence $y_4^n \in \mathcal{Y}_4^n$.

We assume that the message M is uniformly distributed over $[1 : 2^{nR}]$. The average probability of error is defined as $P_e^{(n)} = \mathbb{P}\{\hat{M} \neq M\}$. A rate R is said to be achievable for the diamond channel if there exists a sequence of $(2^{nR}, n)$ codes such that $\lim_{n \rightarrow \infty} P_e^{(n)} = 0$. The capacity C of the diamond channel is the supremum of the achievable rates R .

The capacity of the diamond channel is not known in general. Schein and Gallager [39] characterized inner bounds on the capacity region based on decode–forward, compress–forward, and amplify–forward.

Hybrid coding yields the following lower bound on the capacity, the proof of which is given in Appendix D.

Theorem 4: The capacity of the diamond channel $p(y_2, y_3|x_1)p(y_4|x_2, x_3)$ is lower bounded as

$$C \geq \max \min \{ I(X_1; U_2, U_3, Y_4), I(X_1, U_2; U_3, Y_4) - I(U_2; Y_2|X_1), \\ I(X_1, U_3; U_2, Y_4) - I(U_3; Y_3|X_1), I(X_1, U_2, U_3; Y_4) - I(U_2, U_3; Y_2, Y_3|X_1) \}, \quad (9)$$

where the maximum is over all conditional pmfs $p(x_1)p(u_2|y_2)p(u_3|y_3)$ and functions $x_2(u_2, y_2)$, $x_3(u_3, y_3)$.

Remark 2: Theorem 4 includes both the noisy network coding inner bound, which is recovered by setting $U_j = (X_j, \hat{Y}_j)$ with $p(x_j)p(\hat{y}_j|y_j)$, $j = 2, 3$, and the amplify-forward inner bound, which is obtained by setting $U_j = \emptyset$ for $j = 2, 3$, and the inclusion can be strict in general, as demonstrated below.

1) *Deterministic Diamond Channel:* Consider the special case where the multiple access channel $p(y_2, y_3|x_1)$ and the broadcast channel $p(y_4|x_2, x_3)$ are deterministic, i.e., the channel outputs are functions of the corresponding inputs. In this case, Theorem 4 simplifies to the following.

Corollary 4: The capacity of the deterministic diamond channel is lower bounded as

$$C \geq \max_{p(x_1)p(x_2|y_2)p(x_3|y_3)} R(Y_2, Y_3, Y_4|X_2, X_3), \quad (10)$$

where

$$R(Y_2, Y_3, Y_4|X_2, X_3) = \min \{ H(Y_2, Y_3), H(Y_2) + H(Y_4|X_2, Y_2), \\ H(Y_3) + H(Y_4|X_3, Y_3), H(Y_4) \}.$$

Proof: Set in Theorem 4 $U_2 = (Y_2, X_2)$, $U_3 = (Y_3, X_3)$, $x_2(u_2, y_2) = x_2$, and $x_3(u_3, y_3) = x_3$ under a pmf $p(x_2|y_2)p(x_3|y_3)$. ■

We can compare the result in Corollary 4 with the existing inner and outer bounds for this channel model. An outer bound on the capacity region is given by the cutset bound [68], which in this case simplifies to

$$C \leq \max_{p(x_1)p(x_2, x_3)} R(Y_2, Y_3, Y_4|X_2, X_3) \quad (11)$$

On the other hand, specializing the scheme in [69] for deterministic relay networks, we obtain the lower bound

$$C \geq \max_{p(x_1)p(x_2)p(x_3)} R(Y_2, Y_3, Y_4 | X_2, X_3). \quad (12)$$

Note that (10), (11), and (12) differ only in the set of allowed maximizing input pmfs. In particular, (10) improves upon the inner bound (12) by allowing X_2 and X_3 to depend on Y_2 and Y_3 and thereby increasing the set of distributions $p(x_2, x_3)$. The following example demonstrates that the inclusion can be strict.

Example 1: Suppose that $p(y_2, y_3 | x_1)$ is the Blackwell broadcast channel (i.e., $X_1 \in \{0, 1, 2\}$ and $p_{Y_2, Y_3 | X_1}(0, 0 | 0) = p_{Y_2, Y_3 | X_1}(0, 1 | 1) = p_{Y_2, Y_3 | X_1}(1 | 2) = 1$) and $p(y_4 | x_2, x_3)$ is the binary erasure multiple access channel (i.e., $X_2, X_3 \in \{0, 1\}$ and $Y_4 = X_2 + X_3 \in \{0, 1, 2\}$). It can be easily seen that the general lower bound reduces to $C \geq 1.5$, while the capacity is $C = \log 3$, which coincides with the hybrid coding lower bound (with $X_2 = Y_2$ and $X_3 = Y_3$). Thus, hybrid coding strictly outperforms the coding scheme by Avestimehr, Diggavi, and Tse [69] and noisy network coding [38].

2) *Hybrid Coding Architecture:* The proof of achievability of Theorem 4 is based on a hybrid coding architecture similar to the one used in the proof of Theorem 2 and can be described as follows. At the source node, the message M is mapped to one of 2^{nR_1} sequences $X_1^n(M)$ i.i.d. $\sim p(x_1)$ as in point-to-point communication. At the relay nodes, the “source” the sequence Y_j^n , $j = 2, 3$, is separately mapped into one of 2^{nR_j} independently generated sequences $U_j^n(M_j)$. Then, the pair $(Y_j^n, U_j^n(M_j))$ is mapped by node j to X_j^n via a symbol-by-symbol map. By the covering lemma, the source encoding operation at the relays is successful if

$$R_2 > I(U_2; Y_2)$$

$$R_3 > I(U_3; Y_3).$$

At the destination node, decoding is performed by joint typicality and indirect decoding of the sequences (U_2^n, U_3^n) , that is, by searching for the unique message $\hat{M} \in [1 : 2^{nR}]$ such that the tuple $(X_1^n(\hat{M}), U_2^n(M_2), U_3^n(M_3), Y_4^n)$ is typical for some $M_2 \in [1 : 2^{nR_2}]$ and $M_3 \in [1 : 2^{nR_3}]$. By the packing lemma, combined with the technique introduced in Section II, the channel decoding operation at the destination node is successful if

$$R < I(X_1; U_2, U_3, Y_4)$$

$$R + R_2 < I(X_1, U_2; U_3, Y_4) + I(X_1; U_2)$$

$$R + R_3 < I(X_1, U_3; U_2, Y_4) + I(X_1; U_3)$$

$$R + R_2 + R_3 < I(X_1, U_2, U_3; Y_4) + I(X_1; U_2) + I(X_1, U_2; U_3).$$

Hence, the lower bound (9) is obtained by combining the conditions for source coding at the relay nodes with those for channel decoding at the destination and by eliminating the auxiliary rates (R_1, R_2) from the resulting system of inequalities.

V. CONCLUDING REMARKS

In this paper we presented a new approach to studying lossy communication of correlated sources over networks based on hybrid analog/digital coding. We first revisited the problem of lossy communication over a point-to-point channel, for which we proposed a hybrid scheme that generalizes both the digital, separate source and channel coding scheme and the analog, uncoded transmission scheme. Similar to Shannon's source-channel separation architecture, the proposed hybrid scheme employs a modular system architecture, whereby the source encoding and channel decoding operations are decoupled. However, unlike the separation architecture, a single codebook is used for both source coding and channel coding, which allows the resulting coding scheme to achieve the performance of the best known joint source-channel coding schemes.

Next, we discussed how the proposed hybrid coding architecture can be generalized to construct joint source-channel coding schemes for lossy communication over multiuser channels. To illustrate the main ideas, we focused on the specific problem of lossy communications over multiple access channels, for which we presented a joint source-channel coding scheme that unifies and generalizes several existing results in the literature. As in the case of point-to-point communication, the proposed scheme is conceptually similar to separate source and channel coding and is obtained by concatenating the source coding scheme in [49], [50] for distributed lossy source coding with a channel code for multiple access communication, except that the same codeword is used for source coding as well as for channel coding. The same design principle can be readily adapted to other joint source-channel coding problems for which separate source coding and channel coding have matching index structures, such as

- communication of a 2-DMS with common part over a DM-MAC (Berger–Tung coding with common part [70], [71] matched to the multiple access channel coding with common message [54]),
- communication of a 2-DMS over a DM broadcast channel (lossy Gray–Wyner system [35] matched to Marton’s coding for a broadcast channel [60]),
- communication of a 2-DMS over a DM interference channel (extension of Berger–Tung coding for a 2-by-2 source network matched to Han–Kobayashi coding for an interference channel [72]).

In all these cases, hybrid coding performs as well as (and sometimes better than) the existing coding schemes [45].

Finally, we explored applications of hybrid coding in the context of relay networks. We introduced a general coding technique for DM relay networks based on hybrid coding, whereby each relay uses the hybrid coding interface to transmit a symbol-by-symbol function of the received sequence and its quantized version (analog-to-analog/digital interface). We demonstrated via two specific examples, the two-relay diamond channel and the two-way relay channel, that the proposed hybrid coding can strictly outperform both amplify–forward (analog-to-analog interface) and compress–forward/noisy network coding (analog-to-digital interfaces). For simplicity, we assumed that the relay nodes do not attempt to decode the message transmitted by the source, but the presented results can be further improved by combining hybrid coding with other coding techniques such as decode–forward and structured coding [63]. In this case, hybrid coding provides a general analog/digital-to-analog/digital interface for relay communication. While we have focused on two specific examples, similar ideas can be applied to general layered network model, provided that the proposed hybrid coding scheme is repeated at each layer in the network [45]. In principle, hybrid coding can also be applied to the relay channel and other nonlayered relay networks. However, in this case hybrid coding (or even amplify–forward) would not yield inner bounds to the capacity region in a single-letter form, due to the dependency between the channel input at each relay node and the previously received analog channel outputs.

APPENDIX A

PROOF OF LEMMA 1

Given \tilde{u}^n and s^n , let $\mathcal{A} = \{U^n(1) = \tilde{u}^n, S^n = s^n\}$ in short. Let $\mathcal{C}' = \{U^n(m) : m \in [3 : 2^{nR}]\}$. Then, by the law of total probability and the Bayes rule, for every u^n ,

$$\begin{aligned}
& \mathbb{P}\{U^n(2) = u^n \mid M = 1, \mathcal{A}\} \\
&= \sum_{\mathcal{C}'} \mathbb{P}\{\mathcal{C}' = c', U^n(2) = u^n \mid M = 1, \mathcal{A}\} \\
&= \sum_{\mathcal{C}'} \mathbb{P}\{\mathcal{C}' = c' \mid M = 1, \mathcal{A}\} \mathbb{P}\{U^n(2) = u^n \mid \mathcal{A}, \mathcal{C}' = c'\} \frac{\mathbb{P}\{M = 1 \mid \mathcal{A}, U^n(2) = u^n, \mathcal{C}' = c'\}}{\mathbb{P}\{M = 1 \mid \mathcal{A}, \mathcal{C}' = c'\}} \\
&= \sum_{\mathcal{C}'} \mathbb{P}\{\mathcal{C}' = c' \mid M = 1, \mathcal{A}\} \left(\prod_{i=1}^n p_{U^n}(u_i) \right) \frac{\mathbb{P}\{M = 1 \mid \mathcal{A}, U^n(2) = u^n, \mathcal{C}' = c'\}}{\mathbb{P}\{M = 1 \mid \mathcal{A}, \mathcal{C}' = c'\}}. \tag{13}
\end{aligned}$$

For each $(\tilde{u}^n, s^n, u^n, c')$ such that $\mathbb{P}\{\mathcal{C}' = c' \mid M = 1, \mathcal{A}\} > 0$, let $n(\tilde{u}^n, s^n, u^n, c') = n(s^n, c') = |\{u'^n \in c' : (u'^n, s^n) \in \mathcal{T}_c^{(n)}\}|$ denote the number of unique sequences in c' that are jointly typical with s^n and

$$i(\tilde{u}^n, s^n, u^n, c') = i(\tilde{u}^n, s^n, c') = \begin{cases} 1, & (\tilde{u}^n, s^n) \notin \mathcal{T}_c^{(n)} \text{ and } n(s^n, c') = 0, \\ 0, & \text{otherwise,} \end{cases}$$

be the indicator function for the case that neither \tilde{u}^n nor any codeword in c' is jointly typical with s^n . Then, by the way the random index M is generated, it can be easily verified that

$$\mathbb{P}\{M = 1 \mid \mathcal{A}, U^n(2) = u^n, \mathcal{C}' = c'\} \leq \frac{1}{2^{nR}} i(\tilde{u}^n, s^n, c') + \frac{1}{n(s^n, c') + 1} (1 - i(\tilde{u}^n, s^n, c')).$$

Similarly, since $U^n(2) \sim \prod_{i=1}^n p_U(u_i)$, independent of S^n and $U^n(m)$, $m \neq 2$,

$$\begin{aligned}
& \mathbb{P}\{M = 1 \mid \mathcal{A}, \mathcal{C}' = c\} \\
& \geq \mathbb{P}\{M = 1 \mid \mathcal{A}, \mathcal{C}' = c', 2 \notin \mathcal{I}\} \cdot \mathbb{P}\{2 \notin \mathcal{I} \mid \mathcal{A}, \mathcal{C}' = c'\} \\
& \geq \mathbb{P}\{M = 1 \mid \mathcal{A}, \mathcal{C}' = c', 2 \notin \mathcal{I}\} \left(1 - 2^{-n(I(U;S) - \delta(\epsilon'))}\right) \\
& = \left(\frac{1}{2^{nR}} i(\tilde{u}^n, s^n, c') + \frac{1}{n(s^n, c') + 1} (1 - i(\tilde{u}^n, s^n, c')) \right) \left(1 - 2^{-n(I(U;S) - \delta(\epsilon'))}\right).
\end{aligned}$$

It follows that

$$\frac{\mathbb{P}\{M = 1 \mid U^n(2) = u^n, E, \mathcal{C}' = c'\}}{\mathbb{P}\{M = 1 \mid E, \mathcal{C}' = c'\}} \leq \frac{1}{1 - 2^{-n(I(U;S) - \delta(\epsilon'))}} \leq 1 + \epsilon \tag{14}$$

for n sufficiently large. By combining (13) and (14), the claim follows.

APPENDIX B
PROOF OF THEOREM 2

For simplicity, we consider the case $Q = \emptyset$. Achievability for an arbitrary Q can be proved using coded time sharing [42, Section 4.5.3].

Codebook generation: Let $\epsilon > \epsilon' > 0$. Fix a pmf $p(u_1|s_1)p(u_2|s_2)$, channel encoding functions $x_1(u_1, s_1)$ and $x_2(u_2, s_2)$, and source decoding functions $\hat{s}_1(u_1, u_2, y)$ and $\hat{s}_2(u_1, u_2, y)$ such that $E(d_j(S_j, \hat{S}_j)) \leq D_j/(1 + \epsilon)$, $j = 1, 2$. For each $j = 1, 2$, randomly and independently generate 2^{nR_j} sequences $u_j^n(m_j)$, $m_j \in [1 : 2^{nR_j}]$, each according to $\prod_{i=1}^n p_{U_j}(u_{ji})$. The codebook $\mathcal{C} = \{(u_1^n(m_1), u_2^n(m_2)) : m_1 \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]\}$ is revealed to both the encoders and the decoder.

Encoding: Upon observing a sequence s_j^n , encoder $j = 1, 2$ finds an index $m_j \in [1 : 2^{nR_j}]$ such that $(s_j^n, u_j^n(m_j)) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there is more than one such index, it chooses one of them at random. If there is no such index, it chooses an arbitrary index at random from $[1 : 2^{nR_j}]$. Encoder j then transmits $x_{ji} = x_j(u_{ji}(m_j), s_{ji})$ for $i \in [1 : n]$.

Decoding: Upon receiving y^n , the decoder finds the unique index pair (\hat{m}_1, \hat{m}_2) such that $(u_1^n(\hat{m}_1), u_2^n(\hat{m}_2), y^n) \in \mathcal{T}_{\epsilon}^{(n)}$ and sets the estimates as $\hat{s}_{ji} = \hat{s}_j(u_{1i}(m_1), u_{2i}(m_2), y_i)$, $i \in [1 : n]$, for $j = 1, 2$.

Analysis of the expected distortion: We bound the distortion averaged over (S_1^n, S_2^n) , the random choice of the codebook \mathcal{C} , and the random index assignments at the encoders. Let M_1 and M_2 be the random variables denoting the chosen indexes at encoder 1 and at encoder 2, respectively. Define the “error” event

$$\mathcal{E} = \{(S_1^n, S_2^n, U_1^n(\hat{M}_1), U_2^n(\hat{M}_2), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}$$

and partition it into

$$\begin{aligned} \mathcal{E}_j &= \{(S_j^n, U_j^n(m_j)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } m_j\}, \quad j = 1, 2, \\ \mathcal{E}_3 &= \{(S_1^n, S_2^n, U_1^n(M_1), U_2^n(M_2), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}, \\ \mathcal{E}_4 &= \{(U_1^n(m_1), U_2^n(m_2), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_1 \neq M_1, m_2 \neq M_2\}, \\ \mathcal{E}_5 &= \{(U_1^n(m_1), U_2^n(M_2), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_1 \neq M_1\}, \\ \mathcal{E}_6 &= \{(U_1^n(M_1), U_2^n(m_2), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_2 \neq M_2\}. \end{aligned}$$

Then by the union of events bound,

$$P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2) + P(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c) + P(\mathcal{E}_4) + P(\mathcal{E}_5) + P(\mathcal{E}_6).$$

As in the case of point-to-point communication studied in Section II, the desired distortion pair is achieved if $P(\mathcal{E})$ tends to zero as $n \rightarrow \infty$. By the covering lemma, $P(\mathcal{E}_1)$ and $P(\mathcal{E}_2)$ tend to zero as $n \rightarrow \infty$, if

$$R_1 > I(U_1; S_1) + \delta(\epsilon'), \quad (15)$$

$$R_2 > I(U_2; S_2) + \delta(\epsilon'). \quad (16)$$

By the Markov lemma [42, Section 12.1.1], the third term tends to zero as $n \rightarrow \infty$.

By the symmetry of random codebook generation and encoding, we analyze the remaining probabilities conditioned on the event $\mathcal{M} = \{M_1 = 1, M_2 = 1\}$. First, we bound $P(\mathcal{E}_4)$. By the union of events bound,

$$P(\mathcal{E}_4 | \mathcal{M}) \leq \sum_{m_1=2}^{2^{nR_1}} \sum_{m_2=2}^{2^{nR_2}} \sum_{(u_1^n, u_2^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} P\{U_1^n(m_1) = u_1^n, U_2^n(m_2) = u_2^n, Y^n = y^n | \mathcal{M}\}. \quad (17)$$

Let $\tilde{U}^n = (U_1^n(1), U_2^n(1), S_1^n, S_2^n)$ and $\tilde{u}^n = (\tilde{u}_1^n, \tilde{u}_2^n, s_1^n, s_2^n)$ in short. Then, by the law of total probability, for $m_1 \neq 1$ and $m_2 \neq 1$,

$$\begin{aligned} & P\{U_1^n(m_1) = u_1^n, U_2^n(m_2) = u_2^n, Y^n = y^n | \mathcal{M}\} \\ &= \sum_{\tilde{u}^n} P\{U_1^n(m_1) = u_1^n, U_2^n(m_2) = u_2^n, Y^n = y^n, \tilde{U}^n = \tilde{u}^n | \mathcal{M}\} \\ &\stackrel{(a)}{=} \sum_{\tilde{u}^n} P\{U_1^n(m_1) = u_1^n | \mathcal{M}, \tilde{U}^n = \tilde{u}^n\} P\{U_2^n(m_2) = u_2^n | \mathcal{M}, U_1^n(m_1) = u_1^n, \tilde{U}^n = \tilde{u}^n\} \\ &\quad \times P\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, Y^n = y^n\} P\{Y^n = y^n | \mathcal{M}\} \\ &\stackrel{(b)}{=} \sum_{(\tilde{u}_1^n, \tilde{u}_2^n, s_1^n, s_2^n)} P\{U_1^n(m_1) = u_1^n | M_1 = 1, U_1^n(1) = \tilde{u}_1^n, S_1^n = s_1^n\} \\ &\quad \times P\{U_2^n(m_2) = u_2^n | M_2 = 1, U_2^n(1) = \tilde{u}_2^n, S_2^n = s_2^n\} \\ &\quad \times P\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, Y^n = y^n\} P\{Y^n = y^n | \mathcal{M}\} \\ &\stackrel{(c)}{\leq} (1 + \epsilon) \sum_{\tilde{u}^n} \left(\prod_{i=1}^n p_{U_1}(u_{1i}) p_{U_2}(u_{2i}) \right) P\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, Y^n = y^n\} P\{Y^n = y^n | \mathcal{M}\} \\ &= (1 + \epsilon) \left(\prod_{i=1}^n p_{U_1}(u_{1i}) p_{U_2}(u_{2i}) \right) P\{Y^n = y^n | \mathcal{M}\} \end{aligned} \quad (18)$$

for n sufficiently large. Here, (a) follows from the fact that given \mathcal{M}

$$(U_1^n(m_1), U_2^n(m_2)) \rightarrow (S_1^n, S_2^n, M_1, M_2, U_1^n(M_1), U_2^n(M_2)) \rightarrow Y^n \quad (19)$$

form a Markov chain for all $m_1 \neq 1$ and $m_2 \neq 1$, while (b) follows by the independence of the sequences and the encoding procedure. For step (c), we apply Lemma 1 twice. Combining (17) and (18), it follows that for n sufficiently large

$$\begin{aligned} \mathbb{P}(\mathcal{E}_4 | \mathcal{M}) &\leq (1 + \epsilon) \sum_{m_1=2}^{2^{nR_1}} \sum_{m_2=2}^{2^{nR_2}} \sum_{(u_1^n, u_2^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} p_{U_1^n}(u_1^n) p_{U_2^n}(u_2^n) \mathbb{P}\{Y^n = y^n | \mathcal{M}\} \\ &\leq (1 + \epsilon) 2^{n(R_1+R_2)} \sum_{y^n \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{Y^n = y^n | \mathcal{M}\} 2^{-n(I(U_1, U_2; Y) + I(U_1; U_2) - \delta(\epsilon))} \\ &\leq (1 + \epsilon) 2^{n(R_1+R_2 - I(U_1, U_2; Y) - I(U_1; U_2) + \delta(\epsilon))}. \end{aligned}$$

Hence, $\mathbb{P}(\mathcal{E}_4)$ tends to zero as $n \rightarrow \infty$ if

$$R_1 + R_2 < I(U_1, U_2; Y) + I(U_1; U_2) - \delta(\epsilon). \quad (20)$$

Following similar steps, $\mathbb{P}(\mathcal{E}_5)$ is upper bounded by

$$\begin{aligned} &\mathbb{P}\{(U_1^n(m_1), U_2^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1 | \mathcal{M}\} \\ &\leq \sum_{m_1=2}^{2^{nR_1}} \mathbb{P}\{(U_1^n(m_1), U_2^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)} | \mathcal{M}\} \\ &= \sum_{m_1=2}^{2^{nR_1}} \sum_{(u_1^n, u_2^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{U_1^n(m_1) = u_1^n, U_2^n(1) = u_2^n, Y^n = y^n | \mathcal{M}\} \\ &\stackrel{(a)}{\leq} (1 + \epsilon) \sum_{m_1=2}^{2^{nR_1}} \sum_{(u_1^n, u_2^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} p_{U_1^n}(u_1^n) \mathbb{P}\{Y^n = y^n, U_2^n(1) = u_2^n | \mathcal{M}\} \\ &\leq (1 + \epsilon) 2^{nR_1} \sum_{(u_2^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{U_2^n(1) = u_2^n, Y^n = y^n | \mathcal{M}\} 2^{-n(I(U_1; Y, U_2) - \delta(\epsilon))} \\ &\leq (1 + \epsilon) 2^{n(R_1 - I(U_1; Y, U_2) + \delta(\epsilon))} \end{aligned}$$

for n sufficiently large, which implies that $\mathbb{P}(\mathcal{E}_5)$ tends to zero as $n \rightarrow \infty$ if

$$R_1 < I(U_1; Y, U_2) - \delta(\epsilon). \quad (21)$$

In the above chain of inequalities step (a) is justified as follows. Let $\tilde{U}^n = (U_1^n(1), S_1^n, S_2^n)$ and $\tilde{u}^n = (\tilde{u}_1^n, s_1^n, s_2^n)$ in short. Then, by the law of total probability, for $m_1 \neq 1$ and $m_2 \neq 1$,

$$\begin{aligned}
& \mathbb{P}\{U_1^n(m_1) = u_1^n, U_2^n(1) = u_2^n, Y^n = y^n | \mathcal{M}\} \\
&= \sum_{\tilde{u}^n} \mathbb{P}\{U_1^n(m_1) = u_1^n, U_2^n(1) = u_2^n, Y^n = y^n, \tilde{U}^n = \tilde{u}^n | \mathcal{M}\} \\
&\stackrel{(b)}{=} \sum_{\tilde{u}^n} \mathbb{P}\{U_1^n(m_1) = u_1^n | \mathcal{M}, U_2^n(1) = u_2^n, \tilde{U}^n = \tilde{u}^n\} \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, U_2^n(1) = u_2^n, Y^n = y^n\} \\
&\quad \times \mathbb{P}\{U_2^n(1) = u_2^n, Y^n = y^n | \mathcal{M}\} \\
&\stackrel{(c)}{=} \sum_{\tilde{u}^n} \mathbb{P}\{U_1^n(m_1) = u_1^n | M_1 = 1, U_1^n(1) = \tilde{u}_1^n, S_1^n = s_1^n\} \\
&\quad \times \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, U_2^n(1) = u_2^n, Y^n = y^n\} \mathbb{P}\{U_2^n(1) = u_2^n, Y^n = y^n | \mathcal{M}\} \\
&\stackrel{(d)}{\leq} (1 + \epsilon) \sum_{\tilde{u}^n} \left(\prod_{i=1}^n p_{U_1}(u_{1i}) \right) \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, U_2^n(1) = u_2^n, Y^n = y^n\} \\
&\quad \times \mathbb{P}\{U_2^n(1) = u_2^n, Y^n = y^n | \mathcal{M}\} \\
&= (1 + \epsilon) \left(\prod_{i=1}^n p_{U_1}(u_{1i}) \right) \mathbb{P}\{U_2^n(1) = u_2^n, Y^n = y^n | \mathcal{M}\}
\end{aligned}$$

for n sufficiently large. Here, (b) follows from the fact that (19) form a Markov chain for all $m_1 \neq 1$ and $m_2 \neq 1$, (c) follows by the independence of the sequences and the encoding procedure, while (d) follows from Lemma 1.

Finally, $\mathbb{P}(\mathcal{E}_6)$ can be bounded in a similar manner, provided that the subscripts 1 and 2 are interchanged in the upper bound for $\mathbb{P}(\mathcal{E}_5)$. It follows that $\mathbb{P}(\mathcal{E}_6)$ tends to zero as $n \rightarrow \infty$ if

$$R_2 < I(U_2; Y, U_1) - \delta(\epsilon). \quad (22)$$

Therefore, if (15), (16), (20), (21), and (22), the probability of “error” tends to zero as $n \rightarrow \infty$ and the average distortions over the random codebooks is bounded as desired. Thus, there exists at least one sequence of codes achieving the desired distortions. By letting $\epsilon \rightarrow 0$ and eliminating the intermediate rate pair (R_1, R_2) , the sufficient condition in Theorem 2 (with $Q = \emptyset$) for lossy communication over a DM-MAC via hybrid coding is established.

APPENDIX C

PROOF OF THEOREM 3

We use b transmission blocks, each consisting of n transmissions, as in the proof of the multihop lower bound for the relay channel [42, Section 16.4.1]. A sequence of $(b - 1)$ message

pairs $(m_{1j}, m_{2j}) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$, $j \in [1 : b - 1]$, each selected independently and uniformly over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$ is sent over b blocks. Note that the average rate pair over the b blocks is $((b - 1)/b)(R_1, R_2)$, which can be made arbitrarily close to (R_1, R_2) by letting $b \rightarrow \infty$.

Codebook generation: Let $\epsilon > \epsilon' > 0$. Fix $p(x_1)p(x_2)p(u_3|y_3)$ and an encoding function $x_3(u_3, y_3)$. We randomly and independently generate a codebook for each block. For $j \in [1 : b]$, randomly and independently generate 2^{nR_3} sequences $u_3^n(l_{3j})$, $l_{3j} \in [1 : 2^{nR_3}]$, each according to $\prod_{i=1}^n p_{U_3}(u_{3i})$. For each $k = 1, 2$, randomly and independently generate 2^{nR_k} sequences $x_k^n(m_{kj})$, $m_{kj} \in [1 : 2^{nR_k}]$, each according to $\prod_{i=1}^n p_{X_k}(x_{ki})$. This defines the codebook

$$\mathcal{C}_j = \{(x_1^n(m_{1j}), x_2^n(m_{2j}), u_3^n(l_{3j})) : m_1 \in [1 : 2^{nR_1}], m_2 \in [1 : 2^{nR_2}], l_3 \in [1 : 2^{nR_3}]\}$$

for $j \in [1 : b]$.

Encoding: Let $m_{kj} \in [1 : 2^{nR_k}]$ be the independent message to be sent in block $j \in [1 : b - 1]$ by node $k = 1, 2$. Then, node k transmits $x_k^n(m_{kj})$ from codebook \mathcal{C}_j .

Relay encoding: Upon receiving $y_3^n(j)$ in block $j \in [1 : b - 1]$, relay node 3 finds an index l_{3j} such that $(u_3^n(l_{3j}), y_3^n(j)) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there is more than one index, it chooses one of them at random. If there is no such index, it chooses an arbitrary index at random from $[1 : 2^{nR_3}]$. In block $j + 1$, relay 3 then transmits $x_{3i} = x_{3i}(u_{3i}(l_{3j}), y_{3i}(j))$ for $i \in [1 : n]$.

Decoding: Upon receiving $y_1^n(j)$, $j \in [2 : b]$, decoder 1 finds the unique message $\hat{m}_{2,(j-1)}$ such that

$$(x_1^n(m_{j-1}), u_3^n(l_{3,(j-1)}), x_2^n(\hat{m}_{2,(j-1)}), y_1^n(j)) \in \mathcal{T}_{\epsilon}^{(n)},$$

for some $l_{3,(j-1)} \in [1 : 2^{nR_3}]$. Decoding at node 2 is performed in a similar manner.

Analysis of the probability of error: We analyze the probability of decoding error at node 1 in block $j = 2, \dots, b$, averaged over the random codebooks and index assignment in the encoding procedure at the relay. Let $L_{3,(j-1)}$ be the random variable denoting the index chosen in block $j - 1$ at relay 3. Decoder 1 makes an error only if one or more of the following events occur:

$$\mathcal{E}_1 = \{(Y_3^n(j - 1), U_3^n(l)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l\},$$

$$\mathcal{E}_2 = \{(X_1^n(M_{1,(j-1)}), U_3^n(L_{3,(j-1)}), X_2^n(M_{2,(j-1)}), Y_1^n(j)) \notin \mathcal{T}_{\epsilon}^{(n)}\},$$

$$\mathcal{E}_3 = \{(X_1^n(M_{1,(j-1)}), U_3^n(L_{3,(j-1)}), X_2^n(m), Y_1^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m \neq M_{2,(j-1)}\},$$

$$\mathcal{E}_4 = \{(X_1^n(M_{1,(j-1)}), U_3^n(l), X_2^n(m), Y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \neq M_{2,(j-1)}, l \neq L_{3,(j-1)}\},$$

Then, by the union of events bound the probability of decoding error is upper bounded as

$$\mathbb{P}(\hat{M}_{2,(j-1)} \neq M_{2,(j-1)}) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c) + \mathbb{P}(\mathcal{E}_3) + \mathbb{P}(\mathcal{E}_4).$$

By the covering lemma, $\mathbb{P}(\mathcal{E}_1)$ tends to zero as $n \rightarrow \infty$, if

$$R_3 > I(U_3; Y_3) + \delta(\epsilon').$$

By the Markov lemma, $\mathbb{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c)$ tends to zero as $n \rightarrow \infty$. By the symmetry of the random codebooks generation and random index assignment at the relays, it suffices to consider the conditional probabilities of the remaining error events conditioned on the event that

$$\mathcal{M} = \{M_{1,(j-1)} = 1, M_{2,(j-1)} = 1, L_{3,(j-1)} = 1\}. \quad (23)$$

Then, by the packing lemma, $\mathbb{P}(\mathcal{E}_3)$ tends to zero as $n \rightarrow \infty$ if

$$R_2 < I(X_2; Y_1, U_3 | X_1) - \delta(\epsilon).$$

Next, for n large enough $\mathbb{P}(\mathcal{E}_4)$ is upper bounded by

$$\begin{aligned} & \mathbb{P}\{(X_1^n(1), U_3^n(l), X_2^n(m), Y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } l \neq 1, m \neq 1 | \mathcal{M}\} \\ & \leq \sum_{m=2}^{2^{nR_2}} \sum_{l=2}^{2^{nR_3}} \sum_{(x_1^n, u_3^n, x_2^n, y_1^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{X_1^n(1) = x_1^n, U_3^n(l) = u_3^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n | \mathcal{M}\} \\ & \stackrel{(a)}{\leq} (1 + \epsilon) \sum_{m=2}^{2^{nR_2}} \sum_{l=2}^{2^{nR_3}} \sum_{(x_1^n, u_3^n, x_2^n, y_1^n) \in \mathcal{T}_\epsilon^{(n)}} p_{U_3^n}(u_3^n) \cdot \mathbb{P}\{X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n | \mathcal{M}\} \\ & = (1 + \epsilon) \sum_{m=2}^{2^{nR_2}} \sum_{l=2}^{2^{nR_2}} \sum_{(x_1^n, u_3^n, x_2^n, y_1^n) \in \mathcal{T}_\epsilon^{(n)}} p_{U_3^n}(u_3^n) \cdot p_{X_2^n}(x_2^n) \cdot \mathbb{P}\{X_1^n(1) = x_1^n, Y_1^n(j) = y_1^n | \mathcal{M}\} \\ & \leq (1 + \epsilon) 2^{n(R_2+R_3)} \sum_{(x_1^n, y_1^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{X_1^n(1) = x_1^n, Y_1^n(j) = y_1^n | \mathcal{M}\} 2^{-n(I(X_2, U_3; Y_1, X_1) + I(X_2; U_3) - \delta(\epsilon))} \\ & \leq (1 + \epsilon) 2^{n(R_2+R_3 - I(X_2, U_3; Y_1, X_1) - I(X_2; U_3) + \delta(\epsilon))}. \end{aligned}$$

Here, step (a) is justified as follows. Let $\tilde{U}^n = (U_3^n(1), Y_3^n(j-1))$ and $\tilde{u}^n = (\tilde{u}_3^n, \tilde{y}_3^n)$ in short.

Then, by the law of total probability, for $l \neq 1$ and $m \neq 1$,

$$\mathbb{P}\{X_1^n(1) = x_1^n, U_3^n(l) = u_3^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n | \mathcal{M}\}$$

$$\begin{aligned}
&= \sum_{\tilde{u}^n} \mathbb{P}\{X_1^n(1) = x_1^n, U_3^n(l) = u_3^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n, \tilde{U}^n = \tilde{u}^n | \mathcal{M}\} \\
&\stackrel{(b)}{=} \sum_{\tilde{u}^n} \mathbb{P}\{U_3^n(l) = u_3^n | \mathcal{M}, X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, \tilde{U}^n = \tilde{u}^n\} \\
&\quad \times \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n\} \\
&\quad \times \mathbb{P}\{X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n | \mathcal{M}\} \\
&\stackrel{(c)}{=} \sum_{\tilde{u}^n} \mathbb{P}\{U_3^n(l) = u_3^n | L_{3,(j-1)} = 1, U_3^n(1) = \tilde{u}_3^n, Y_3^n(j-1) = \tilde{y}_3^n\} \\
&\quad \times \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n\} \\
&\quad \times \mathbb{P}\{X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n | \mathcal{M}\} \\
&\stackrel{(d)}{\leq} (1 + \epsilon) \sum_{\tilde{u}^n} \left(\prod_{i=1}^n p_{U_3}(u_{3i}) \right) \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n\} \\
&\quad \times \mathbb{P}\{X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n | \mathcal{M}\} \\
&= (1 + \epsilon) \left(\prod_{i=1}^n p_{U_1}(u_{1i}) \right) \mathbb{P}\{X_1^n(1) = x_1^n, X_2^n(m) = x_2^n, Y_1^n(j) = y_1^n | \mathcal{M}\}
\end{aligned}$$

for n sufficiently large, where (b) follows from the fact that given \mathcal{M}

$$U_3^n(l) \rightarrow (X_1^n(1), X_2^n(m), U_3^n(L_{3,(j-1)}), Y_3^n(j-1)) \rightarrow Y_1^n(j)$$

form a Markov chain for all $l_2 \neq 1$ and $m \neq 1$, step (c) follows by the independence of the sequences and the encoding procedure, while (d) follows from Lemma 1. It follows that $\mathbb{P}(\mathcal{E}_4)$ tends to zero as $n \rightarrow \infty$ if

$$R_2 + R_3 < I(X_2, U_3; X_1, Y_1) + I(X_2; U_3) - \delta(\epsilon).$$

By similar steps, the decoding error probability at node 2 goes to zero as $n \rightarrow \infty$ if $R_1 < I(X_1, U_3; Y_2 | X_2) - \delta(\epsilon)$ and

$$R_1 + R_3 < I(X_1, U_3; X_2, Y_2) + I(X_1; U_3) - \delta(\epsilon).$$

Finally, by eliminating R_3 from the above inequalities, the probability of error tends to zero as $n \rightarrow \infty$ if the conditions in Theorem 3 are satisfied.

APPENDIX D
PROOF OF THEOREM 4

The achievability proof of Theorem 4 uses b transmissions blocks, each consisting of n transmissions, as in the proof of the multihop lower bound for the relay channel [42]. A sequence of $(b - 1)$ messages $m_j \in [1 : 2^{nR}]$, $j \in [1 : b - 1]$, each selected independently and uniformly over $[1 : 2^{nR}]$ is sent over b blocks. Note that the average rate over the b blocks is $\frac{(b-1)}{b}R$, which can be made arbitrarily close to R by letting $b \rightarrow \infty$.

Codebook Generation: Let $\epsilon > \epsilon' > 0$. Fix $p(x_1)p(u_2|y_2)p(u_3|y_3)$ and two encoding functions $x_2(u_2, y_2)$ and $x_3(u_3, y_3)$. We randomly and independently generate a codebook for each block. For $j \in [1 : b]$, randomly and independently generate 2^{nR} sequences $x_1^n(m_j)$, $m_j \in [1 : 2^{nR}]$, each according to $\prod_{i=1}^n p_{X_1}(x_{1i})$. For each $k = 1, 2$, randomly and independently generate 2^{nR_k} sequences $u_k^n(l_{kj})$, $l_{kj} \in [1 : 2^{nR_k}]$, each according to $\prod_{i=1}^n p_{U_k}(u_{ki})$.

Encoding: Let $m_j \in [1 : 2^{nR}]$ be the independent message to be sent in block $j \in [1 : b - 1]$. Then, the source node transmits $x_1^n(m_j)$.

Relay Encoding: Upon receiving $y_k^n(j)$ in block $j \in [1 : b - 1]$ relay node k , $k = 1, 2$, finds an index l_{kj} such that $(u_k^n(l_{kj}), y_k^n(j)) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there is more than one index, it chooses one of them at random. If there is no such index, it chooses an arbitrary index at random from $[1 : 2^{nR_k}]$. In block $j + 1$, relay k then transmits $x_{ki} = x_{ki}(u_{ki}(l_{kj}), y_{ki}(j))$ for $i \in [1 : n]$.

Decoding: Upon receiving $y_4^n(j)$, $j \in [2 : b]$, the decoder finds the unique message \hat{m}_{j-1} such that

$$(x_1^n(\hat{m}_{j-1}), u_2^n(l_{2,(j-1)}), u_3^n(l_{3,(j-1)}), y_4^n(j)) \in \mathcal{T}_{\epsilon}^{(n)},$$

for some $l_{2,(j-1)} \in [1 : 2^{nR_2}]$ and $l_{3,(j-1)} \in [1 : 2^{nR_3}]$.

Analysis of the probability of error: We analyze the probability of decoding error for the message M_{j-1} in block j , $j = 2, \dots, b$, averaged over the random codebooks and index assignments. Let $L_{2,(j-1)}$ and $L_{3,(j-1)}$ be the random variables denoting the indexes chosen in block $j - 1$ at relay 2 and 3, respectively. The decoder makes an error only if one or more of the following events occur:

$$\mathcal{E}_1 = \{(Y_2^n(j-1), U_2^n(l_2)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l_2\},$$

$$\begin{aligned}
\mathcal{E}_2 &= \{(Y_3^n(j-1), U_3^n(l_3)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } l_3\}, \\
\mathcal{E}_3 &= \{(X_1^n(M_{j-1}), U_2^n(L_{2,(j-1)}), U_3^n(L_{3,(j-1)}), Y_4^n(j)) \notin \mathcal{T}_\epsilon^{(n)}\}, \\
\mathcal{E}_4 &= \{(X_1^n(m), U_2^n(L_{2,(j-1)}), U_3^n(L_{3,(j-1)}), Y_4^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \neq M_{j-1}\}, \\
\mathcal{E}_5 &= \{(X_1^n(m), U_2^n(l_2), U_3^n(L_{3,(j-1)}), Y_4^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } l_2 \neq L_{2,(j-1)}, m \neq M_{j-1}\}, \\
\mathcal{E}_6 &= \{(X_1^n(m), U_2^n(L_{2,(j-1)}), U_3^n(l_3), Y_4^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } l_3 \neq L_{3,(j-1)}, m \neq M_{j-1}\}, \\
\mathcal{E}_7 &= \{(X_1^n(m), U_2^n(l_2), U_3^n(l_3), Y_4^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } l_2 \neq L_{2,(j-1)}, l_3 \neq L_{3,(j-1)}, \\
&\quad m \neq M_{j-1}\}.
\end{aligned}$$

Then by the union of events bound, the probability of decoding error is upper bounded as

$$\mathbb{P}(\hat{M}_{j-1} \neq M_{j-1}) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) + \mathbb{P}(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c) + \mathbb{P}(\mathcal{E}_4) + \mathbb{P}(\mathcal{E}_5) + \mathbb{P}(\mathcal{E}_6) + \mathbb{P}(\mathcal{E}_7).$$

By the covering lemma, $\mathbb{P}(\mathcal{E}_1)$ and $\mathbb{P}(\mathcal{E}_2)$ tend to zero as $n \rightarrow \infty$, if

$$R_2 > I(U_2; Y_2) + \delta(\epsilon'),$$

$$R_3 > I(U_3; Y_3) + \delta(\epsilon'),$$

respectively. By the Markov lemma, $\mathbb{P}(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c)$ tends to zero as $n \rightarrow \infty$.

By the symmetry of the random codebooks generation and random index assignment at the relays, it suffices to consider the conditional probabilities of the remaining error events conditioned on the event that

$$\mathcal{M} = \{M_{j-1} = 1, L_{2,(j-1)} = 1, L_{3,(j-1)} = 1\}. \quad (24)$$

Then, by the packing lemma, $\mathbb{P}(\mathcal{E}_4)$ tends to zero as $n \rightarrow \infty$ if

$$R < I(X_1; U_2, U_3, Y_4) - \delta(\epsilon).$$

Next, we next bound $\mathbb{P}(\mathcal{E}_5)$. Let $\tilde{U}^n = (U_2^n(1), Y_2^n(j-1), Y_3^n(j-1))$ and $\tilde{u}^n = (\tilde{u}_2^n, \tilde{y}_2^n, \tilde{y}_3^n)$ in short. Then, by the law of total probability, for $l \neq 1$ and $m \neq 1$,

$$\begin{aligned}
&\mathbb{P}\{X_1^n(m) = x_1^n, U_2^n(l_2) = u_2^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
&= \sum_{\tilde{u}^n} \mathbb{P}\{X_1^n(m) = x_1^n, U_2^n(l_2) = u_2^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n, \tilde{U}^n = \tilde{u}^n | \mathcal{M}\} \\
&= \sum_{\tilde{u}^n} \mathbb{P}\{U_2^n(l_2) = u_2^n | \mathcal{M}, X_1^n(m) = x_1^n, U_3^n(1) = u_3^n, \tilde{U}^n = \tilde{u}^n\}
\end{aligned}$$

$$\begin{aligned}
& \times \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, X_1^n(m) = x_1^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n\} \\
& \times \mathbb{P}\{X_1^n(m) = x_1^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
& = \sum_{\tilde{u}^n} \mathbb{P}\{U_2^n(l_2) = u_2^n | L_{2,(j-1)} = 1, U_2^n(1), Y_2^n(j-1)\} \\
& \quad \times \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, X_1^n(m) = x_1^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n\} \\
& \quad \times \mathbb{P}\{X_1^n(m) = x_1^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
& \leq (1 + \epsilon) \left(\prod_{i=1}^n p_{U_2}(u_{2i}) \right) \mathbb{P}\{X_1^n(m) = x_1^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\}
\end{aligned}$$

for n sufficiently large. Here, (a) follows from the fact that given \mathcal{M}

$$U_2^n(l_2) \rightarrow (U_2^n(L_{2,(j-1)}), U_3^n(L_{3,(j-1)}), Y_2^n(j-1), Y_3^n(j-1)) \rightarrow Y_4^n(j)$$

form a Markov chain for all $l_2 \neq L_{2,(j-1)}$, (b) follows by the independence of the sequences and the encoding procedure, and (c) follows by Lemma 1. It follows that, for n large enough,

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_5) &= \mathbb{P}\{(X_1^n(m), U_2^n(l_2), U_3^n(1), Y_4^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } l_2 \neq 1, m \neq 1 | \mathcal{M}\} \\
&\leq \sum_{m=2}^{2^{nR}} \sum_{l_2=2}^{2^{nR_2}} \sum_{(x_1^n, u_2^n, u_3^n, y_4^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{X_1^n(m) = x_1^n, U_2^n(l_2) = u_2^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
&\leq (1 + \epsilon) \sum_{m=2}^{2^{nR}} \sum_{l_2=2}^{2^{nR_2}} \sum_{(x_1^n, u_2^n, u_3^n, y_4^n) \in \mathcal{T}_\epsilon^{(n)}} p_{U_2^n}(u_2^n) \cdot \mathbb{P}\{X_1^n(m) = x_1^n, U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
&= (1 + \epsilon) \sum_{m=2}^{2^{nR}} \sum_{l_2=2}^{2^{nR_2}} \sum_{(x_1^n, u_2^n, u_3^n, y_4^n) \in \mathcal{T}_\epsilon^{(n)}} p_{U_2^n}(u_2^n) \cdot p_{X_1^n}(x_1^n) \mathbb{P}\{U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\}, \\
&= (1 + \epsilon) 2^{n(R+R_2)} \sum_{(u_2^n, y_4^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{U_3^n(1) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\} 2^{-n(I(X_1, U_2; U_3, Y_4) + I(X_1; U_2) - \delta(\epsilon))},
\end{aligned}$$

which implies that $\mathbb{P}(\mathcal{E}_5)$ tends to zero as $n \rightarrow \infty$ if

$$R + R_2 < I(X_1, U_2; U_3, Y_4) + I(X_1; U_2) - \delta(\epsilon).$$

$\mathbb{P}(\mathcal{E}_6)$ can be bounded in a similar manner, provided that the subscripts 1 and 2 are interchanged in the upper bound for $\mathbb{P}(\mathcal{E}_5)$. It follows that $\mathbb{P}(\mathcal{E}_6)$ tends to zero as $n \rightarrow \infty$ if

$$R + R_3 < I(X_1, U_3; U_2, Y_4) + I(X_1; U_3) - \delta(\epsilon)$$

Next, we bound $\mathbb{P}(\mathcal{E}_7)$. Let $\tilde{U}^n = (U_2^n(1), U_3^n(1), Y_2^n(j-1), Y_3^n(j-1))$ and $\tilde{u}^n = (\tilde{u}_2^n, \tilde{u}_3^n, \tilde{y}_2^n, \tilde{y}_3^n)$ in short. Then, by the law of total probability, for $l \neq 1$ and $m \neq 1$,

$$\mathbb{P}\{X_1^n(m) = x_1^n, U_2^n(l_2) = u_2^n, U_3^n(l_3) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\}$$

$$\begin{aligned}
&= \sum_{\tilde{u}^n} \mathbb{P}\{X_1^n(m) = x_1^n, U_2^n(l_2) = u_2^n, U_3^n(l_3) = u_3^n, Y_4^n(j) = y_4^n, \tilde{U}^n = \tilde{u}^n | \mathcal{M}\} \\
&= \sum_{\tilde{u}^n} \mathbb{P}\{U_2^n(l_2) = u_2^n, U_3^n(l_3) = u_3^n | \mathcal{M}, X_1^n(m) = x_1^n, \tilde{U}^n = \tilde{u}^n\} \\
&\quad \times \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, X_1^n(m) = x_1^n, Y_4^n(j) = y_4^n\} \\
&\quad \times \mathbb{P}\{X_1^n(m) = x_1^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
&= \sum_{\tilde{u}^n} \mathbb{P}\{U_2^n(l_2) = u_2^n | L_{2,(j-1)} = 1, U_2^n(1), Y_2^n(j-1)\} \\
&\quad \times \mathbb{P}\{U_3^n(l_3) = u_3^n | L_{3,(j-1)} = 1, U_3^n(1), Y_3^n(j-1)\} \\
&\quad \times \mathbb{P}\{\tilde{U}^n = \tilde{u}^n | \mathcal{M}, X_1^n(m) = x_1^n, Y_4^n(j) = y_4^n\} \\
&\quad \times \mathbb{P}\{X_1^n(m) = x_1^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
&\leq (1 + \epsilon) \left(\prod_{i=1}^n p_{U_2}(u_{2i}) p_{U_3}(u_{3i}) \right) \mathbb{P}\{X_1^n(m) = x_1^n, Y_4^n(j) = y_4^n | \mathcal{M}\}
\end{aligned}$$

for n sufficiently large. given \mathcal{M}

$$(U_2^n(l_2), U_2^n(l_3)) \rightarrow (U_2^n(L_{2,(j-1)}), U_3^n(L_{3,(j-1)}), Y_2^n(j-1), Y_3^n(j-1)) \rightarrow Y_4^n(j)$$

form a Markov chain for all $l_2 \neq L_{2,(j-1)}$ and $l_3 \neq L_{3,(j-1)}$, (b) follows by the independence of the sequences and the encoding procedure, and (c) follows by applying Lemma 1 twice. It follows that, for n large enough,

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_5) &= \mathbb{P}\{(X_1^n(m), U_2^n(l_2), U_3^n(l_3), Y_4^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } l_2 \neq 1, l_3 \neq 1, m \neq 1 | \mathcal{M}\} \\
&\leq \sum_{m=2}^{2^{nR}} \sum_{l_2=2}^{2^{nR_2}} \sum_{l_3=2}^{2^{nR_3}} \sum_{(x_1^n, u_2^n, u_3^n, y_4^n) \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{X_1^n(m) = x_1^n, U_2^n(l_2) = u_2^n, U_3^n(l_3) = u_3^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
&\leq (1 + \epsilon) \sum_{m=2}^{2^{nR}} \sum_{l_2=2}^{2^{nR_2}} \sum_{l_3=2}^{2^{nR_3}} \sum_{(x_1^n, u_2^n, u_3^n, y_4^n) \in \mathcal{T}_\epsilon^{(n)}} p_{U_2^n}(u_2^n) p_{U_3^n}(u_3^n) \cdot \mathbb{P}\{X_1^n(m) = x_1^n, Y_4^n(j) = y_4^n | \mathcal{M}\} \\
&\leq (1 + \epsilon) \sum_{m=2}^{2^{nR}} \sum_{l_2=2}^{2^{nR_2}} \sum_{l_3=2}^{2^{nR_3}} \sum_{(x_1^n, u_2^n, u_3^n, y_4^n) \in \mathcal{T}_\epsilon^{(n)}} p_{U_2^n}(u_2^n) p_{U_3^n}(u_3^n) p_{X_1^n}(x_1^n) \cdot \mathbb{P}\{Y_4^n(j) = y_4^n | \mathcal{M}\} \\
&= (1 + \epsilon) 2^{n(R+R_2+R_3)} \sum_{y_4^n \in \mathcal{T}_\epsilon^{(n)}} \mathbb{P}\{Y_4^n(j) = y_4^n | \mathcal{M}\} 2^{-n(I(X_1, U_2, U_3; Y_4) + I(X_1; U_2) + I(X_1, U_2; U_3) - \delta(\epsilon))},
\end{aligned}$$

which implies that $\mathbb{P}(\mathcal{E}_5)$ tends to zero as $n \rightarrow \infty$ if

$$R + R_2 + R_3 < I(X_1, U_2, U_3; Y_4) + I(X_1; U_2) + I(X_1, U_2; U_3) - \delta(\epsilon),$$

Finally, by eliminating R_2 and R_3 , the probability of error tends to zero as $n \rightarrow \infty$ if the conditions in Theorem 4 are satisfied.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
- [2] —, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Int. Conv. Rec., part 4*, 1959, vol. 7, pp. 142–163, reprinted with changes in *Information and Decision Processes*, R. E. Machol, Ed. New York: McGraw-Hill, 1960, pp. 93–126.
- [3] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [4] L. Song, R. Yeung, and N. Cai, "A separation theorem for single-source network coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1861–1871, May 2006.
- [5] A. Ramamoorthy, K. Jain, P. Chou, and M. Effros, "Separating distributed source coding from network coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2785–2795, June 2006.
- [6] M. Agarwal and S. K. Mitter, "Communication to within a fidelity criterion over unknown networks by reduction to reliable communication problems over unknown networks," 2010. [Online]. Available: <http://arxiv.org/abs/1002.1300>
- [7] S. Jalali and M. Effros, "On the separation of lossy source-network coding and channel coding in wireline networks," in *Proc. IEEE International Symposium on Information Theory*, June 2010, pp. 500–504.
- [8] C. Tian, J. Chen, S. Diggavi, and S. Shamai, "Optimality and approximate optimality of source–channel separation in networks," 2010, submitted to *IEEE Trans. Inf. Theory*, 2010. [Online]. Available: <http://arxiv.org/abs/1004.2648>
- [9] T. M. 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.
- [10] K. de Bruyn, V. Prelov, and E. van der Meulen, "Reliable transmission of two correlated sources over an asymmetric multiple-access channel (corresp.)," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 716 – 718, Sep. 1987.
- [11] R. Rajesh, V. Varshneya, and V. Sharma, "Distributed joint source channel coding on a multiple access channel with side information," in *Proc. IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008, pp. 2707–2711.
- [12] R. Rajesh, V. Sharma, and V. K. Varshenya, "Joint source-channel coding on a multiple access channel with side information," 2009, submitted to *IEEE Trans. Inf. Theory*, 2007. [Online]. Available: <http://arxiv.org/pdf/0904.4006/>
- [13] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian over a Gaussian MAC," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2714–2752, 2010.
- [14] —, "Sending a bivariate Gaussian source over a Gaussian MAC with feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1852–1864, 2010.
- [15] S. H. Lim, P. Minero, and Y.-H. Kim, "Lossy communication of correlated sources over multiple access channels," in *Proc. 48th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, Oct. 2010.
- [16] A. Jain, D. Gündüz, S. Kulkarni, H. Poor, and S. Verdú, "Energy-distortion tradeoffs in Gaussian joint source-channel coding problems," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3153–3168, May 2012.
- [17] T. S. Han and M. M. H. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 641–650, Sep. 1987.
- [18] U. Mittal and N. Phamdo, "Hybrid digital-analog (HDA) joint source-channel codes for broadcasting and robust communications," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1082–1102, May 2002.
- [19] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.

- [20] K. Wei and G. Kramer, "Broadcast channel with degraded source random variables and receiver side information," in *Proc. IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008, pp. 1711–1715.
- [21] G. Kramer and C. Nair, "Comments on "Broadcast channels with arbitrarily correlated sources"," in *Proc. IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009, pp. 2777–2779.
- [22] P. Minero and Y.-H. Kim, "Correlated sources over broadcast channels," in *Proc. IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009, pp. 2780–2784.
- [23] G. Kramer, Y. Liang, and S. Shamai, "Outer bounds on the admissible source region for broadcast channels with dependent sources," in *Information Theory and Applications Workshop, 2009*, Feb. 2009, pp. 169–172.
- [24] R. Soundararajan and S. Vishwanath, "Hybrid coding for Gaussian broadcast channels with Gaussian sources," in *Proc. IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009, pp. 2790 –2794.
- [25] C. Tian, S. Diggavi, and S. Shamai, "The achievable distortion region of bivariate Gaussian source on Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6419–6427, Oct. 2011.
- [26] J. Nayak, E. Tuncel, and D. Gündüz, "Wyner-Ziv coding over broadcast channels: Digital schemes," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1782–1799, Apr. 2010.
- [27] Y. Gao and E. Tuncel, "Wyner-Ziv coding over broadcast channels: hybrid digital/analog schemes," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5660–5672, Sept. 2011.
- [28] W. Liu and B. Chen, "Communicating correlated sources over interference channels: The lossy case," in *Proc. IEEE International Symposium on Information Theory*, Austin, Texas, June 2010, pp. 345 –349.
- [29] —, "Interference channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 8027–8037, Dec. 2011.
- [30] D. Gündüz, E. Erkip, A. Goldsmith, and H. Poor, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3927–3944, Sept. 2009.
- [31] T. Coleman, E. Martinian, and E. Ordentlich, "Joint source-channel coding for transmitting correlated sources over broadcast networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3864–3868, Aug. 2009.
- [32] D. Gündüz, E. Erkip, A. Goldsmith, and H. Poor, "Reliable joint source–channel cooperative transmission over relay networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2442–2458, Apr. 2013.
- [33] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [34] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, 1958.
- [35] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell System Tech. J.*, vol. 53, pp. 1681–1721, 1974.
- [36] M. Wilson, K. Narayanan, and G. Caire, "Joint source channel coding with side information using hybrid digital analog codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4922–4940, Oct. 2010.
- [37] G. Dueck, "A note on the multiple access channel with correlated sources," *IEEE Trans. Inf. Theory*, vol. 27, no. 2, pp. 232–235, 1981.
- [38] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.
- [39] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *IEEE International Symposium on Information Theory, 2000*, 2000, p. 22.
- [40] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

- [41] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE International Symposium on Information Theory*, Seattle, WA, Jul. 2006, pp. 1668–1672.
- [42] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [43] A. Orłitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, 2001.
- [44] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. 11, pp. 558–567, 1965.
- [45] P. Minero, S. H. Lim, and Y.-H. Kim, "Hybrid coding: An interface for joint source-channel coding and network communication," 2013, to be submitted to *Foundations and Trends in Communications and Information Theory*.
- [46] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [47] N. Merhav and S. Shamai, "On joint source-channel coding for the Wyner-Ziv source and the Gelfand-Pinsker channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2844–2855, Nov. 2003.
- [48] A. Sutivong, M. Chiang, T. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1486–1495, April 2005.
- [49] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978.
- [50] S.-Y. Tung, "Multiterminal source coding," Ph.D. Thesis, Cornell University, Ithaca, NY, 1978.
- [51] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Control Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.
- [52] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, pp. 100–113, 1975.
- [53] A. B. Wagner, B. G. Kelly, and Y. Altuğ, "Distributed rate-distortion with common components," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4035–4057, Aug. 2011.
- [54] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell System Tech. J.*, vol. 52, pp. 1037–1076, Sep. 1973.
- [55] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, 1996.
- [56] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, 1997.
- [57] V. Prabhakaran, D. N. C. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. IEEE International Symposium on Information Theory*, Chicago, IL, June/July 2004, p. 117.
- [58] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.
- [59] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network," *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.
- [60] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, 1979.
- [61] M. R. Aref, "Information flow in relay networks," Ph.D. Thesis, Stanford University, Stanford, CA, Oct. 1980.
- [62] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

- [63] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.
- [64] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity of the Gaussian two-way relay channel within 1/2 bit," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5488–5494, Nov. 2010.
- [65] J. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.
- [66] M. Khormuji and M. Skoglund, "On instantaneous relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3378–3394, July 2010.
- [67] —, "Hybrid digital-analog noisy network coding," in *2011 International Symposium on Network Coding (NetCod)*, July 2011, pp. 1–5.
- [68] A. El Gamal, "On information flow in relay networks," in *Proc. IEEE National Telecom Conference*, Nov. 1981, vol. 2, pp. D4.1.1–D4.1.4.
- [69] S. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
- [70] A. H. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 828–840, 1982.
- [71] A. B. Wagner, B. G. Kelly, and Y. Altuğ, "The lossy one-helper conjecture is false," in *Proc. 47th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, Sep. 2009.
- [72] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, 1981.