Noisy Network Coding

Sung Hoon Lim, Young-Han Kim, Abbas El Gamal, and Sae-Young Chung

Abstract

A noisy network coding scheme for sending multiple sources over a general noisy network is presented. For multi-source multicast networks, the scheme naturally extends both network coding over noiseless networks by Ahlswede, Cai, Li, and Yeung, and compress–forward coding for the relay channel by Cover and El Gamal to general discrete memoryless and Gaussian networks. The scheme also recovers as special cases the results on coding for wireless relay networks and deterministic networks by Avestimehr, Diggavi, and Tse, and coding for wireless erasure networks by Dana, Gowaikar, Palanki, Hassibi, and Effros. The scheme involves message repetition coding, relay signal compression, and simultaneous decoding. Unlike previous compress–forward schemes, where independent messages are sent over multiple blocks, the same message is sent multiple times using independent codebooks as in the network coding scheme for cyclic networks. Furthermore, the relays do not use Wyner–Ziv binning as in previous compress–forward schemes, and each decoder performs simultaneous joint typicality decoding on the received signals from all the blocks without explicitly decoding the compression indices. A consequence of this new scheme is that achievability is proved simply and more generally without resorting to time expansion to extend results for acyclic networks to networks with cycles. The noisy network coding scheme is then extended to general multi-source networks by combining it with decoding techniques for interference channels. For the Gaussian multicast network, noisy network coding improves the previously established gap to the cutset bound. We also demonstrate through two popular AWGN network examples that noisy network coding can outperform conventional compress–forward, amplify–forward, and hash–forward coding schemes.

S. H. Lim and S.-Y. Chung are with the Department of Electrical Engineering, KAIST, Daejeon 305-701, Korea (e-mail: sunghlim@kaist.ac.kr and sychung@ee.kaist.ac.kr).

Y.-H. Kim is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA (e-mail: yhk@ucsd.edu).

A. El Gamal is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA (e-mail: abbas@ee.stanford.edu).

I. INTRODUCTION

Consider the N-node discrete memoryless network depicted in Figure 1. Each node wishes to send a message to a set of destination nodes while acting as a relay for messages from other nodes. What is the capacity region of this network, that is, the set of rates at which the nodes can reliably communicate their messages? What is the coding scheme that achieves the capacity region? These questions are at the heart of network information theory, yet complete answers remain elusive.
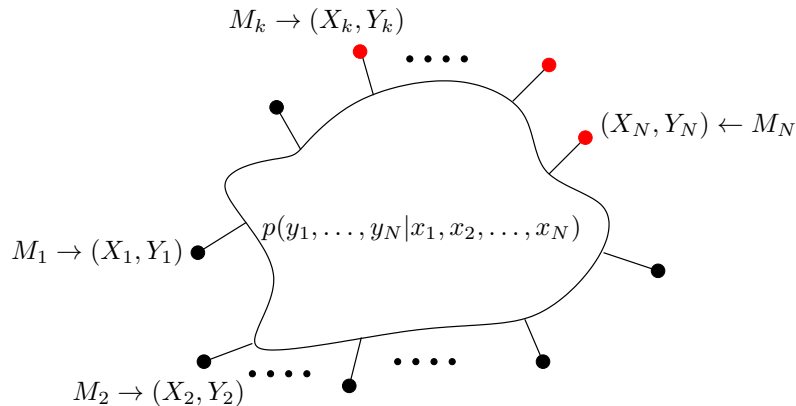

Fig. 1. An N-node discrete memoryless network.

Some progress has been made toward answering these questions in the past forty years. In [1], [2], a general cutset outer bound on the capacity region of this network was established. This bound generalizes the max-flow min-cut theorem for noiseless single-source unicast networks [3], [4], and has been shown to be tight for several other classes of networks.

In their seminal paper on network coding [5], Ahlswede, Cai, Li, and Yeung showed that the capacity of noiseless single-source multicast networks coincides with the cutset bound, thus generalizing the max-flow min-cut theorem to multiple destinations. Each relay in the network coding scheme sends a function of its incoming signals over each outgoing link instead of simply forwarding incoming signals. Their proof of the *network coding theorem* is done in two steps. For acyclic networks, relay mappings are randomly generated and they show that the message is correctly decoded with high probability provided the rate is below the cutset bound. This proof is then extended to cyclic networks by constructing an acyclic *time-expanded* network and relating achievable rates and codes for the time-expanded network to those for the original cyclic network.

The network coding theorem has been extended in several directions. Dana, Gowaikar, Palanki, Hassibi, and Effros [6] studied the multiple-source multicast erasure network as a simple model for a wireless

data network with packet loss. They showed that the capacity region coincides with the cutset bound and is achieved via network coding. Ratnakar and Kramer [7] extended network coding to characterize the multicast capacity for single-source deterministic networks with broadcast but no interference at the receivers. Avestimehr, Diggavi, and Tse [8] further extended this result to deterministic networks with broadcast *and* interference to obtain a lower bound on capacity that coincides with the cutset bound when the channel output is a linear function of input signals over a finite field. Their proof is again done in two steps. As in the original proof of the network coding theorem, random coding is used to establish the lower bound for *layered* deterministic networks. A time-expansion technique is then used to extend the capacity lower bound to arbitrary nonlayered deterministic networks.

In an earlier and seemingly unrelated line of investigation, van der Meulen [9] introduced the relay channel with a single source X_1, single destination Y_3, and single relay with transmitter–receiver pair (X_2, Y_2). Although the capacity for this channel is still not known in general, several nontrivial upper and lower bounds have been developed. In [10], Cover and El Gamal proposed the compress–forward coding scheme in which the relay compresses its noisy observation of the source signal and forwards the compressed description to the destination. Despite its simplicity, compress–forward was shown to be optimal for classes of deterministic [11] and modulo-sum [12] relay channels. The Cover–El Gamal compress–forward lower bound on capacity has the form

$$C \geq \max_{p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)} I(X_1; \hat{Y}_2, Y_3 | X_2), \tag{1}$$

where the maximum is over all pmfs $p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)$ such that $I(X_2; Y_3) \geq I(Y_2; \hat{Y}_2 | X_2, Y_3)$. This lower bound was established using a block Markov coding scheme—in each block the sender transmits a new message, and the relay compresses its received signal and sends the bin index of the compression index to the receiver using Wyner–Ziv coding [13]. Decoding is performed sequentially. At the end of each block, the receiver first decodes the compression index and then uses it to decode the message sent in the previous block. Kramer, Gastpar, and Gupta [14] used an extension of this scheme to establish a compress–forward lower bound on the capacity of general relay networks. Around the same time, El Gamal, Mohseni, and Zahedi [15] put forth the equivalent characterization of the compress–forward lower bound

$$C \geq \max_{p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)} \min\{I(X_1; \hat{Y}_2, Y_3 | X_2),\ I(X_1, X_2; Y_3) - I(Y_2; \hat{Y}_2 | X_1, X_2, Y_3)\}. \tag{2}$$

As we will see, this characterization motivates a more general way to extend compress–forward to networks.

In this paper, we describe a noisy network coding scheme that extends and unifies the above results. On the one hand, the scheme naturally extends compress–forward coding to noisy networks. The resulting inner bound on the capacity region extends the equivalent characterization in (2), rather than the original characterization in (1). On the other hand, our scheme includes network coding and its variants as special cases. Hence, while the coding schemes for deterministic networks and erasure networks can be viewed as bottom-up generalizations of network coding to more complicated networks, our coding scheme represents a top-down approach for general noisy networks.

The noisy coding scheme employs block Markov message repetition coding, relay signal compression, and simultaneous decoding. Instead of sending different messages over multiple blocks and decoding one message at a time as in previous compress–forward coding schemes [10], [14], the source transmits the same message over multiple blocks using independently generated codebooks. Although a similar message repetition scheme is implicitly used in the time expansion technique for cyclic noiseless networks [5] and nonlayered deterministic networks [8], our achievability proof does not require a two-step approach that depends on the network topology. The relay operation is also simpler than previous compress–forward schemes—the compression index of the received signal in each block is sent without Wyner–Ziv binning. After receiving the signals from all the blocks, each destination node performs simultaneous joint typicality decoding of the messages without explicitly decoding the compression indices. As we will demonstrate, this results in better performance than previous schemes in [14], [16], [17], [18], [19] for networks with more than one relay node or multiple messages.

The simplicity of our scheme makes it straightforward to combine with decoding techniques for interference channels. Indeed, the noisy network coding scheme can be viewed as transforming a multi-hop relay network into a single-hop interference network where the channel outputs are compressed versions of the received signals. We develop two coding schemes for general multiple source networks based on this observation. At one extreme, noisy network coding is combined with decoding all messages, while at the other, interference is treated as noise.

We apply these noisy network coding schemes to Gaussian networks. For the multiple-source multi-cast case, we establish an inner bound that improves upon previous capacity approximation results by Avestimehr, Diggavi, and Tse [8] and Perron [20] with a tighter gap to the cutset bound. We then show that noisy network coding can outperform other specialized schemes for two-way relay channels [16], [17] and interference relay channels [18], [19].

The rest of the paper is organized as follows. In the next section, we formally define the problem of communicating multiple sources over a general network and discuss the main results. We also show that

previous results on network coding are special cases of our main theorems and compare noisy network coding to other schemes. In Section III, we present the noisy network coding scheme for multiple-source multicast networks. In Section IV, the scheme is extended to general multiple-source networks. Results on Gaussian networks are discussed in Section V.

Throughout the paper, we follow the notation in [21]. In particular, a sequence of random variables with node index k and time index $i \in [1 : n]$ is denoted as $X_k^n = (X_{k1}, \ldots, X_{kn})$. A set of random variables is denoted as $X(\mathcal{A}) = \{X_k : k \in \mathcal{A}\}$.

II. PROBLEM SETUP AND MAIN RESULTS

The N-node discrete memoryless network (DMN) $(\prod_{k=1}^{N} \mathcal{X}_k, p(y^N | x^N), \prod_{k=1}^{N} \mathcal{Y}_k)$ depicted in Figure 1 consists of N sender–receiver alphabet pairs $(\mathcal{X}_k, \mathcal{Y}_k)$, $k \in [1 : N] := \{1, \ldots, N\}$, and a collection of conditional pmfs $p(y_1, \ldots, y_N | x_1, \ldots, x_N)$. Each node $k \in [1 : N]$ wishes to send a message M_k to a set of destination nodes, $\mathcal{D}_k \subseteq [1 : N]$. Formally, a $(2^{nR_1}, \ldots, 2^{nR_N}, n)$ code for a DMN consists of N message sets $[1 : 2^{nR_1}], \ldots, [1 : 2^{nR_N}]$, a set of encoders with encoder $k \in [1 : N]$ that assigns an input symbol x_{ki} to each pair (m_k, y_k^{i-1}) for $i \in [1 : n]$, and a set of decoders with decoder $d \in \cup_{k=1}^{N} \mathcal{D}_k$ that assigns message estimates $(\hat{m}_{kd} : k \in \mathcal{S}_d)$ to each (y_d^n, m_d), where $\mathcal{S}_d := \{k : d \in \mathcal{D}_k\}$ is the set of nodes that send messages to destination d. For simplicity we assume $d \in \mathcal{S}_d$ for all destination nodes.

We assume that the messages M_k, $k \in [1 : N]$, are independent of each other and each message is uniformly distributed over its message set. The average probability of error is defined as

$$P_e^{(n)} = \mathsf{P}\{\hat{M}_{kd} \neq M_k \text{ for some } d \in \mathcal{D}_k, k \in [1 : N]\}.$$

A rate tuple $(R_1, \ldots, R_N)$ is said to be achievable if there exists a sequence of $(2^{nR_1}, \ldots, 2^{nR_N}, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity region of the DMN is the closure of the set of achievable rate tuples.

We are ready to state our main results.

Multiple-source multicast networks: In Section III, we establish the following noisy network coding theorem for multicasting multiple sources over a DMN. The coding scheme and techniques used to prove this theorem, which we highlighted earlier, constitute the key contributions of our paper.

Theorem 1: Let $\mathcal{D} = \mathcal{D}_1 = \cdots = \mathcal{D}_N$. A rate tuple $(R_1, \ldots, R_N)$ is achievable for the DMN $p(y^N | x^N)$ if there exists some joint pmf $p(q) \prod_{k=1}^{N} p(x_k | q) p(\hat{y}_k | y_k, x_k, q)$ such that

$$R(\mathcal{S}) < \min_{d \in \mathcal{S}^c \cap \mathcal{D}} I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c), Q) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d, Q) \tag{3}$$

for all *cutsets* $\mathcal{S} \subseteq [1 : N]$ with $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$, where $R(\mathcal{S}) = \sum_{k \in \mathcal{S}} R_k$.

This inner bound has a similar structure to the cutset outer bound given by

$$R(\mathcal{S}) \leq I(X(\mathcal{S}); Y(\mathcal{S}^c)|X(\mathcal{S}^c)) \tag{4}$$

for all $\mathcal{S} \subseteq [1 : N]$ with $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$. The first term of (4), however, has Y replaced by the "compressed" version $\hat{Y}$. Another difference between the bounds is the negative term appearing in (3), which quantifies the rate requirement to convey the compressed version. In addition, the maximum in (3) is only over independent X^N.

Theorem 1 can be specialized to several important network models as follows:

Noiseless networks: Consider a noiseless network modeled by a weighted directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{C})$, where $\mathcal{N} = [1 : N]$ is the set of nodes, $\mathcal{E} \subseteq [1 : N] \times [1 : N]$ is the set of edges, and $\mathcal{C} = \{C_{jk} \in \mathbb{R}^+ : (j, k) \in \mathcal{E}\}$ is the set of link capacities. Each edge $(j, k) \in \mathcal{E}$ carries an input symbol $x_{jk} \in \mathcal{X}_{jk}$ with link capacity $C_{jk} = \log |\mathcal{X}_{jk}|$, resulting in the channel output at node k as $Y_k = \{X_{jk} : (j, k) \in \mathcal{E}\}$. By setting $\hat{Y}_k = Y_k$ for all k and evaluating Theorem 1 with the uniform pmf on X^N, it can be easily shown that inner bound (3) coincides with the cutset bound, and thus the capacity region is the set of rate tuples $(R_1, \ldots, R_N)$ such that

$$R(\mathcal{S}) \leq \sum_{\substack{(j,k) \in \mathcal{E} \\ j \in \mathcal{S}, k \in \mathcal{S}^c}} C_{jk}. \tag{5}$$

This recovers previous results in [5] for the single-source case and [6] for the multiple-source case.

Relay channels: Consider the relay channel $p(y_2, y_3|x_1, x_2)$. It can be easily shown that the inner bound (3) reduces to the alternative characterization of the compress–forward lower bound in (2).

Erasure networks: Consider the erasure multiple-source multicast network in which the channel output at node $k \in [1 : N]$ is $Y_k = \{Y_{jk} : j \in [1 : N]\}$, where $Y_{jk} = \varepsilon$ if it is erased, and $Y_{jk} = X_j$, otherwise. Assume further that the network erasure pattern is known at the destination nodes. Taking $\hat{Y}_k = Y_k$, $k \in [1 : N]$ and the uniform pmf on X^N as in the noiseless case, inner bound (3) reduces to

$$R(\mathcal{S}) \leq \sum_{j \in \mathcal{S}} \left(\log |\mathcal{X}_j| (1 - P\{\text{link } (j, k) \text{ is erased for all } k \in \mathcal{S}^c\}) \right). \tag{6}$$

It can be also shown that the inner bound coincides with the cutset bound and thus characterizes the capacity region. This recovers the previous result in [6].

Deterministic networks: Suppose $Y_k = g_k(X_1, \ldots, X_N)$, $k \in [1 : N]$. By setting $\hat{Y}_k = Y_k$, $k \in [1 : N]$, Theorem 1 implies that a rate tuple $(R_1, \ldots, R_N)$ is achievable for the deterministic network if there

exists some pmf $p(q) \prod_{k=1}^{N} p(x_k|q)$ such that

$$R(\mathcal{S}) < I(X(\mathcal{S}); Y(\mathcal{S}^c)|X(\mathcal{S}^c), Q) = H(Y(\mathcal{S}^c)|X(\mathcal{S}^c), Q) \tag{7}$$

for all $\mathcal{S} \subseteq [1:N]$ with $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$. This recovers previous results in [8] for the single-source case and in [20] for the multiple-source case. Note that the lower bound (7) is tight when the cutset bound is attained by the product pmf, for example, as in the deterministic network without interference [7] or the finite-field linear deterministic network $Y_k = \sum_{j=1}^{N} g_{jk} X_j$ [8].

Note that in all the above special cases, the channel output at node k can be expressed as a deterministic function of the input symbols $(X_1, \ldots, X_N)$ and the destination output symbol Y_d, i.e.,

$$Y_k = g_{dk}(X_1, \ldots, X_N, Y_d) \quad \text{for every } k \in [1:N] \text{ and } d \in \mathcal{D}. \tag{8}$$

Under this structure, the inner bound in Theorem 1 can be simplified by substituting $\hat{Y}_k = Y_k$ for $k \in [1:N]$ in (3) to obtain the following generalization.

Corollary 1: Let $\mathcal{D} = \mathcal{D}_1 = \cdots = \mathcal{D}_N$. A rate tuple $(R_1, \ldots, R_N)$ is achievable for the semideterministic DMN (8) if there exists some joint pmf $p(q) \prod_{k=1}^{N} p(x_k|q)$ such that

$$R(\mathcal{S}) < I(X(\mathcal{S}); Y(\mathcal{S}^c)|X(\mathcal{S}^c), Q) \tag{9}$$

for all $\mathcal{S} \subseteq [1:N]$ with $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$.

We also show in Appendix C that our noisy network coding scheme can strictly outperform the extension of the original compress–forward scheme for the relay channel to networks in [14, Th 3].

General multiple-source networks: We extend the noisy network coding theorem to general multiple-source networks. As a first step, we note that Theorem 1 continues to hold for general networks with *multicast completion* of destination nodes, that is, when every message is decoded by all destination nodes $\mathcal{D} = \cup_{k=1}^{N} \mathcal{D}_k$. Thus, we can obtain an inner bound on the capacity region for the DMN in the same form as (3) with $\mathcal{D} = \cup_{k=1}^{N} \mathcal{D}_k$.

This multicast-completion inner bound can be improved by noting that noisy network coding transforms a multi-hop relay network $p(y^N|x^N)$ into a single-hop interference network $p(\tilde{y}^N|x^N)$, where the effective channel output at decoder k is $\tilde{Y}_k = (X_k, Y_k, \hat{Y}_1, \ldots, \hat{Y}_N)$ and the compressed channel outputs $(\hat{Y}_1, \ldots, \hat{Y}_N)$ are conveyed to decoders with some rate penalty. This observation leads to a modified decoding rule that does not require each destination to decode unintended messages correctly, resulting in the following improved inner bound.

Theorem 2: A rate tuple $(R_1, \ldots, R_N)$ is achievable for the DMN if there exists some joint pmf $p(q) \prod_{k=1}^{N} p(x_k|q) p(\hat{y}_k|y_k, x_k, q)$ such that

$$R(\mathcal{S}) < \min_{d \in \mathcal{S}^c \cap \mathcal{D}(\mathcal{S})} I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c), Q) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d, Q) \quad (10)$$

for all cutsets $\mathcal{S} \subseteq [1:N]$ with $\mathcal{S}^c \cap \mathcal{D}(\mathcal{S}) \neq \emptyset$, where $\mathcal{D}(\mathcal{S}) := \cup_{k \in \mathcal{S}} \mathcal{D}_k$.

The proof of this theorem is given in Subsection IV-A.

As an alternative, each destination node can simply treat interference as noise rather than decoding it. Using this approach, we establish the following inner bound on the capacity region.

Theorem 3: A rate tuple $(R_1, \ldots, R_N)$ is achievable for the DMN if there exists some joint pmf $p(q) \prod_{k=1}^{N} p(u_k, x_k|q) p(\hat{y}_k|y_k, u_k, q)$ with

$$R(\mathcal{T}) < I(X(\mathcal{T}), U(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{T}^c), U(\mathcal{S}^c), Q) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X(\mathcal{S}_d), U^N, \hat{Y}(\mathcal{S}^c), Y_d, Q) \quad (11)$$

for all $\mathcal{S} \subseteq [1:N]$, $d \in \mathcal{D}(\mathcal{S})$, and $\mathcal{S} \cap \mathcal{S}_d \subseteq \mathcal{T} \subseteq \mathcal{S}_d$ such that $\mathcal{S}^c \cap \mathcal{D}(\mathcal{S}) \neq \emptyset$, where $\mathcal{T}^c = \mathcal{S}_d \backslash \mathcal{T}$.

Unlike the coding schemes in Theorems 1 and 2 where each node maps both its own message and the compression index to a single codeword, here each node applies superposition coding [22] for forwarding the compression index along with its own message. (Note that when a node does not have its own message and it acts only as a relay, there is no difference in the relay operation from the previous schemes.) The details are given in Subsection IV-B.

Gaussian networks: In Section V, we present an extension of the above results to Gaussian networks and compare the performance of noisy network coding to other specialized coding schemes for two popular Gaussian networks.

Consider the Gaussian network

$$Y^N = GX^N + Z^N, \quad (12)$$

where $G \in \mathbb{R}^{N \times N}$ is the channel gain matrix and Z^N is a vector of independent Gaussian random variables with zero mean and unit variance. We further assume average power constraint P on each sender X_k.

In Subsection V-A, we establish the following result on the multicast capacity region of this general Gaussian network.

Theorem 4: Let $\mathcal{D} = \mathcal{D}_1 = \cdots = \mathcal{D}_N$. For any rate tuple $(R_1, \ldots, R_N)$ in the cutset outer bound, there exists $(R'_1, \ldots, R'_N)$ in the inner bound in Theorem 1 for the AWGN network (12) such that

$$\sum_{k \in \mathcal{S}} (R_k - R'_k) \leq \frac{|\mathcal{S}|}{2} + \frac{\min\{|\mathcal{S}|, |\mathcal{S}^c|\}}{2} \log(2|\mathcal{S}|)$$

for all $\mathcal{S} \subseteq [1:N]$ with $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$.

This theorem implies that the gap between the cutset bound and our inner bound is less than or equal to $(N/4)\log(2N)$ for $N > 3$, regardless of the values of the channel gain matrix G and power constraint P.

We also demonstrate through the following two examples that noisy network coding can outperform previous coding schemes, some of which are developed specifically for these channel models:

Two-way relay channel (Subsection V-B): Consider the AWGN two-way relay channel

$$Y_1 = g_{21}X_2 + g_{31}X_3 + Z_1,$$
$$Y_2 = g_{12}X_1 + g_{32}X_3 + Z_2, \tag{13}$$
$$Y_3 = g_{13}X_1 + g_{23}X_2 + Z_3,$$

where the channel gains are $g_{12} = g_{21} = 1$, $g_{13} = g_{31} = d^{-\gamma/2}$ and $g_{23} = g_{32} = (1-d)^{-\gamma/2}$, and $d \in [0,1]$ is the location of the relay node between nodes 1 and 2 (which are unit distance apart). Source nodes 1 and 2 wish to exchange messages reliably with the help of relay node 3. Various coding schemes for this channel have been investigated in [16], [17]. In Figure 2, we compare the performance of noisy network coding (Theorem 2) to amplify–forward (AF) and an extension of compress–forward (CF) for $d \in [0, 1/2]$ and $\gamma = 3$. Note that noisy network coding outperforms the other two schemes coinciding with the compress–forward only when the relay is midway between nodes 1 and 2 ($d = 1/2$) and when it coincides with node 1 ($d = 0$).
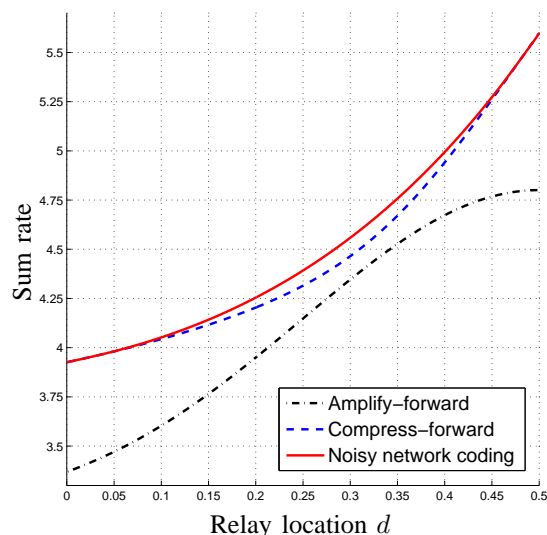

Fig. 2. Comparison of coding schemes for $P = 10$.

Interference relay channel (Subsection V-C): Consider the AWGN interference relay channel with orthogonal receiver components in Figure 3.
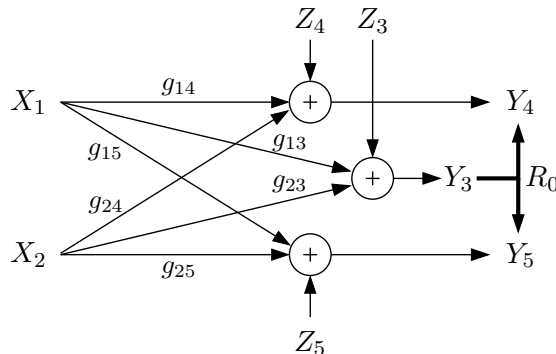

Fig. 3. AWGN interference relay channel.

The channel outputs are

$$Y_j = g_{1j}X_1 + g_{2j}X_2 + Z_j, \quad j = 3, 4, 5,$$

where g_{ij} is the channel gain of link (i, j). Source node 1 wishes to send a message to destination node 4, while source node 2 wishes to send a message to destination node 5. Relay node 3 helps the communication of this interference channel by sending some information about Y_3 over a common noiseless link of rate R_0 to both destination nodes. In Figure 4, we compare noisy network coding (Theorems 2 and 3) to compress–forward (CF) and hash–forward (HF) in [19]. The curve representing noisy network coding depicts the maximum of achievable sum rates in Theorems 2 and 3. Note that, although not shown in the figure, Theorem 3 alone outperforms the other two schemes for all channel gains and power constraints. At high signal-to-noise ratio (SNR), Theorem 2 provides further improvement, since decoding other messages is a better strategy when interference is strong.

III. NOISY NETWORK CODING FOR MULTICAST

To illustrate the main idea of the noisy network coding scheme and highlight the differences from the standard compress–forward coding scheme [10], [14], we first prove Theorem 1 for the 3-node relay channel and then extend the proof to general multicast networks.

Let $\mathbf{x}_{kj}$ denote $(x_{k,(j-1)n+1}, \ldots, x_{k,jn})$, $j \in [1 : b]$; thus $x_k^{bn} = (x_{k1}, \ldots, x_{k,nb}) = (\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kb}) = \mathbf{x}_k^b$. To send a message $m \in [1 : 2^{nbR}]$, the source node transmits $\mathbf{x}_{1j}(m)$ for each block $j \in [1 : b]$. In block j, the relay finds a "compressed" version $\hat{\mathbf{y}}_{2j}$ of the relay output $\mathbf{y}_{2j}$ conditioned on $\mathbf{x}_{2j}$, and
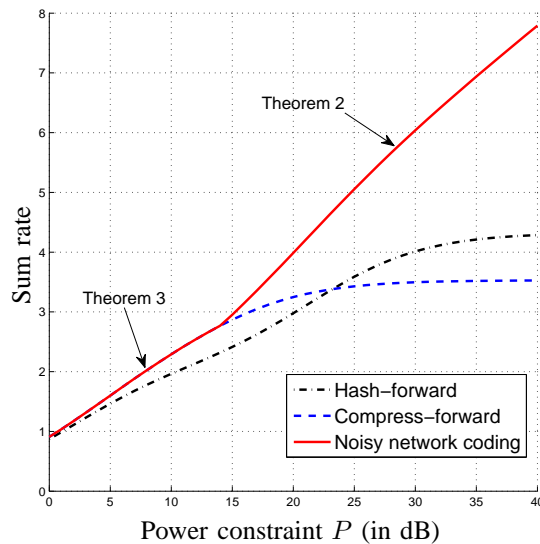
Fig. 4. Comparison of coding schemes for $g_{14} = g_{25} = 1$, $g_{15} = g_{24} = g_{13} = 0.5$, $g_{13} = 0.1$.

transmits a codeword $\mathbf{x}_{2,j+1}(\hat{\mathbf{y}}_{2j})$ in the next block. After b block transmissions, the decoder finds the correct message $m \in [1 : 2^{nbR}]$ using $(\mathbf{y}_{31}, \ldots, \mathbf{y}_{3b})$ by joint typical decoding for each of b blocks simultaneously. The details are as follows.

Codebook generation: Fix $p(x_1)p(x_2)p(\hat{y}_2|y_2, x_2)$. We randomly and independently generate a codebook for each block.

For each $j \in [1 : b]$, randomly and independently generate 2^{nbR} sequences $\mathbf{x}_{1j}(m)$, $m \in [1 : 2^{nbR}]$, each according to $\prod_{i=1}^{n} p_{X_1}(x_{1,(j-1)n+i})$. Similarly, randomly and independently generate $2^{n\hat{R}_2}$ sequences $\mathbf{x}_{2j}(l_{j-1})$, $l_{j-1} \in [1 : 2^{n\hat{R}_2}]$, each according to $\prod_{i=1}^{n} p_{X_2}(x_{2,(j-1)n+i})$. For each $\mathbf{x}_{2j}(l_{j-1})$, $l_{j-1} \in [1 : 2^{n\hat{R}_2}]$, randomly and conditionally independently generate $2^{n\hat{R}_2}$ sequences $\hat{\mathbf{y}}_{2j}(l_j|l_{j-1})$, $l_j \in [1 : 2^{n\hat{R}_2}]$, each according to $\prod_{i=1}^{n} p_{\hat{Y}_2|X_2}(\hat{y}_{2,(j-1)n+i}|x_{2,(j-1)n+i}(l_{j-1}))$.

This defines the codebook

$$\mathcal{C}_j = \left\{ \mathbf{x}_{1j}(m), \mathbf{x}_{2j}(l_{j-1}), \hat{\mathbf{y}}_{2j}(l_j|l_{j-1}) : m \in [1 : 2^{nbR}], l_j, l_{j-1} \in [1 : 2^{n\hat{R}_2}] \right\}$$

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

Encoding and decoding are explained with the help of Table I.

Encoding: Let m be the message to be sent. The relay, upon receiving $\mathbf{y}_{2j}$ at the end of block $j \in [1 : b]$,

Block	1	2	3	$\cdots$	$b-1$	b
X_1	$\mathbf{x}_{11}(m)$	$\mathbf{x}_{12}(m)$	$\mathbf{x}_{13}(m)$	$\ldots$	$\mathbf{x}_{1,b-1}(m)$	$\mathbf{x}_{1b}(m)$
Y_2	$\hat{\mathbf{y}}_{21}(l_1\|1), l_1$	$\hat{\mathbf{y}}_{22}(l_2\|l_1), l_2$	$\hat{\mathbf{y}}_{23}(l_3\|l_2), l_3$	$\ldots$	$\hat{\mathbf{y}}_{2,b-1}(l_{b-1}\|l_{b-2}), l_{b-1}$	$\hat{\mathbf{y}}_{2b}(l_b\|l_{b-1}), l_b$
X_2	$\mathbf{x}_{21}(1)$	$\mathbf{x}_{22}(l_1)$	$\mathbf{x}_{23}(l_2)$	$\ldots$	$\mathbf{x}_{2,b-1}(l_{b-2})$	$\mathbf{x}_{2b}(l_{b-1})$
Y_3	$\emptyset$	$\emptyset$	$\emptyset$	$\ldots$	$\emptyset$	$\hat{m}$

TABLE I

NOISY NETWORK CODING FOR THE RELAY CHANNEL.

finds an index l_j such that

$$(\hat{\mathbf{y}}_{2j}(l_j|l_{j-1}), \mathbf{y}_{2j}, \mathbf{x}_{2j}(l_{j-1})) \in \mathcal{T}_{\epsilon'}^{(n)},$$

where $l_0 = 1$ by convention. If there is more than one such index, choose one of them at random. If there is no such index, choose an arbitrary index at random from $[1 : 2^{n\hat{R}_2}]$. The codeword pair $(\mathbf{x}_{1j}(m), \mathbf{x}_{2j}(l_{j-1}))$ is transmitted in block $j \in [1 : b]$.

Decoding: Let $\epsilon > \epsilon'$. At the end of block b, the decoder finds a unique message $\hat{m} \in [1 : 2^{nbR}]$ such that

$$(\mathbf{x}_{1j}(\hat{m}), \hat{\mathbf{y}}_{2j}(\hat{l}_j|\hat{l}_{j-1}), \mathbf{x}_{2j}(\hat{l}_{j-1}), \mathbf{y}_{3j}) \in \mathcal{T}_{\epsilon}^{(n)} \quad \text{for all } j \in [1 : b]$$

for some $\hat{l}_1, \hat{l}_2, \ldots, \hat{l}_b$. If there is none or more than one such message, it declares an error.

Analysis of the probability of error: Let M denote the message sent at the source node and L_j denote the indices chosen by the relay at block $j \in [1 : b]$. Define

$$\mathcal{E}_0 := \bigcup_{j=1}^{b} \big\{ (\hat{\mathbf{Y}}_{2j}(l_j|L_{j-1}), \mathbf{X}_{2j}(L_{j-1}), \mathbf{Y}_{2j}) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l_j \big\},$$

$$\mathcal{E}_m := \big\{ (\mathbf{X}_{1j}(m), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_{\epsilon}^{(n)}, j \in [1 : b] \text{ for some } l_1, l_2, \ldots, l_b \big\}.$$

To bound the probability of error, assume without loss of generality that $M = 1$. Then the probability of error is upper bounded by

$$\mathsf{P}(\mathcal{E}) \leq \mathsf{P}(\mathcal{E}_0) + \mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_1^c) + \mathsf{P}(\cup_{m \neq 1} \mathcal{E}_m).$$

By the covering lemma [21], $\mathsf{P}(\mathcal{E}_0) \to 0$ as $n \to \infty$, if $R_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(\epsilon')$. By the conditional typicality lemma [21], $\mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_1^c) \to 0$ as $n \to \infty$.

To bound $\mathsf{P}(\cup_{m \neq 1} \mathcal{E}_m)$, assume without loss of generality that $(L_1, \ldots, L_b) = (1, \ldots, 1)$; recall the symmetry of the codebook construction. For $j \in [1 : b]$, $m \in [1 : 2^{nbR}]$, and $l_{j-1}, l_j \in [1 : 2^{n\hat{R}_2}]$, define the events

$$\mathcal{A}_j(m, l_{j-1}, l_j) := \big\{ (\mathbf{X}_{1j}(m), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)} \big\}.$$

Then,

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_m) &= \mathsf{P}(\cup_{l^b} \cap_{j=1}^b \mathcal{A}_j(m, l_{j-1}, l_j)) \\
&\leq \sum_{l^b} \mathsf{P}(\cap_{j=1}^b \mathcal{A}_j(m, l_{j-1}, l_j)) \\
&= \sum_{l^b} \prod_{j=1}^b \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) \\
&\leq \sum_{l^b} \prod_{j=2}^b \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)),
\end{aligned} \tag{14}$$

where equality (14) follows since the codebook is generated independently for each block j and the channel is memoryless. Note that if $m \neq 1$ and $l_{j-1} = 1$, then $\mathbf{X}_{1j}(m) \sim \prod_{i=1}^n p_{X_1}(x_{1,(j-1)n+i})$ is independent of $(\hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j})$ (given $L_{j-1} = L_j = 1$). Hence, by the joint typicality lemma [21],

$$\begin{aligned}
\mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) &= \mathsf{P}\big\{ (\mathbf{X}_{1j}(m), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)} \big\} \\
&\leq 2^{-n(I(X_1; \hat{Y}_2, Y_3 | X_2) - \delta(\epsilon))} \\
&=: 2^{-n(I_1 - \delta(\epsilon))}.
\end{aligned} \tag{15}$$

Similarly, if $m \neq 1$ and $l_{j-1} \neq 1$, then

$$(\mathbf{X}_{1j}(m), \mathbf{X}_{2j}(l_{j-1}), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1})) \sim \prod_{i=1}^n p_{X_1}(x_{1,(j-1)n+i}) p_{X_2, \hat{Y}_2}(x_{2,(j-1)n+i}, \hat{y}_{2,(j-1)n+i})$$

is independent of $\mathbf{Y}_{3j}$ (given $L_{j-1} = L_j = 1$). Hence, by the joint typicality lemma

$$\mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) \leq 2^{-n(I(X_1, X_2; Y_3) + I(\hat{Y}_2; X_1, Y_3 | X_2) - \delta(\epsilon))} =: 2^{-n(I_2 - \delta(\epsilon))}. \tag{16}$$

If the binary sequence l^{b-1} has k 1s, then by (15) and (16),

$$\prod_{j=2}^b \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) \leq 2^{-n(kI_1 + (b-1-k)I_2 - (b-1)\delta(\epsilon))}.$$

Therefore

$$\sum_{l^b} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) = \sum_{l_b} \sum_{l^{b-1}} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j))$$

$$\leq \sum_{l_b} \sum_{k=0}^{b-1} \binom{b-1}{k} 2^{n(b-1-k)\hat{R}_2} 2^{-n(kI_1 + (b-1-k)I_2 - (b-1)\delta(\epsilon))}$$

$$= \sum_{l_b} \sum_{k=0}^{b-1} \binom{b-1}{k} 2^{-n(kI_1 + (b-1-k)(I_2 - \hat{R}_2) - (b-1)\delta(\epsilon))}$$

$$\leq \sum_{l_b} \sum_{k=0}^{b-1} \binom{b-1}{k} 2^{-n((b-1)(\min\{I_1, I_2 - \hat{R}_2\} - \delta(\epsilon)))}$$

$$= 2^{n\hat{R}_2} 2^{b-1} \cdot 2^{-n((b-1)(\min\{I_1, I_2 - \hat{R}_2\} - \delta(\epsilon)))}.$$

Thus,

$$\sum_{m \neq 1} \sum_{l^b} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j) \mid L_{j-1} = L_j = 1) \to 0$$

as $n \to \infty$, provided that

$$R < \frac{b-1}{b}(\min\{I_1, I_2 - \hat{R}_2\} - \delta(\epsilon)) - \frac{1}{b}\hat{R}_2.$$

Finally, by eliminating $\hat{R}_2 > I(\hat{Y}_2; Y_2 | X_2) + \delta(\epsilon')$ and letting $b \to \infty$, we have shown the achievability of any rate

$$R < \min\{I(X_1; \hat{Y}_2, Y_3 | X_2), I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, Y_3)\} - \delta(\epsilon) - \delta(\epsilon').$$

This concludes the proof of Theorem 1 for the special case of the relay channel.

We now describe the noisy network coding scheme for multiple-source multicast over a general DMN $p(y^N | x^N)$. For simplicity of notation, we consider the case $Q = \emptyset$. Achievability for an arbitrary time-sharing random variable Q can be proved using the coded time sharing technique [21].

Codebook generation: Fix $\prod_{k=1}^{N} p(x_k)p(\hat{y}_k | y_k, x_k)$. We randomly and independently generate a codebook for each block. For each $j \in [1 : b]$ and $k \in [1 : N]$, randomly and independently generate $2^{nbR_k} \times 2^{n\hat{R}_k}$ sequences $\mathbf{x}_{k,j}(m_k, l_{k,j-1})$, $m_k \in [1 : 2^{nbR_k}]$, $l_{k,j-1} \in [1 : 2^{n\hat{R}_k}]$, each according to $\prod_{i=1}^{n} p_{X_k}(x_{k,(j-1)n+i})$. For each node $k \in [1 : N]$ and each $\mathbf{x}_{kj}(m_k, l_{k,j-1})$, $m_k \in [1 : 2^{nbR_k}]$, $l_{k,j-1} \in [1 : 2^{n\hat{R}_k}]$, randomly and conditionally independently generate $2^{n\hat{R}_k}$ sequences $\hat{\mathbf{y}}_{kj}(l_{kj} | m_k, l_{k,j-1})$, $l_{kj} \in [1 : 2^{n\hat{R}_k}]$, each according to $\prod_{i=1}^{n} p_{\hat{Y}_k | X_k}(\hat{y}_{k,(j-1)n+i} | x_{k,(j-1)n+i}(m_k, l_{k,j-1}))$. This defines the codebook

$$\mathcal{C}_j = \left\{ \mathbf{x}_{kj}(m_k, l_{k,j-1}), \hat{\mathbf{y}}_{kj}(l_{kj} | m_k, l_{k,j-1}) : m_k \in [1 : 2^{nbR_k}], l_{kj}, l_{k,j-1} \in [1 : 2^{n\hat{R}_k}], k \in [1 : N] \right\}$$

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

Encoding: Let $(m_1, \ldots, m_N)$ be the messages to be sent. Each node $k \in [1:N]$, upon receiving $\mathbf{y}_{kj}$ at the end of block $j \in [1:b]$, finds an index l_{kj} such that

$$(\hat{\mathbf{y}}_{kj}(l_{kj}|m_k, l_{k,j-1}), \mathbf{y}_{kj}, \mathbf{x}_{kj}(m_k, l_{k,j-1})) \in \mathcal{T}_{\epsilon'}^{(n)},$$

where $l_{k0} = 1$, $k \in [1:N]$, by convention. If there is more than one such index, choose one of them at random. If there is no such index, choose an arbitrary index at random from $[1:2^{n\hat{R}_k}]$. Then each node $k \in [1:N]$ transmits the codeword $\mathbf{x}_{kj}(m_k, l_{k,j-1})$ in block $j \in [1:b]$.

Decoding: Let $\epsilon > \epsilon'$. At the end of block b, decoder $d \in \mathcal{D}$ finds a unique index tuple $(\hat{m}_{1d}, \ldots, \hat{m}_{Nd})$, where $\hat{m}_{kd} \in [1:2^{nbR_k}]$ for $k \neq d$ and $\hat{m}_{dd} = m_d$, such that there exist some $(\hat{l}_{1j}, \ldots, \hat{l}_{Nj})$, $\hat{l}_{kj} \in [1:2^{n\hat{R}_k}]$, $k \neq d$ and $\hat{l}_{dj} = l_{dj}$, $j \in [1:b]$, satisfying

$$(\mathbf{x}_{1j}(\hat{m}_{1d}, \hat{l}_{1,j-1}), \ldots, \mathbf{x}_{Nj}(\hat{m}_{Nd}, \hat{l}_{N,j-1}),$$
$$\hat{\mathbf{y}}_{1j}(\hat{l}_{1j}|\hat{m}_{1d}, \hat{l}_{1,j-1}), \ldots, \hat{\mathbf{y}}_{Nj}(\hat{l}_{Nj}|\hat{m}_{Nd}, \hat{l}_{N,j-1}), \mathbf{y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)}$$

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

Analysis of the probability of error: Let M_k denote the message sent at node $k \in [1:N]$ and L_{kj}, $k \in [1:N]$, $j \in [1:b]$, denote the index chosen by node k for block j. To bound the probability of error for decoder $d \in \mathcal{D}$, assume without loss of generality that $(M_1, \ldots, M_N) = (1, \ldots, 1) =: \mathbf{1}$. Define

$$\mathcal{E}_0 := \bigcup_{j=1}^{b} \bigcup_{k=1}^{N} \left\{ (\hat{\mathbf{Y}}_{kj}(l_{kj}|1, L_{k,j-1}), \mathbf{X}_{kj}(1, L_{k,j-1}), \mathbf{Y}_{kj}) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l_{kj} \right\}$$

$$\mathcal{E}_{\mathbf{m}} := \left\{ (\mathbf{X}_{1j}(m_1, l_{1,j-1}), \ldots, \mathbf{X}_{Nj}(m_N, l_{N,j-1}), \right.$$
$$\hat{\mathbf{Y}}_{1j}(l_{1j}|m_1, l_{1,j-1}), \ldots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|m_N, l_{N,j-1}), \mathbf{Y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)},$$
$$\left. j \in [1:b], \text{ for some } (\mathbf{l}_1, \ldots, \mathbf{l}_b), \text{ where } l_{dj} = L_{dj}, j \in [1:b] \right\}.$$

Here, $\mathbf{l}_j = (l_{1j}, \ldots, l_{Nj})$ for $j \in [1:b]$. Then the probability of error is upper bounded as

$$\mathsf{P}(\mathcal{E}) \leq \mathsf{P}(\mathcal{E}_0) + \mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{\mathbf{1}}^c) + \mathsf{P}(\cup_{\mathbf{m} \neq \mathbf{1}} \mathcal{E}_{\mathbf{m}}), \tag{17}$$

where $\mathbf{m} := (m_1, \ldots, m_N)$ such that $m_d = 1$. As in the 3-node case, by the covering lemma, $\mathsf{P}(\mathcal{E}_0) \to 0$ as $n \to \infty$, if $\hat{R}_k > I(\hat{Y}_k; Y_k|X_k) + \delta(\epsilon')$, $k \in [1:N]$, and by the conditional typicality lemma $\mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{\mathbf{1}}^c) \to 0$ as $n \to \infty$. For the third term, assume without loss of generality that $\mathbf{L}_1 = \cdots = \mathbf{L}_b = \mathbf{1}$,

where $\mathbf{L}_j := (L_{1j}, \ldots, L_{Nj})$. Define the events

$$\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j) := \big\{ (\mathbf{X}_{1j}(m_1, l_{1,j-1}), \ldots, \mathbf{X}_{Nj}(m_N, l_{N,j-1}),$$
$$\hat{\mathbf{Y}}_{1j}(l_{1j}|m_1, l_{1,j-1}), \ldots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|m_N, l_{N,j-1}), \mathbf{Y}_{dj}) \in \mathcal{T}_\epsilon^{(n)} \big\}$$

for $\mathbf{m} \neq \mathbf{1}$ and all $\mathbf{l}_j$. Then,

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_{\mathbf{m}}) &= \mathsf{P}(\cup_{\mathbf{l}^b} \cap_{j=1}^b \mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\leq \sum_{\mathbf{l}^b} \mathsf{P}(\cap_{j=1}^b \mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&= \sum_{\mathbf{l}^b} \prod_{j=1}^b \mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\leq \sum_{\mathbf{l}^b} \prod_{j=2}^b \mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)),
\end{aligned} \tag{18}$$

where (18) follows since the codebook is generated independently for each block j and the channel is memoryless.

For each $\mathbf{l}^b$ and $j \in [2:b]$, let $\mathcal{S}_j(\mathbf{m}, \mathbf{l}^b) \subseteq [1:N]$ such that $\mathcal{S}_j(\mathbf{m}, \mathbf{l}^b) = \{k : m_k \neq 1 \text{ or } l_{k,j-1} \neq 1\}$. Note that $\mathcal{S}_j(\mathbf{m}, \mathbf{l}^b)$ depends only on $(\mathbf{m}, \mathbf{l}_{j-1})$ and hence we write it as $\mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1})$. We further define $\mathcal{T}(\mathbf{m}) \subseteq [1:N]$ such that $\mathcal{T}(\mathbf{m}) = \{k : m_k \neq 1\}$. From the definitions we can see that $\mathcal{T}(\mathbf{m}) \subseteq \mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1})$ and $d \in \mathcal{S}_j^c(\mathbf{m}, \mathbf{l}_{j-1}) \subseteq \mathcal{T}^c(\mathbf{m})$.

Define $\mathbf{X}_j(\mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1}))$ to be the set of $\mathbf{X}_{kj}(m_k, l_{k,j-1})$, $k \in \mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1})$, where m_k and $l_{k,j-1}$ are the corresponding elements in $\mathbf{m}$ and $\mathbf{l}^b$, respectively. Similarly define $\hat{\mathbf{Y}}_j(\mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1}))$ and $\mathbf{Y}_j(\mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1}))$. Then, by the joint typicality lemma and the fact that

$$\big(\mathbf{X}(\mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1})), \hat{\mathbf{Y}}(\mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1}))\big) \sim \prod_{k \in \mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1})} \prod_{i=1}^n p_{X_k}(x_{k,(j-1)n+i}) \, p_{\hat{Y}_k|X_k}(\hat{y}_{k,(j-1)n+i}|x_{k,(j-1)n+i})$$

is independent of $\big(\mathbf{X}(\mathcal{S}_j^c(\mathbf{m}, \mathbf{l}_{j-1})), \hat{\mathbf{Y}}(\mathcal{S}_j^c(\mathbf{m}, \mathbf{l}_{j-1})), \mathbf{Y}_{dj}\big)$ (given $\mathbf{L}_{j-1} = \mathbf{L}_j = \mathbf{1}$), we have

$$\mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \leq 2^{-n(I_1(\mathcal{S}(\mathbf{m}, \mathbf{l}_{j-1})) + I_2(\mathcal{S}(\mathbf{m}, \mathbf{l}_{j-1})) - \delta(\epsilon))},$$

where

$$I_1(\mathcal{S}) := I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c)),$$
$$I_2(\mathcal{S}) := \sum_{k \in \mathcal{S}} I(\hat{Y}_k; \hat{Y}(\mathcal{S}^c \cup \{k' \in \mathcal{S} : k' < k\}), Y_d, X^N | X_k).$$

Furthermore, from the definitions of $\mathcal{T}(\mathbf{m})$ and $\mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1})$, if $\mathbf{m} \neq \mathbf{1}$ with $m_d = 1$, then

$$\sum_{\mathbf{l}_{j-1}} 2^{-n(I_1(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) + I_2(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) - \delta(\epsilon))}$$

$$\leq \sum_{\substack{\mathcal{S} \subseteq [1:N] \\ \mathcal{T}(\mathbf{m}) \subseteq \mathcal{S}, d \in \mathcal{S}^c}} \sum_{\mathbf{l}_{j-1}: \mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1}) = \mathcal{S}} 2^{-n(I_1(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) + I_2(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) - \delta(\epsilon))}$$

$$\leq \sum_{\substack{\mathcal{S} \subseteq [1:N] \\ \mathcal{T}(\mathbf{m}) \subseteq \mathcal{S}, d \in \mathcal{S}^c}} 2^{-n(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon))}$$

$$\leq 2^{N-1} 2^{-n(\min_{\mathcal{S}}(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon)))},$$

where the minimum is over $\mathcal{S} \subseteq [1:N]$ such that $\mathcal{T}(\mathbf{m}) \subseteq \mathcal{S}$ and $d \in \mathcal{S}^c$. Hence,

$$\sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j))$$

$$= \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}_b} \sum_{\mathbf{l}^{b-1}} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j))$$

$$= \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}_b} \prod_{j=2}^{b} \sum_{\mathbf{l}_{j-1}} \mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j))$$

$$\leq \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}_b} \prod_{j=2}^{b} \left(\sum_{\mathbf{l}_{j-1}} 2^{-n(I_1(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) + I_2(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) - \delta(\epsilon))} \right)$$

$$\leq \sum_{\substack{\mathcal{T} \subseteq [1:N] \\ \mathcal{T} \neq \emptyset, d \in \mathcal{T}^c}} 2^{\sum_{k \in \mathcal{T}} nbR_k} 2^{\sum_{k \neq d} n\hat{R}_k} 2^{(N-1)(b-1)} 2^{n\left(-(b-1)\min_{\mathcal{S}}(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon))\right)}, \quad (19)$$

where the minimum in (19) is over $\mathcal{S} \subseteq [1:N]$ such that $\mathcal{T} \subseteq \mathcal{S}, d \in \mathcal{S}^c$. Hence, (19) tends to zero as $n \to \infty$ if

$$R(\mathcal{T}) < \frac{b-1}{b} \left(\min_{\substack{\mathcal{S} \subseteq [1:N] \\ \mathcal{T} \subseteq \mathcal{S}, d \in \mathcal{S}^c}} \left(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} \hat{R}_k \right) - \delta(\epsilon) \right) - \frac{1}{b} \left(\sum_{k \neq d} \hat{R}_k \right)$$

for all $\mathcal{T} \subseteq [1:N]$ such that $\mathcal{T} \neq \emptyset$ and $d \in \mathcal{T}^c$. By eliminating $\hat{R}_k > I(\hat{Y}_k; Y_k | X_k) + \delta(\epsilon')$ and letting $b \to \infty$, the probability of error tends to zero as $n \to \infty$ if

$$R(\mathcal{T}) < \min_{\substack{\mathcal{S} \subseteq [1:N] \\ \mathcal{T} \subseteq \mathcal{S}, d \in \mathcal{S}^c}} \left(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} I(\hat{Y}_k; Y_k | X_k) \right) - (N-1)\delta(\epsilon') - \delta(\epsilon)$$

for all $\mathcal{T} \subseteq [1 : N]$ such that $d \in \mathcal{T}^c$. Finally, note that

$$I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} I(\hat{Y}_k; Y_k | X_k) = -\sum_{k \in \mathcal{S}} I(\hat{Y}_k; Y_k | X^N, \hat{Y}(\mathcal{S}^c), Y_d, \hat{Y}(\{k' \in \mathcal{S} : k' < k\}))$$

$$= -\sum_{k \in \mathcal{S}} I(\hat{Y}_k; Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d, \hat{Y}(\{k' \in \mathcal{S} : k' < k\}))$$

$$= -I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d).$$

Therefore, the probability of error tends to zero as $n \to \infty$ if

$$R(\mathcal{T}) < I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c)) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d) - (N - 1)\delta(\epsilon') - \delta(\epsilon) \quad (20)$$

for all $\mathcal{S}, \mathcal{T} \subseteq [1 : N]$ such that $\emptyset \neq \mathcal{T} \subseteq \mathcal{S}$ and $d \in \mathcal{S}^c$. Since for every $\mathcal{S} \subseteq [1 : N]$ such that $\mathcal{S} \neq \emptyset$ and $d \in \mathcal{S}^c$ the inequalities with $\mathcal{T} \subsetneq \mathcal{S}$ are inactive due to the inequality with $\mathcal{T} = \mathcal{S}$ in (20), the set of inequalities can be further simplified to

$$R(\mathcal{S}) < I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c)) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d) - (N - 1)\delta(\epsilon') - \delta(\epsilon) \quad (21)$$

for all $\mathcal{S} \subseteq [1 : N]$ such that $d \in \mathcal{S}^c$. Thus, the probability of decoding error tends to zero for each destination node $d \in \mathcal{D}$ as $n \to \infty$, provided that the rate tuple satisfies (21).

Hence, by the union of events bound, the probability of error for *all* destinations tends to zero as $n \to \infty$ if the rate tuple $(R_1, \ldots, R_N)$ satisfies

$$R(\mathcal{S}) < \min_{d \in \mathcal{S}^c \cap \mathcal{D}} I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c)) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d)$$

for all $\mathcal{S} \subseteq [1 : N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$ for some $\prod_{k=1}^N p(x_k) p(\hat{y}_k | y_k, x_k)$. Finally, by coded time sharing, the probability of error tends to zero as $n \to \infty$ if the rate tuple $(R_1, \ldots, R_N)$ satisfies

$$R(\mathcal{S}) < \min_{d \in \mathcal{S}^c \cap \mathcal{D}} I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c), Q) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d, Q)$$

for all $\mathcal{S} \subseteq [1 : N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$ for some $\prod_{k=1}^N p(q) p(x_k | q) p(\hat{y}_k | y_k, x_k, q)$. This completes the proof of Theorem 1.

IV. EXTENSIONS TO GENERAL MULTIPLE-SOURCE NETWORKS

A. Proof of Theorem 2 via Multicast Completion with Implicit Decoding

We modify the decoding rule in the previous section to establish Theorem 2 as follows.

Decoding: At the end of block b, decoder $d \in \cup_{k=1}^N \mathcal{D}_k$ finds a unique index tuple $\{\hat{m}_{kd} : k \in \mathcal{S}_d\}$ such that there exist some $(\hat{m}_{kd} : k \in \mathcal{S}_d^c)$ and $(\hat{l}_{1j}, \ldots, \hat{l}_{Nj})$, where $\hat{m}_{kd} \in [1 : 2^{nbR_k}]$ for $k \neq d$, $\hat{m}_{dd} = m_d$,

$\hat{l}_{kj} \in [1 : 2^{n\hat{R}_k}]$ for $k \neq d$, and $\hat{l}_{dj} = l_{dj}$, $j \in [1 : b]$, satisfying

$$(\mathbf{x}_{1j}(\hat{m}_{1d}, \hat{l}_{1,j-1}), \dots, \mathbf{x}_{Nj}(\hat{m}_{Nd}, \hat{l}_{N,j-1}),$$

$$\hat{\mathbf{y}}_{1j}(\hat{l}_{1j}|\hat{m}_{1d}, \hat{l}_{1,j-1}), \dots, \hat{\mathbf{y}}_{Nj}(\hat{l}_{Nj}|\hat{m}_{Nd}, \hat{l}_{N,j-1}), \mathbf{y}_{dj}) \in \mathcal{T}_\epsilon^{(n)}$$

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

The analysis of the probability of error is similar to that for Theorem 1 in Section III. For completeness, the details are given in Appendix A.

B. Proof of Theorem 3 via Treating Interference as Noise

Codebook generation: Again we consider the case $Q = \emptyset$. Fix $\prod_{k=1}^N p(u_k, x_k) p(\hat{y}_k|y_k, u_k)$. We randomly and independently generate a codebook for each block. For each $j \in [1 : b]$ and $k \in [1 : N]$, randomly and independently generate $2^{n\hat{R}_k}$ sequences $\mathbf{u}_{kj}(l_{k,j-1})$, $l_{k,j-1} \in [1 : 2^{n\hat{R}_k}]$, each according to $\prod_{i=1}^n p_{U_k}(u_{k,(j-1)n+i})$. For each $k \in [1 : N]$ and each $\mathbf{u}_{kj}(l_{k,j-1})$, $l_{k,j-1} \in [1 : 2^{n\hat{R}_k}]$, randomly and conditionally independently generate 2^{nbR_k} sequences $\mathbf{x}_{kj}(m_k|l_{k,j-1})$, $m_k \in [1 : 2^{nbR_k}]$, each according to $\prod_{i=1}^n p_{X_k|U_k}(x_{k,(j-1)n+i}|u_{k,(j-1)n+i}(l_{k,j-1}))$. For each $k \in [1 : N]$ and each $\mathbf{u}_{kj}(l_{k,j-1})$, $l_{k,j-1} \in [1 : 2^{n\hat{R}_k}]$, randomly and conditionally independently generate $2^{n\hat{R}_k}$ sequences $\hat{\mathbf{y}}_{kj}(l_{kj}|l_{k,j-1})$, $l_{kj} \in [1 : 2^{n\hat{R}_k}]$, each according to $\prod_{i=1}^n p_{\hat{Y}_k|U_k}(\hat{y}_{k,(j-1)n+i}|u_{k,(j-1)n+i}(l_{k,j-1}))$. This defines the codebook

$$\mathcal{C}_j = \left\{ \mathbf{u}_{kj}(l_{k,j-1}), \mathbf{x}_{kj}(m_k|l_{k,j-1}), \hat{\mathbf{y}}_{kj}(l_{kj}|l_{k,j-1}) : m_k \in [1 : 2^{nbR_k}], l_{kj}, l_{k,j-1} \in [1 : 2^{n\hat{R}_k}], k \in [1 : N] \right\}$$

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

Encoding: Let $(m_1, \dots, m_N)$ be the messages to be sent. Each node $k \in [1 : N]$, upon receiving $\mathbf{y}_{kj}$ at the end of block $j \in [1 : b]$, finds an index l_{kj} such that

$$(\hat{\mathbf{y}}_{kj}(l_{kj}|l_{k,j-1}), \mathbf{y}_{kj}, \mathbf{u}_{kj}(l_{k,j-1})) \in \mathcal{T}_{\epsilon'}^{(n)},$$

where $l_{k0} = 1$, $k \in [1 : N]$, by convention. If there is more than one such index, choose one of them at random. If there is no such index, choose an arbitrary index at random from $[1 : 2^{n\hat{R}_k}]$. Then each node $k \in [1 : N]$ transmits the codeword $\mathbf{x}_{kj}(m_k|l_{k,j-1})$ in block $j \in [1 : b]$.

Similarly as before, decoding is done by simultaneous joint typical decoding, however, since we are treating interference as noise, codewords corresponding to the unintended messages $(m_k : k \in \mathcal{S}_d^c)$ are discarded, which leads to the following.

Decoding: At the end of block b, decoder $d \in \cup_{k=1}^{N} \mathcal{D}_k$ finds a unique index tuple $(\hat{m}_{kd} : k \in \mathcal{S}_d)$ such that there exist some $(\hat{l}_{1j}, \dots, \hat{l}_{Nj})$, where $\hat{m}_{kd} \in [1 : 2^{nbR_k}]$ and $k \neq d$ and $\hat{m}_{dd} = m_d$, $\hat{l}_{kj} \in [1 : 2^{n\hat{R}_k}]$, $k \neq d$ and $\hat{l}_{dj} = l_{dj}$, $j \in [1 : b]$, satisfying

$$((\mathbf{x}_{kj}(\hat{m}_{kd}|\hat{l}_{k,j-1}) : k \in \mathcal{S}_d), \mathbf{u}_{1j}(\hat{l}_{1,j-1}), \dots, \mathbf{u}_{Nj}(\hat{l}_{N,j-1}),$$

$$\hat{\mathbf{y}}_{1j}(\hat{l}_{1j}|\hat{l}_{1,j-1}), \dots, \hat{\mathbf{y}}_{Nj}(\hat{l}_{Nj}|\hat{l}_{N,j-1}), \mathbf{y}_{dj}) \in \mathcal{T}_\epsilon^{(n)}$$

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

The analysis of the probability of error is delegated to Appendix B.

V. GAUSSIAN NETWORKS

We consider the additive white Gaussian noise (AWGN) network in which the channel output vector for an input vector X^N is $Y^N = GX^N + Z^N$, where $G \in \mathbb{R}^{N \times N}$ is the channel gain matrix and Z^N is a vector of independent additive white Gaussian noise with zero mean and unit variance. We assume average power constraint P on each sender, i.e.,

$$\sum_{i=1}^{n} \mathsf{E}\left(x_{ki}^2(m_k, Y_k^{i-1})\right) \leq nP$$

for all $k \in [1 : N]$ and $m_k \in [1 : 2^{nR_k}]$. For each cutset $\mathcal{S} \subseteq [1 : N]$, define a channel gain submatrix $G(\mathcal{S})$ such that

$$Y(\mathcal{S}^c) = G(\mathcal{S})X(\mathcal{S}) + G'(\mathcal{S})X(\mathcal{S}^c) + Z(\mathcal{S}^c).$$

In the following subsection, we prove Theorem 4. In Subsections V-B and V-C, we provide the capacity inner bounds for the AWGN two-way relay channel and the AWGN interference relay channel used in Figures 2 and 4.

A. AWGN Multicast Capacity Gap (Proof of Theorem 4)

The cutset outer bound for the AWGN multiple-source multicast network simplifies to the set of rate tuples such that

$$\sum_{k \in \mathcal{S}} R_k \leq \frac{1}{2} \log \left| I + \frac{P}{2} G(\mathcal{S}) G(\mathcal{S})^T \right| + \frac{1}{2} \min\{|\mathcal{S}|, |\mathcal{S}^c|\} \log(2|\mathcal{S}|) \tag{22}$$

for all $\mathcal{S} \subseteq [1 : N]$ with $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$. To show this, first note that the cutset outer bound (4) continues to hold with the set of input distributions satisfying $\mathsf{E}(X_k^2) \leq P$, $k \in [1 : N]$. For each $\mathcal{S} \subseteq [1 : N]$ such

that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$, we can further loosen the cutset outer bound as

$$
\begin{aligned}
R(\mathcal{S}) &\leq I(X(\mathcal{S}); Y(\mathcal{S}^c)|X(\mathcal{S}^c)) \\
&= h(Y(\mathcal{S}^c)|X(\mathcal{S}^c)) - h(Y(\mathcal{S}^c)|X^N) \\
&= h\big(G(\mathcal{S})X(\mathcal{S}) + Z(\mathcal{S}^c)|X(\mathcal{S}^c)\big) - h(Y(\mathcal{S}^c)|X^N) \\
&= h\big(G(\mathcal{S})X(\mathcal{S}) + Z(\mathcal{S}^c)\big) - h(Y(\mathcal{S}^c)|X^N) \\
&= \frac{1}{2}\log(2\pi e)^{|\mathcal{S}^c|}\Big|I + G(\mathcal{S})K_{X(\mathcal{S})}G(\mathcal{S})^T\Big| - \frac{|\mathcal{S}^c|}{2}\log(2\pi e) \\
&\leq \frac{1}{2}\log\Big|I + \mathrm{tr}(K_{X(\mathcal{S})})G(\mathcal{S})G(\mathcal{S})^T\Big| && (23) \\
&\leq \frac{1}{2}\log\Big|I + |\mathcal{S}|P \cdot G(\mathcal{S})G(\mathcal{S})^T\Big| && (24) \\
&\leq \frac{1}{2}\log\Big|2|\mathcal{S}| \cdot I + 2|\mathcal{S}|\frac{P}{2} \cdot G(\mathcal{S})G(\mathcal{S})^T\Big| \\
&\leq \frac{1}{2}\log\Big|I + \frac{P}{2}G(\mathcal{S})G(\mathcal{S})^T\Big| + \frac{|\mathcal{S}^c|}{2}\log(2|\mathcal{S}|),
\end{aligned}
$$

where $K_{X(\mathcal{S})}$ is the covariance matrix of $X(\mathcal{S})$, (23) follows since $\mathrm{tr}(K)I - K$ is positive semidefinite for any covariance matrix K [23, Theorem 7.7.3], and (24) follows since $\mathrm{tr}(K_{X(\mathcal{S})}) \leq |\mathcal{S}|P$, from the power constraint. By rewriting (24) as

$$
\frac{1}{2}\log\Big|I + |\mathcal{S}|P \cdot G(\mathcal{S})G(\mathcal{S})^T\Big| = \frac{1}{2}\log\Big|I + |\mathcal{S}|P \cdot G(\mathcal{S})^T G(\mathcal{S})\Big|
$$

and following similar steps, we also have

$$
\begin{aligned}
R(\mathcal{S}) &\leq \frac{1}{2}\log\left|I + \frac{P}{2}G(\mathcal{S})^T G(\mathcal{S})\right| + \frac{|\mathcal{S}|}{2}\log(2|\mathcal{S}|) \\
&= \frac{1}{2}\log\left|I + \frac{P}{2}G(\mathcal{S})G(\mathcal{S})^T\right| + \frac{|\mathcal{S}|}{2}\log(2|\mathcal{S}|).
\end{aligned}
$$

On the other hand, the noisy network coding inner bound in Theorem 1 yields the inner bound characterized by the set of inequalities

$$
R(\mathcal{S}) \leq \frac{1}{2}\log\left|I + \frac{P}{2}G(\mathcal{S})G(\mathcal{S})^T\right| - \frac{|\mathcal{S}|}{2} \tag{25}
$$

for all $\mathcal{S} \subseteq [1 : N]$ with $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$. To show this, first note that by the standard procedure [21], Theorem 1 for the discrete memoryless network can be easily adapted for the AWGN network with power constraint, which gives the inner bound (3) on the capacity region with (product) input distributions satisfying $\mathsf{E}(X_k^2) \leq P$, $k \in [1 : N]$.

Let $Q = \emptyset$ and X_k, $k \in [1 : N]$, be i.i.d. Gaussian with zero mean and variance P. Let

$$
\hat{Y}_k = Y_k + \hat{Z}_k, \quad k \in [1 : N],
$$

where $\hat{Z}_k$, $k \in [1:N]$, are i.i.d. Gaussian with zero mean and unit variance. Then for each $\mathcal{S} \subseteq [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$ and $d \in \mathcal{D}$,

$$
\begin{aligned}
I(\hat{Y}(\mathcal{S}); Y(\mathcal{S})|X^N, \hat{Y}(\mathcal{S}^c), Y_d) &\leq I(\hat{Y}(\mathcal{S}); Y(\mathcal{S})|X^N) \\
&= h(\hat{Y}(\mathcal{S})|X^N) - h(\hat{Y}(\mathcal{S})|Y(\mathcal{S}), X^N) \\
&= \frac{|\mathcal{S}|}{2} \log(4\pi e) - \frac{|\mathcal{S}|}{2} \log(2\pi e) \\
&= \frac{|\mathcal{S}|}{2},
\end{aligned}
$$

where the first inequality is due to the Markovity $(\hat{Y}(\mathcal{S}^c), Y_d) \to (X^N, Y(\mathcal{S})) \to \hat{Y}(\mathcal{S})$. Furthermore,

$$
\begin{aligned}
I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d|X(\mathcal{S}^c)) &\geq I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c)|X(\mathcal{S}^c)) \\
&= h(\hat{Y}(\mathcal{S}^c)|X(\mathcal{S}^c)) - h(\hat{Y}(\mathcal{S}^c)|X^N) \\
&= \frac{1}{2}\log(2\pi e)^{|\mathcal{S}^c|} \left|2I + G(\mathcal{S})PG(\mathcal{S})^T\right| - \frac{|\mathcal{S}^c|}{2}\log(4\pi e) \\
&= \frac{1}{2}\log\left|I + \frac{P}{2}G(\mathcal{S})G(\mathcal{S})^T\right|.
\end{aligned}
$$

Therefore, by Theorem 1, a rate tuple $(R_1, \ldots, R_N)$ is achievable if

$$
R(\mathcal{S}) < \frac{1}{2}\log\left|I + \frac{P}{2}G(\mathcal{S})G(\mathcal{S})^T\right| - \frac{|\mathcal{S}|}{2}
$$

for all $\mathcal{S} \subseteq [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$.

Comparing the cutset outer bound (22) and inner bound (25) completes the proof of Theorem 4.

B. AWGN Two-Way Relay Channels

Recall the model for the AWGN two-way relay channel (13) in Section II.

Rankov and Wittenben [16] showed that the amplify–forward (AF) coding scheme results in the inner bound on the capacity region that consists of all rate pairs (R_1, R_2) such that

$$
R_k < \frac{1}{2}\log\left(\frac{a_k + \sqrt{a_k^2 - b_k^2}}{2}\right), \quad k \in \{1, 2\}
$$

for some $\alpha \leq \sqrt{P/(g_{13}^2 P + g_{23}^2 P + 1)}$, where $a_1 := 1 + \frac{P(g_{12}^2 + \alpha^2 g_{32}^2 g_{13}^2)}{g_{32}^2 \alpha^2 + 1}$, $a_2 := 1 + \frac{P(g_{21}^2 + \alpha^2 g_{31}^2 g_{23}^2)}{g_{31}^2 \alpha^2 + 1}$, $b_1 := \frac{2P\alpha g_{32} g_{13} g_{12}}{g_{32}^2 \alpha^2 + 1}$, and $b_2 := \frac{2P\alpha g_{31} g_{23} g_{21}}{g_{31}^2 \alpha^2 + 1}$. They also showed that an extension of the original compress–forward (CF) coding scheme for the relay channel to the two-way relay channel results in the following

inner bound on the capacity region that consists of all rate pairs (R_1, R_2) such that

$$R_1 < \mathrm{C}\left(\frac{g_{13}^2 P + (1+\sigma^2)g_{12}^2 P}{1+\sigma^2}\right),$$

$$R_2 < \mathrm{C}\left(\frac{g_{23}^2 P + (1+\sigma^2)g_{21}^2 P}{1+\sigma^2}\right)$$

for some

$$\sigma^2 \geq \max\left\{\frac{(1+g_{12}^2 P)(1+g_{13}^2 P) - (g_{12}g_{13}P)^2}{\min\{g_{32}^2, g_{31}^2\}P}, \frac{(1+g_{21}^2 P)(1+g_{23}^2 P) - (g_{21}g_{23}P)^2}{\min\{g_{32}^2, g_{31}^2\}P}\right\}.$$

Specializing Theorem 2 to the two-way relay channel gives the inner bound that consists of all rate pairs (R_1, R_2) such that

$$R_1 \leq \min\{I(X_1; Y_2, \hat{Y}_3 | X_2, X_3, Q), I(X_1, X_3; Y_2 | X_2, Q) - I(Y_3; \hat{Y}_3 | X_1, X_2, X_3, Y_2, Q)\}$$

$$R_2 \leq \min\{I(X_2; Y_1, \hat{Y}_3 | X_1, X_3, Q), I(X_2, X_3; Y_1 | X_1, Q) - I(Y_3; \hat{Y}_3 | X_1, X_2, X_3, Y_1, Q)\}$$

for some $p(q)p(x_1|q)p(x_2|q)p(x_3|q)p(\hat{y}_3|y_3, x_3, q)$. By setting $Q = \emptyset$ and $\hat{Y}_3 = Y_3 + \hat{Z}$ with $\hat{Z} \sim \mathrm{N}(0, \sigma^2)$, this inner bound simplifies to the set of rate pairs (R_1, R_2) such that

$$R_1 < \min\left\{\mathrm{C}\left(\frac{g_{13}^2 P + (1+\sigma^2)g_{12}^2 P}{1+\sigma^2}\right), \mathrm{C}(g_{12}^2 P + g_{32}^2 P) - \mathrm{C}(1/\sigma^2)\right\},$$

$$R_2 < \min\left\{\mathrm{C}\left(\frac{g_{23}^2 P + (1+\sigma^2)g_{21}^2 P}{1+\sigma^2}\right), \mathrm{C}(g_{21}^2 P_1 + g_{31}^2 P) - \mathrm{C}(1/\sigma^2)\right\} \qquad (26)$$

for some $\sigma^2 > 0$.

C. AWGN Interference Relay Channels

Recall the model for the AWGN interference relay channel with orthogonal receiver components in Figure 4.

Djeumou, Belmaga, and Lasaulce [18], and Razaghi and Yu [19] showed that an extension of the original compress–forward (CF) coding scheme for the relay channel to the two-way relay channel results in the inner bound on the capacity region that consists of all rate pairs (R_1, R_2) such that

$$R_1 < \mathrm{C}\left(\frac{(g_{13}^2 + (1+\sigma^2)g_{14}^2)P + (g_{23}g_{14} - g_{24}g_{13})^2 P^2}{1+\sigma^2 + (g_{23}^2 + (1+\sigma^2)g_{24}^2)P}\right),$$

$$R_2 < \mathrm{C}\left(\frac{(g_{23}^2 + (1+\sigma^2)g_{25}^2)P + (g_{13}g_{25} - g_{15}g_{23})^2 P^2}{1+\sigma^2 + (g_{13}^2 + (1+\sigma^2)g_{15}^2)P}\right)$$

for some

$$\sigma^2 \geq \frac{1}{2^{2R_0} - 1} \cdot \max\left\{\frac{(g_{13}g_{24} - g_{23}g_{14})^2 P^2 + a_1}{(g_{14}^2 P + g_{24}^2 P + 1)}, \frac{(g_{13}g_{25} - g_{23}g_{15})^2 P^2 + a_2}{(g_{15}^2 P + g_{25}^2 P + 1)}\right\},$$

where

$$a_1 := (g_{13}^2 + g_{14}^2)P + (g_{23}^2 + g_{24}^2)P + 1,$$

$$a_2 := (g_{13}^2 + g_{15}^2)P + (g_{23}^2 + g_{25}^2)P + 1.$$

Razaghi and Yu [19] generalized the hash–forward coding scheme [24], [11] for the relay channel to the interference relay channel, in which the relay sends the bin index (hash) of its noisy observation and destination nodes use list decoding. This generalized hash–forward scheme gives the inner bound on the capacity region that consists of the set of rate pairs (R_1, R_2) such that

$$R_1 < C\left(\frac{g_{14}^2 P}{g_{24}^2 P + 1}\right) + R_0 - C\left(\frac{(g_{23}^2 + g_{24}^2)P + 1}{(g_{24}^2 P + 1)\sigma^2}\right),$$

$$R_2 < C\left(\frac{g_{25}^2 P}{g_{15}^2 P + 1}\right) + R_0 - C\left(\frac{(g_{13}^2 + g_{15}^2)P + 1}{(g_{15}^2 P + 1)\sigma^2}\right)$$

for some $\sigma^2 > 0$ satisfying

$$\sigma^2 \leq \frac{1}{2^{2R_0} - 1} \cdot \min\left\{\frac{(g_{13}g_{24} - g_{23}g_{14})^2 P^2 + a_1}{(g_{14}^2 P + g_{24}^2 P + 1)}, \frac{(g_{13}g_{25} - g_{23}g_{15})^2 P^2 + a_2}{(g_{15}^2 P + g_{25}^2 P + 1)}\right\},$$

where a_1 and a_2 are the same as above.

Specializing Theorem 2 by setting $\hat{Y}_3 = Y_3 + \hat{Z}$ with $\hat{Z} \sim N(0, \sigma^2)$ gives the inner bound that consists of all rate pairs (R_1, R_2) such that

$$R_1 < \min\left\{C(g_{14}^2 P) + R_0 - C(1/\sigma^2), \ C\left(\frac{(g_{13}^2 + (1 + \sigma^2)g_{14}^2)P}{1 + \sigma^2}\right)\right\},$$

$$R_2 < \min\left\{C(g_{25}^2 P) + R_0 - C(1/\sigma^2), \ C\left(\frac{(g_{23}^2 + (1 + \sigma^2)g_{25}^2)P}{1 + \sigma^2}\right)\right\},$$

$$R_1 + R_2 < C((g_{14}^2 + g_{24}^2)P) + R_0 - C(1/\sigma^2),$$

$$R_1 + R_2 < C\left(\frac{(g_{13}^2 + g_{23}^2)P + (1 + \sigma^2)(g_{14}^2 + g_{24}^2)P + (g_{13}g_{24} - g_{23}g_{14})^2 P^2}{1 + \sigma^2}\right),$$

$$R_1 + R_2 < C((g_{15}^2 + g_{25}^2)P) + R_0 - C(1/\sigma^2),$$

$$R_1 + R_2 < C\left(\frac{(g_{13}^2 + g_{23}^2)P + (1 + \sigma^2)(g_{25}^2 + g_{15}^2)P + (g_{23}g_{15} - g_{13}g_{25})^2 P^2}{1 + \sigma^2}\right)$$

for some $\sigma^2 > 0$. By the same choice of $\hat{Y}_3$, the inner bound in Theorem 3 can be specialized to the set of rate pairs (R_1, R_2) such that

$$R_1 < C\left(\frac{g_{14}^2 P}{g_{24}^2 P + 1}\right) + R_0 - C\left(\frac{(g_{23}^2 + g_{24}^2)P + 1}{(g_{24}^2 P + 1)\sigma^2}\right),$$

$$R_1 < C\left(\frac{(g_{13}^2 + (1 + \sigma^2)g_{14}^2)P + (g_{23}g_{14} - g_{24}g_{13})^2 P^2}{1 + \sigma^2 + (g_{23}^2 + (1 + \sigma^2)g_{24}^2)P}\right),$$

$$R_2 < C\left(\frac{g_{25}^2 P}{g_{15}^2 P + 1}\right) + R_0 - C\left(\frac{(g_{13}^2 + g_{15}^2)P + 1}{(g_{15}^2 P + 1)\sigma^2}\right),$$

$$R_2 < \mathrm{C}\left(\frac{(g_{23}^2 + (1+\sigma^2)g_{25}^2)P + (g_{13}g_{25} - g_{15}g_{23})^2 P^2}{1 + \sigma^2 + (g_{13}^2 + (1+\sigma^2)g_{15}^2)P}\right)$$

for some $\sigma^2 > 0$.

VI. Concluding Remarks

We presented a new noisy network coding scheme and used it to establish inner bounds on the capacity region of general multiple-source noisy networks. This scheme unifies and extends previous results on network coding and its extensions, and on compress–forward for the relay channel. We demonstrated that the noisy network coding scheme can outperform previous network compress–forward schemes. The reasons are: first, the relays do not use Wyner–Ziv coding (no binning index to decode), second, the destinations are not required to decode the compression indices correctly, and third, simultaneous decoding over all blocks is used.

How good is noisy network coding as a general purpose scheme? As we have seen, noisy network coding is optimal in some special cases. It also performs generally well under high SNR conditions in the network. In addition, it is a robust and scalable scheme in the sense that the relay operations do not depend on the specific codebooks used by the sources and destinations or even the topology of the network. Noisy network coding, however, is not always the best strategy. For example, for a cascade of AWGN channels with low SNR, the optimal strategy is for the relay to decode the message and then forward it to the final destination. This simple multi-hop scheme can be improved by using the information from multiple paths and coherent cooperation as in the decode–forward scheme for the relay channel [10] and its extensions to networks [25], [14]. Further improvement can be obtained by only partial decoding of messages at the relays [10], and by combining noisy network coding with partial decode–forward to obtain the type of hybrid schemes in [10] and [14].

Another important direction to improve noisy network coding for multiple sources is to use more sophisticated interference coding schemes, such as interference alignment [26] and Han–Kobayashi superposition coding [27].

APPENDIX A

ERROR PROBABILITY ANALYSIS FOR THEOREM 2

The analysis follows the same steps of the multicast case except that the union in the third error term of (17) is over all $\mathbf{m}$ such that $(m_k : k \in \mathcal{S}_d) \neq (1, \ldots, 1)$. Thus,

$$
\begin{aligned}
&\mathsf{P}(\cup_{\mathbf{m}} \mathcal{E}_{\mathbf{m}}) \\
&\leq \sum_{\mathbf{m}} \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\leq \sum_{\substack{\mathcal{T} \subseteq [1:N] \\ \mathcal{T} \cap \mathcal{S}_d \neq \emptyset, d \in \mathcal{T}^c}} 2^{\sum_{k \in \mathcal{T}} nbR_k} 2^{\sum_{k \neq d} n\hat{R}_k} 2^{(N-1)(b-1)} 2^{n\left(-(b-1)\min_{\mathcal{S}}(I_1(\mathcal{S})+I_2(\mathcal{S})-\sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon))\right)},
\end{aligned}
$$

(27)

where the minimum in (27) is over $\mathcal{S} \subseteq [1 : N]$ such that $\mathcal{T} \subseteq \mathcal{S}, d \in \mathcal{S}^c$. Hence, (27) tends to zero as $n \to \infty$ if

$$
R(\mathcal{T}) < \frac{b-1}{b} \left(\min_{\substack{\mathcal{S} \subseteq [1:N] \\ \mathcal{T} \subseteq \mathcal{S}, d \in \mathcal{S}^c}} \left(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} \hat{R}_k \right) - \delta(\epsilon) \right) - \frac{1}{b} \left(\sum_{k \neq d} \hat{R}_k \right)
$$

for all $\mathcal{T} \subseteq [1 : N]$ such that $\mathcal{T} \cap \mathcal{S}_d \neq \emptyset$ and $d \in \mathcal{T}^c$. By eliminating $\hat{R}_k > I(\hat{Y}_k; Y_k | X_k) + \delta(\epsilon')$, letting $b \to \infty$, and getting rid of inactive inequalities, the probability of error tends to zero as $n \to \infty$ if

$$
R(\mathcal{S}) < I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c)) - I(\hat{Y}(\mathcal{S}); Y(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d) - (N-1)\delta(\epsilon') - \delta(\epsilon),
$$

(28)

for all $\mathcal{S} \subseteq [1 : N]$ such that $\mathcal{S} \cap \mathcal{S}_d \neq \emptyset$ and $d \in \mathcal{S}^c$. Thus, the probability of decoding error tends to zero for each destination node $d \in \mathcal{D}$ as $n \to \infty$, provided that the rate tuple satisfies (28). Finally, by the union of events bound, the probability of error for all destinations tends to zero as $n \to \infty$ if the rate tuple $(R_1, \ldots, R_N)$ satisfies

$$
R(\mathcal{S}) < \min_{d \in \mathcal{S}^c \cap \mathcal{D}(\mathcal{S})} I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c)) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d)
$$

for all $\mathcal{S} \subseteq [1 : N]$ such that $\mathcal{S}^c \cap \mathcal{D}(\mathcal{S}) \neq \emptyset$ for some $\prod_{k=1}^{N} p(x_k) p(\hat{y}_k | y_k, x_k)$. This completes the proof of Theorem 2 for $Q = \emptyset$. The proof for the general Q follows by coded time sharing.

APPENDIX B

ERROR PROBABILITY ANALYSIS FOR THEOREM 3

Let M_k denote the message sent at node $k \in [1 : N]$ and L_{kj}, $k \in [1 : N]$, $j \in [1 : b]$, denote the index chosen by node k for block j. To bound the probability of error for decoder $d \in \mathcal{D}$, assume without loss

of generality that $(M_1, \ldots, M_N) = (1, \ldots, 1) = \mathbf{1}$. Define

$$\mathcal{E}_0 := \bigcup_{j=1}^{b} \bigcup_{k=1}^{N} \left\{ (\hat{\mathbf{Y}}_{kj}(l_{kj}|L_{k,j-1}), \mathbf{U}_{kj}(L_{k,j-1}), \mathbf{Y}_{kj}) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l_{kj} \right\}$$

$$\mathcal{E}_{\mathbf{m}} := \big\{ (\{\mathbf{X}_{kj}(m_k|l_{k,j-1}) : k \in \mathcal{S}_d\}, \mathbf{U}_{1j}(l_{1,j-1}), \ldots, \mathbf{U}_{Nj}(l_{N,j-1}),$$

$$\hat{\mathbf{Y}}_{1j}(l_{1j}|l_{1,j-1}), \ldots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|l_{N,j-1}), \mathbf{Y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)}, j \in [1:b],$$

$$\text{for some } (\mathbf{l}_1, \ldots, \mathbf{l}_b), \text{ where } l_{dj} = L_{dj}, j \in [1:b] \big\}.$$

Here, $\mathbf{m} := (m_k : k \in \mathcal{S}_d)$ and $\mathbf{l}_j = (l_{1j}, \ldots, l_{Nj})$ for $j \in [1:b]$. Then the probability of error is upper bounded as $\mathsf{P}(\mathcal{E}) \leq \mathsf{P}(\mathcal{E}_0) + \mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{\mathbf{1}}^c) + \mathsf{P}(\cup_{\mathbf{m} \neq \mathbf{1}} \mathcal{E}_{\mathbf{m}})$, where $m_d = 1$ in $\mathbf{m}$. By the covering lemma, $\mathsf{P}(\mathcal{E}_0) \to 0$ as $n \to \infty$, if $\hat{R}_k > I(\hat{Y}_k; Y_k|U_k) + \delta(\epsilon')$, $k \in [1:N]$, and by the conditional typicality lemma $\mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_{\mathbf{1}}^c) \to 0$ as $n \to \infty$. For the third term, assume that $\mathbf{L}_1 = \cdots = \mathbf{L}_b = \mathbf{1}$. Define

$$\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j) := \{ (\{\mathbf{X}_{kj}(m_k|l_{k,j-1}) : k \in \mathcal{S}_d\}, \mathbf{U}_{1j}(l_{1,j-1}), \ldots, \mathbf{U}_{Nj}(l_{N,j-1}),$$

$$\hat{\mathbf{Y}}_{1j}(l_{1j}|l_{1,j-1}), \ldots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|l_{N,j-1}), \mathbf{Y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)} \}$$

for $\mathbf{m} \neq \mathbf{1}$ and all $\mathbf{l}_j$. Then, from similar steps to the multicast case,

$$\mathsf{P}(\mathcal{E}_{\mathbf{m}}) \leq \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)).$$

For each $\mathbf{l}^b$ and $j \in [2:b]$, let $\mathcal{S}_j(\mathbf{l}^b) \subseteq [1:N]$ such that $\mathcal{S}_j(\mathbf{l}^b) = \{k : l_{k,j-1} \neq 1\}$. We further define $\mathcal{T}(\mathbf{m}) \subseteq [1:N]$ such that $\mathcal{T}(\mathbf{m}) = \{k : k \in \mathcal{S}_d, m_k \neq 1\}$. By definition, $d \in \mathcal{T}^c(\mathbf{m}) \cap \mathcal{S}_j^c(\mathbf{l}_{j-1})$, where $\mathcal{T}^c(\mathbf{m}) := \mathcal{S}_d \backslash \mathcal{T}(\mathbf{m})$. Then, by the joint typicality lemma, we can show that

$$\mathsf{P}(\mathcal{A}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \leq 2^{-n(I_1(\mathcal{S}(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) + I_2(\mathcal{S}(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) - \delta(\epsilon))},$$

where

$$I_1(\mathcal{S}, \mathcal{T}) := I(X((\mathcal{S} \cup \mathcal{T}) \cap \mathcal{S}_d), U(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X((\mathcal{S}^c \cap \mathcal{T}^c) \cap \mathcal{S}_d), U(\mathcal{S}^c)), \text{ and}$$

$$I_2(\mathcal{S}, \mathcal{T}) := \sum_{k \in \mathcal{S}} I(\hat{Y}_k; \hat{Y}(\mathcal{S}^c \cup \{k' \in \mathcal{S} : k' < k\}), Y_d, X(\mathcal{S}_d), U^N | U_k).$$

Furthermore from the definitions of $\mathcal{T}(\mathbf{m})$ and $\mathcal{S}_j(\mathbf{l}_{j-1})$, if $\mathbf{m} \neq \mathbf{1}$ with $m_d = 1$, then

$$\sum_{\mathbf{l}_{j-1}} 2^{-n(I_1(\mathcal{S}_j(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) + I_2(\mathcal{S}_j(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) - \delta(\epsilon))}$$

$$\leq \sum_{\mathcal{S} \subseteq [1:N] : d \in \mathcal{S}^c} 2^{-n(I_1(\mathcal{S}, \mathcal{T}(\mathbf{m})) + I_2(\mathcal{S}, \mathcal{T}(\mathbf{m})) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon))}$$

$$\leq 2^{N-1} 2^{-n(\min_{\mathcal{S}}(I_1(\mathcal{S}, \mathcal{T}(\mathbf{m})) + I_2(\mathcal{S}, \mathcal{T}(\mathbf{m})) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon)))}.$$

Hence,

$$\sum_{\mathbf{m}\neq\mathbf{1}}\sum_{\mathbf{l}^b}\prod_{j=2}^{b}\mathsf{P}(\mathcal{A}_j(\mathbf{m},\mathbf{l}_{j-1},\mathbf{l}_j))$$

$$=\sum_{\mathbf{m}\neq\mathbf{1}}\sum_{\mathbf{l}_b}\sum_{\mathbf{l}^{b-1}}\prod_{j=2}^{b}\mathsf{P}(\mathcal{A}_j(\mathbf{m},\mathbf{l}_{j-1},\mathbf{l}_j))$$

$$=\sum_{\mathbf{m}\neq\mathbf{1}}\sum_{\mathbf{l}_b}\prod_{j=2}^{b}\sum_{\mathbf{l}_{j-1}}\mathsf{P}(\mathcal{A}_j(\mathbf{m},\mathbf{l}_{j-1},\mathbf{l}_j))$$

$$\leq\sum_{\mathbf{m}\neq\mathbf{1}}\sum_{\mathbf{l}_b}\prod_{j=2}^{b}\left(\sum_{\mathbf{l}_{j-1}}2^{-n(I_1(\mathcal{S}_j(\mathbf{l}_{j-1}),\mathcal{T}(\mathbf{m}))+I_2(\mathcal{S}_j(\mathbf{l}_{j-1}),\mathcal{T}(\mathbf{m}))-\delta(\epsilon))}\right)$$

$$\leq\sum_{\mathbf{m}\neq\mathbf{1}}\sum_{\mathbf{l}_b}\prod_{j=2}^{b}\left(2^{N-1}2^{-n(\min_{\mathcal{S}}(I_1(\mathcal{S},\mathcal{T}(\mathbf{m}))+I_2(\mathcal{S},\mathcal{T}(\mathbf{m}))-\sum_{k\in\mathcal{S}}\hat{R}_k-\delta(\epsilon)))}\right)$$

$$\leq\sum_{\substack{\mathcal{T}\subseteq\mathcal{S}_d\\\mathcal{T}\neq\emptyset,d\in\mathcal{T}^c}}2^{\sum_{k\in\mathcal{T}}nbR_k}\,2^{\sum_{k\neq d}n\hat{R}_k}2^{(N-1)(b-1)}\,2^{n\left(-(b-1)\min_{\mathcal{S}}(I_1(\mathcal{S},\mathcal{T})+I_2(\mathcal{S},\mathcal{T})-\sum_{k\in\mathcal{S}}\hat{R}_k-\delta(\epsilon))\right)},\quad(29)$$

where the minimum in (29) is over $\mathcal{S}\subseteq[1:N]$ such that $d\in\mathcal{S}^c$. Hence, (29) tends to zero as $n\to\infty$ if

$$R(\mathcal{T})<\frac{b-1}{b}\left(\min_{\mathcal{S}\subseteq[1:N],d\in\mathcal{S}^c}\left(I_1(\mathcal{S},\mathcal{T})+I_2(\mathcal{S},\mathcal{T})-\sum_{k\in\mathcal{S}}\hat{R}_k\right)-\delta(\epsilon)\right)-\frac{1}{b}\left(\sum_{k\neq d}\hat{R}_k\right)$$

for all $\mathcal{T}\subseteq\mathcal{S}_d$ such that $d\in\mathcal{T}^c$. By eliminating $\hat{R}_k>I(\hat{Y}_k;Y_k|U_k)+\delta(\epsilon')$ and letting $b\to\infty$, the probability of error tends to zero as $n\to\infty$ if

$$R(\mathcal{T})<\min_{\mathcal{S}\subseteq[1:N],d\in\mathcal{S}^c}\left(I_1(\mathcal{S},\mathcal{T})+I_2(\mathcal{S},\mathcal{T})-\sum_{k\in\mathcal{S}}I(\hat{Y}_k;Y_k|U_k)\right)-(N-1)\delta(\epsilon')-\delta(\epsilon)$$

for all $\mathcal{T}\subseteq\mathcal{S}_d$ such that $d\in\mathcal{T}^c$. Finally, note that

$$I_2(\mathcal{S},\mathcal{T})-\sum_{k\in\mathcal{S}}I(\hat{Y}_k;Y_k|U_k)=-\sum_{k\in\mathcal{S}}I(\hat{Y}_k;Y_k|X(\mathcal{S}_d),U^N,\hat{Y}(\mathcal{S}^c),Y_d,\hat{Y}(\{k'\in\mathcal{S}:k'<k\}))$$

$$=-\sum_{k\in\mathcal{S}}I(\hat{Y}_k;Y(\mathcal{S})|X(\mathcal{S}_d),U^N,\hat{Y}(\mathcal{S}^c),Y_d,\hat{Y}(\{k'\in\mathcal{S}:k'<k\}))$$

$$=-I(\hat{Y}(\mathcal{S});Y(\mathcal{S})|X(\mathcal{S}_d),U^N,\hat{Y}(\mathcal{S}^c),Y_d).$$

Therefore, the probability of error tends to zero as $n\to\infty$ if

$$R(\mathcal{T})<I(X((\mathcal{S}\cup\mathcal{T})\cap\mathcal{S}_d),U(\mathcal{S});\hat{Y}(\mathcal{S}^c),Y_d|X((\mathcal{S}^c\cap\mathcal{T}^c)\cap\mathcal{S}_d),U(\mathcal{S}^c))$$

$$-I(\hat{Y}(\mathcal{S});Y(\mathcal{S})|X(\mathcal{S}_d),U^N,\hat{Y}(\mathcal{S}^c),Y_d)-(N-1)\delta(\epsilon')-\delta(\epsilon)\quad(30)$$

for all $\mathcal{S} \subseteq [1:N]$ and $\mathcal{T} \subseteq \mathcal{S}_d$ such that $d \in \mathcal{S}^c$ and $d \in \mathcal{T}^c$. Since for every $\mathcal{S} \subseteq [1:N]$, $d \in \mathcal{S}^c$ the inequalities corresponding to $\mathcal{T} \subsetneq (\mathcal{S} \cap \mathcal{S}_d)$ are inactive due to the inequality with $\mathcal{T} = \mathcal{S} \cap \mathcal{S}_d$ in (30), the set of inequalities can be further simplified to

$$R(\mathcal{T}) < I(X(\mathcal{T}), U(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{T}^c), U(\mathcal{S}^c))$$
$$- I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X(\mathcal{S}_d), U^N, \hat{Y}(\mathcal{S}^c), Y_d) - (N-1)\delta(\epsilon') - \delta(\epsilon) \tag{31}$$

for all $\mathcal{S} \subseteq [1:N]$ and $\mathcal{S} \cap \mathcal{S}_d \subseteq \mathcal{T} \subseteq \mathcal{S}_d$ such that $d \in \mathcal{S}^c$, where $\mathcal{T}^c = \mathcal{S}_d \backslash \mathcal{T}$. Thus, the probability of decoding error tends to zero for each destination node $d \in \mathcal{D}$ as $n \to \infty$, provided that the rate tuple satisfies (31). By the union of events bound, the probability of error tends to zero as $n \to \infty$ if the rate tuple $(R_1, \ldots, R_N)$ satisfies

$$R(\mathcal{T}) < I(X(\mathcal{T}), U(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{T}^c), U(\mathcal{S}^c)) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X(\mathcal{S}_d), U^N, \hat{Y}(\mathcal{S}^c), Y_d)$$

for all $\mathcal{S} \subseteq [1:N]$, $d \in \mathcal{D}(\mathcal{S})$, and $\mathcal{S} \cap \mathcal{S}_d \subseteq \mathcal{T} \subseteq \mathcal{S}_d$ such that $\mathcal{S}^c \cap \mathcal{D}(\mathcal{S}) \neq \emptyset$, where $\mathcal{T}^c = \mathcal{S}_d \backslash \mathcal{T}$ for some $\prod_{k=1}^N p(x_k)p(\hat{y}_k|y_k, x_k)$. This completes the proof of Theorem 3 for $Q = \emptyset$. The proof for the general Q follows by coded time sharing.

APPENDIX C
COMPARISON TO A PREVIOUS EXTENSION OF THE ORIGINAL COMPRESS–FORWARD SCHEME

For a DM single-source (node 1) multicast network with destination nodes $\mathcal{D} \subseteq [2:N]$, a hybrid scheme proposed by Kramer, Gastpar, and Gupta [14, Theorem 3] gives the capacity lower bound

$$C \geq \max \min_{d \in \mathcal{D}} I(X_1; \hat{Y}_2^N, Y_d | U_2^N, X_2^N), \tag{32}$$

where the maximum is over $p(x_1) \prod_{k=2}^N p(u_k, x_k) p(\hat{y}_k | u_2^N, x_k, y_k)$ such that

$$I(\hat{Y}(\mathcal{T}); Y(\mathcal{T}) | U_2^N, X_2^N, \hat{Y}(\mathcal{T}^c), Y_d) + \sum_{k \in \mathcal{T}} I(\hat{Y}_k; X_2^N | U_2^N, X_k)$$
$$\leq I(X(\mathcal{T}); Y_d | U(\mathcal{T}), X(\mathcal{T}^c), U_d, X_d) + \sum_{t=1}^T I(U(\mathcal{K}_t); Y_{r(t)} | U(\mathcal{K}_t^c), X_{r(t)}) \tag{33}$$

for all $\mathcal{T} \subseteq [2:N]$, all partitions $\{\mathcal{K}_t\}_{t=1}^T$ of $[2:N]$, and all $r(t) \in [2:N]$ such that $r(t) \notin \mathcal{K}_t$. The complements $\mathcal{T}^c$ and $\mathcal{K}_t^c$ are the complements of the respective $\mathcal{T}$ and $\mathcal{K}_t$ in $[2:N]$.

The hybrid coding scheme achieving lower bound (32) uses an extension of the original compress–forward scheme for the relay channel as well as decoding of the compression indices at the relays. The pure compress–forward scheme without decoding gives the capacity lower bound

$$C \geq R^* = \max \min_{d \in \mathcal{D}} I(X_1; \hat{Y}_2^N, Y_d | X_2^N), \tag{34}$$

where the maximum is over all pmfs $\prod_{k=1}^{N} p(x_k)p(y_k|x_k)$ such that

$$I(Y(\mathcal{T}); \hat{Y}(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d) + \sum_{k\in\mathcal{T}} I(X_2^N; \hat{Y}_k|X_k) \leq I(X(\mathcal{T}); Y_d|X(\mathcal{T}^c), X_d)$$

for all $\mathcal{T} \subseteq [2:N]$ and $\mathcal{T}^c = [2:N]\backslash\mathcal{T}$. This is identical to (32) with $U_j = \emptyset$, $j \in [2:N]$.

In the following we show that the noisy network coding lower bound in Theorem 1 is uniformly better than lower bound (34) for every $p(y_2^N|x^N)$. By using similar steps to those in [15, Appendix C] and some algebra, lower bound (34) can be upper bounded as

$$R^* \leq \max \min_{d\in\mathcal{D}} \min_{\mathcal{T}\subseteq[2:N]} I(X_1; \hat{Y}_2^N, Y_d|X_2^N) + I(X(\mathcal{T}); Y_d|X(\mathcal{T}^c), X_d)$$

$$- I(\hat{Y}(\mathcal{T}); Y(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d) - \sum_{k\in\mathcal{T}} I(\hat{Y}_k; X_2^N|X_k)$$

$$= \max \min_{d\in\mathcal{D}} \min_{\mathcal{T}\subseteq[2:N]} I(X_1, X(\mathcal{T}); \hat{Y}(\mathcal{T}^c), Y_d|X(\mathcal{T}^c), X_d) - I(\hat{Y}(\mathcal{T}); Y(\mathcal{T})|X^N, \hat{Y}(\mathcal{T}), Y_d)$$

$$- I(X(\mathcal{T}); \hat{Y}(\mathcal{T}^c)|Y_d, X(\mathcal{T}^c), X_d) - \sum_{k\in\mathcal{T}} I(\hat{Y}_k; X_2^N|X_k), \quad (35)$$

where the maximums are over $p(x_1)\prod_{k=2}^{N} p(x_k)p(\hat{y}_k|x_k, y_k)$. Here equality (35) follows since

$$I(X_1; \hat{Y}_2^N, Y_d|X_2^N) + I(X(\mathcal{T}); Y_d|X(\mathcal{T}^c), X_d) - I(\hat{Y}(\mathcal{T}); Y(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$= I(X_1; \hat{Y}(\mathcal{T}^c), Y_d|X_2^N) + I(X_1; \hat{Y}(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$+ I(X(\mathcal{T}); Y_d|X(\mathcal{T}^c), X_d) - I(\hat{Y}(\mathcal{T}); Y(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$= I(X_1; \hat{Y}(\mathcal{T}^c), Y_d|X_2^N) + I(X_1; \hat{Y}(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d) + I(X(\mathcal{T}); \hat{Y}(\mathcal{T}^c), Y_d|X(\mathcal{T}^c), X_d)$$

$$- I(X(\mathcal{T}); \hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c), Y_d, X_d) - I(\hat{Y}(\mathcal{T}); Y(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$= I(X_1, X(\mathcal{T}); \hat{Y}(\mathcal{T}^c), Y_d|X(\mathcal{T}^c), X_d) + I(X_1; \hat{Y}(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$- I(X(\mathcal{T}); \hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c), Y_d, X_d) - I(\hat{Y}(\mathcal{T}); Y(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$= I(X_1, X(\mathcal{T}); \hat{Y}(\mathcal{T}^c), Y_d|X(\mathcal{T}^c), X_d) + I(X_1, Y(\mathcal{T}); \hat{Y}(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$- I(Y(\mathcal{T}); \hat{Y}(\mathcal{T})|X_1, X_2^N, \hat{Y}(\mathcal{T}^c), Y_d) - I(X(\mathcal{T}); \hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c), Y_d, X_d)$$

$$- I(\hat{Y}(\mathcal{T}); Y(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$= I(X_1, X(\mathcal{T}); \hat{Y}(\mathcal{T}^c), Y_d|X(\mathcal{T}^c), X_d) + I(X_1; \hat{Y}(\mathcal{T})|X_2^N, \hat{Y}(\mathcal{T}^c), Y(\mathcal{T}), Y_d)$$

$$- I(Y(\mathcal{T}); \hat{Y}(\mathcal{T})|X_1, X_2^N, \hat{Y}(\mathcal{T}^c), Y_d) - I(X(\mathcal{T}); \hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c), Y_d, X_d)$$

$$= I(X_1, X(\mathcal{T}); \hat{Y}(\mathcal{T}^c), Y_d|X(\mathcal{T}^c), X_d) - I(Y(\mathcal{T}); \hat{Y}(\mathcal{T})|X_1, X_2^N, \hat{Y}(\mathcal{T}^c), Y_d)$$

$$- I(X(\mathcal{T}); \hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c), Y_d, X_d)$$

for all $\mathcal{T} \subseteq [2 : N]$, $\mathcal{T}^c = [2 : N] \backslash \mathcal{T}$ and $d \in \mathcal{D}$, where the last equality follows from the Markovity $(X_1, X(\mathcal{T}^c), X_d, \hat{Y}(\mathcal{T}^c), Y_d) \to (X(\mathcal{T}), Y(\mathcal{T})) \to \hat{Y}(\mathcal{T})$. On the other hand, Theorem 1 can be simplified by setting $Q = \emptyset$ and $R_2 = \cdots = R_N = 0$ as

$$C \geq \max_{\prod_{k=1}^{N} p(x_k) p(\hat{y}_k | y_k, x_k)} \min_{d \in \mathcal{D}} \min_{\mathcal{T} \subseteq [2:N]} I(X_1, X(\mathcal{T}); \hat{Y}(\mathcal{T}^c), Y_d | X(\mathcal{T}^c), X_d) \tag{36}$$

$$- I(Y(\mathcal{T}); \hat{Y}(\mathcal{T}) | X^N, \hat{Y}(\mathcal{T}^c), Y_d),$$

where $\mathcal{T}^c = [2 : N] \backslash \mathcal{T}$. Thus, Theorem 1 achieves a higher rate than (34) with gap

$$I(X(\mathcal{T}); \hat{Y}(\mathcal{T}^c) | Y_d, X(\mathcal{T}^c), X_d) + \sum_{k \in \mathcal{T}} I(\hat{Y}_k; X_2^N | X_k)$$

for each $d \in \mathcal{D}$ and $\mathcal{T} \subseteq [2 : N]$.

We now present a simple example for which noisy network coding performs strictly better than the general hybrid scheme (32). Consider a 4-node noiseless network, where $\mathcal{D} = \{4\}$, $R_2 = R_3 = R_4 = 0$, and $Y_2 = X_1, Y_3 = X_2, Y_4 = X_3$ are all binary. From (5), we know that that the noisy network coding lower bound achieves the capacity $C = 1$. On the other hand, applying (32) to the above noiseless network we get

$$I(X_1; \hat{Y}_2, \hat{Y}_3, Y_4 | U_2, U_3, X_2, X_3) = I(X_1; \hat{Y}_2, \hat{Y}_3 | U_2, U_3, X_2, X_3) \tag{37}$$

$$= I(X_1; \hat{Y}_2 | U_2, U_3, X_2, X_3, \hat{Y}_3) \tag{38}$$

where (37) follows from the channel and (38) follows from the Markovity $X_1 \to (U_2, U_3, X_3, Y_3) \to \hat{Y}_3$. The constraint (33) corresponding to $\mathcal{T} = \{2\}$ and $r(1) = 4$ is

$$I(\hat{Y}_2; Y_2 | U_2, U_3, X_2, X_3, \hat{Y}_3, Y_4) + I(\hat{Y}_2; X_3 | U_2, U_3, X_2) \leq I(X_2; Y_4 | U_2, U_3, X_3) + I(U_2; Y_4 | U_3),$$

which can be simplified as

$$I(X_1; \hat{Y}_2 | U_2, U_3, X_2, X_3, \hat{Y}_3) \leq I(U_2; X_3 | U_3) - I(\hat{Y}_2; X_3 | U_2, U_3, X_2)$$

$$\leq I(U_2; X_3 | U_3)$$

$$= 0,$$

where the equality follows from $U_2 \to U_3 \to X_3$. Thus, the achievable rate of the hybrid scheme is zero for this particular example. It can be easily seen that our noisy network coding scheme outperforms the hybrid scheme for noiseless networks with more than two relays. Note that in general, due to decoding at the relay nodes, the hybrid scheme can sometimes perform better than the noisy network coding scheme without similar augmentation.

REFERENCES

[1] 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.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[3] L. R. Ford, Jr. and D. R. Fulkerson, "Maximal flow through a network," *Canad. J. Math.*, vol. 8, pp. 399–404, 1956.

[4] P. Elias, A. Feinstein, and C. E. Shannon, "A note on the maximum flow through a network," *IRE Trans. Inf. Theory*, vol. 2, no. 4, pp. 117–119, Dec. 1956.

[5] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[6] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789–804, 2006.

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

[8] S. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," 2009, submitted to *IEEE Trans. Inf. Theory,* 2009. [Online]. Available: http://arxiv.org/abs/0906.5394

[9] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[10] 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.

[11] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.

[12] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 921–930, Mar. 2009.

[13] 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.

[14] 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.

[15] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, 2006.

[16] B. Rankov and A. Wittenben, "Achievable rate region for the two-way relay channel," in *Proc. IEEE International Symposium on Information Theory*, Seattle, WA, July 2006.

[17] S. Katti, I. Maric, A. Goldsmith, D. Katabi, and M. Médard, "Joint relaying and network coding in wireless networks," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007, pp. 1101–1105.

[18] B. Djeumou, E. V. Belmega, and S. Lasaulce, "Interference relay channels – part I: Transmission rates," 2009. [Online]. Available: http://arxiv.org/abs/0904.2585v1

[19] P. Razaghi and W. Yu, "Universal relaying for the interference channel," in *Proc. ITA Workshop*, La Jolla, CA, 2010.

[20] E. Perron, "Information-theoretic secrecy for wireless networks," Ph.D. Thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Swiss, Sep. 2009.

[21] A. El Gamal and Y.-H. Kim, "Lecture notes on network information theory," 2010. [Online]. Available: http://arxiv.org/abs/1001.3404

[22] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.

[23] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

[24] T. M. Cover and Y.-H. Kim, "Capacity of a class of deterministic relay channels," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007, pp. 591–595.

[25] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, 2005.

[26] V. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[27] 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.