Noisy Network Coding

Sung Hoon Lim, *Student Member, IEEE*, Young-Han Kim, *Member, IEEE*, Abbas El Gamal, *Fellow, IEEE*, and Sae-Young Chung, *Senior Member, IEEE*

***Abstract*—A noisy network coding scheme for communicating messages between multiple sources and destinations over a general noisy network is presented. For multi-message multicast networks, the scheme naturally generalizes 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 discrete memoryless and Gaussian networks. The scheme also extends 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 lossy compression by the relay as in the compress-forward coding scheme for the relay channel. However, unlike previous compress-forward schemes in which 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 decoding of the received signals from all the blocks without uniquely 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-message networks by combining it with decoding techniques for the interference channel. For the Gaussian multicast network, noisy network coding improves the previously established gap to the cutset bound. We also demonstrate through two popular Gaussian network examples that noisy network coding can outperform conventional compress-forward, amplify-forward, and hash-forward coding schemes.**

***Index Terms*—Compress-forward, discrete memoryless network, Gaussian network, interference relay channel, network coding, noisy network coding, relaying, two-way relay channel.**

I. Introduction

CONSIDER the N-node discrete memoryless network depicted in Fig. 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,

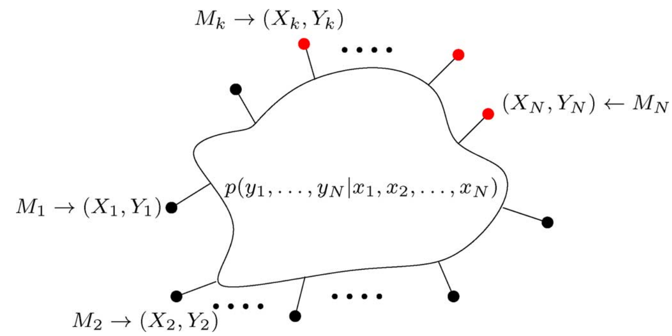

Fig. 1. N-node discrete memoryless network.

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.

Some progress has been made toward answering these questions in the past forty years. In [1] and [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-message 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 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 in two steps. For acyclic networks, the relay mappings are randomly generated and it is shown 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 multi-message multicast erasure network as a simple model for a wireless data network with packet loss. They showed that for the case when the network erasure pattern is known at the destination nodes, 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-message deterministic networks with broadcast but no interference at the receivers. Avestimehr, Diggavi, and Tse further extended this result to deterministic networks with broadcast *and* interference to obtain a lower bound on the multicast capacity that coincides with the cutset bound when each channel output is a linear

Manuscript received March 23, 2010; revised October 15, 2010; accepted January 10, 2011. Date of current version April 20, 2011. This work was supported in part by the DARPA ITMANET program, in part by the MKE/KEIT IT R&D program KI001835, and in part by the NSF CAREER Grant CCF-0747111. The material in this paper was presented in part at the IEEE Information Theory Workshop, Cairo, Egypt, January 2010, and in part at the IEEE International Symposium on Information Theory, Austin, TX, June 2010.

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

Communicated by S. Ulukus, Associate Editor for the special issue on "Interference Networks".

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

Digital Object Identifier 10.1109/TIT.2011.2119930

function of the input signals over a finite field. Their proof again involves 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 result to arbitrary nonlayered deterministic networks. Their coding scheme is further extended to layered and nonlayered Gaussian networks via a quantize-map–forward (QMF) scheme in which scalar quantization transforms a Gaussian network into a deterministic network. A bound on the achievable rate of the QMF scheme is then obtained from a *multi-letter expression*. This bound is shown to be within a constant gap of the cutset bound.

In an earlier and seemingly unrelated line of investigation, van der Meulen [8] introduced the relay channel with a single source X_1, single destination Y_3, and a 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 [9], 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 [10] and modulo-sum [11] 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 [12]. 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 [13] 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 [14] put forth the equivalent characterization of the compress-forward lower bound

$$\begin{aligned} C \geq \max\min\{&I(X_1; \hat{Y}_2, Y_3|X_2), \\ &I(X_1, X_2; Y_3) - I(Y_2; \hat{Y}_2|X_1, X_2, Y_3)\} \end{aligned} \tag{2}$$

where the maximum is over all pmfs $p(x_1)p(x_2)p(\hat{y}_2|y_2, x_2)$. As we will see, this characterization is a special case of a more general extension of compress-forward to networks.

The above two lines of investigation have motivated us to develop the noisy network coding scheme that unifies and extends the above results. On the one hand, our scheme naturally generalizes compress-forward 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 network coding scheme employs lossy compression by the relay as in the previous compress-forward coding schemes in [9] and [13]. However, unlike these schemes, where different messages are sent over multiple blocks and decoded one message at a time, each source node in noisy network coding transmits the same message over multiple blocks using independently generated codebooks. 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 decoding of the messages without uniquely decoding the compression indices. As we will demonstrate throughout the paper, these differences result in better performance than the previous compress-forward schemes [13], [15]–[18] for networks with more than one relay node or more than one message.

Note that each of the three key ideas involved in our scheme has been previously used in other network information theory problems.

1) Sending the same message over multiple blocks has been used implicitly in the time-expansion technique for cyclic noiseless networks [5] and nonlayered deterministic networks [19]. Unlike these time-expansion proofs, however, our achievability proof does not require a two-step approach that depends on the network topology.
2) Relaying the compression indices without binning has again used implicitly in network coding and its extensions.
3) Simultaneous nonunique decoding has been used in various settings, e.g., interference channels [20] and broadcast channels [21].

The key contribution of the paper lies in the manner in which these three ideas are combined and in the careful analysis of the corresponding probability of error, which yields a single-letter characterization of the achievable rate. In fact, using only two of these three ideas fails to achieve the same performance as noisy network coding. 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 multi-message multicast case, noisy network coding yields a single-letter inner bound that is within a tighter gap to the cutset bound than that of the QMF scheme by Avestimehr, Diggavi, and Tse [19] and its extension by Perron [22]. The reason for the tighter gap is that the QMF scheme uses scalar quantization and the bound on the achievable rate is obtained from a multi-letter expression. In comparison, noisy network coding uses a lossy source coding scheme (vector quantization), which, together with more general information theoretic analysis, enables us to obtain the exact single-letter expression of the achievable rate instead of a bound. We also show that noisy network coding can outperform other specialized schemes for two-way relay channels [15], [16] and interference relay channels [17], [18].

The rest of the paper is organized as follows. In the next section, we formally define the problem of communicating multiple messages 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 multi-message multicast networks. In Section IV, the scheme is extended to general multi-message networks. Results on Gaussian networks are discussed in Section V.

Throughout the paper, we follow the notation in [23]. In particular, a sequence of random variables with node index k and time index $i \in [1:n] := \{1,\ldots,n\}$ is denoted as $X_k^n := (X_{k1},\ldots,X_{kn})$. A tuple 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) $(\times_{k=1}^N \mathcal{X}_k, p(y^N|x^N), \times_{k=1}^N \mathcal{Y}_k)$ depicted in Fig. 1 consists of N sender-receiver alphabet pairs $(\mathcal{X}_k,\mathcal{Y}_k)$, $k \in [1:N]$, 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 the 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 $\mathcal{S}_d = \{k \in [1:N] : d \in \mathcal{D}_k\}$ 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.

A. Multi-Message Multicast Networks

In Section III, we establish the following noisy network coding theorem for multicasting multiple messages 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

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

for all subsets $\mathcal{S} \subset [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$ for some pmf $p(q)\prod_{k=1}^N p(x_k|q)p(\hat{y}_k|y_k,x_k,q)$, where $R(\mathcal{S}) := \sum_{k\in\mathcal{S}} R_k$.

This inner bound has a similar structure to the cutset bound

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

for all $\mathcal{S} \subset [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$. The first term of (3), 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 cyclic 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 (links), and $\mathcal{C} = \{C_{jk} \in \mathbb{R}^+ : (j,k) \in \mathcal{E}\}$ is the set of link capacities. Each node $k \in [1:N]$ transmits $X_k = (X_{kl} : (k,l) \in \mathcal{E})$ and receives $Y_k = (X_{jk} : (j,k) \in \mathcal{E})$. Thus, each link $(j,k) \in \mathcal{E}$ carries a symbol $X_{jk} \in \mathcal{X}_{jk}$ noiselessly from node j to node k with link capacity $C_{jk} = \log|\mathcal{X}_{jk}|$. Now for each cut $(\mathcal{S},\mathcal{S}^c)$ separating some source and destination pair

$$\begin{aligned} H(Y(\mathcal{S}^c)|X(\mathcal{S}^c)) &\le \sum_{k\in\mathcal{S}^c} H(X_{jk} : j \in \mathcal{S}, (j,k)\in\mathcal{E}) \\ &\le \sum_{k\in\mathcal{S}^c} \sum_{j\in\mathcal{S},(j,k)\in\mathcal{E}} H(X_{jk}) \\ &\le \sum_{\substack{(j,k)\in\mathcal{E} \\ j\in\mathcal{S},k\in\mathcal{S}^c}} C_{jk} = C(\mathcal{S}) \end{aligned}$$

with equality if $(X_{jk} : (j,k) \in \mathcal{E})$ has a uniform product pmf. Hence, by evaluating the inner bound in Theorem 1 with the uniform pmf on X^N and $\hat{Y}_k = Y_k$ for all k, it can be easily shown that the inner bound coincides with the cutset bound, and thus the capacity region is the set of rate tuples $(R_1,\ldots,R_N)$ such that

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

for all $\mathcal{S} \subset [1:N]$ with $\mathcal{D} \cap \mathcal{S}^c \neq \emptyset$. This recovers previous results in [5].

Relay Channels: Consider the relay channel $p(y_2,y_3|x_1,x_2)$. By taking $N=3$, $R_2 = R_3 = 0$, $Y_1 = X_3 = \emptyset$, and $\mathcal{D} = \{3\}$, it can be readily checked that the inner bound in Theorem 1 reduces to the alternative characterization of the compress-forward lower bound in (2).

Wireless Erasure Networks: Consider a wireless data network with packet loss modeled by a hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{E}, \mathcal{C})$ with random input erasures. Each node $k \in [1:N]$ broadcasts a symbol X_k to a subset of nodes $\mathcal{N}_k$ over a hyperedge (wireless broadcast link) $(k, \mathcal{N}_k)$ and receives $Y_k = (Y_{kj} : k \in \mathcal{N}_j)$ from nodes j for $k \in \mathcal{N}_j$, where

$$Y_{kj} = \begin{cases} \mathrm{e}, & \text{with probability } p_{kj} \\ X_j, & \text{with probability } 1 - p_{kj}. \end{cases}$$

Note that the capacity of each hyperedge $(k, \mathcal{N}_k)$ (with no erasure) is $C_k = \log |\mathcal{X}_k|$. We assume that the erasures are independent of each other. Assume further that the erasure pattern of the entire network is known at each destination node. By evaluating the inner bound in Theorem 1 with the uniform product pmf on X^N and $\hat{Y}_k = Y_k$ for all k, it can be easily shown that the inner bound coincides with the cutset bound and the capacity region is the set of rate tuples $(R_1, \ldots, R_N)$ such that

$$R(\mathcal{S}) \leq \sum_{j \in \mathcal{S}: \mathcal{N}_j \cap \mathcal{S}^c \neq \emptyset} \left(1 - \prod_{k \in \mathcal{N}_j \cap \mathcal{S}^c} p_{kj}\right) C_j \tag{6}$$

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

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

for all $\mathcal{S} \subset [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$ for some pmf $p(q) \prod_{k=1}^N p(x_k|q)$. This recovers previous results in [19] for the single-message case and in [22] for the multi-message case. Note that the inner bound in (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 $Y_k = \sum_{j=1}^N g_{kj} X_j$ [19].

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.,

$$Y_k = g_{dk}(X_1, \ldots, X_N, Y_d) \tag{8}$$

for every $k \in [1:N]$ and $d \in \mathcal{D}$. Under this semideterministic 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 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 $\mathcal{S} \subset [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$.

We also show in Appendix E that our noisy network coding scheme can strictly outperform the extension of the original compress-forward scheme for the relay channel to networks in [13, Theorem 3].

B. General Multi-Message Networks

We extend the noisy network coding theorem to general multi-message 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 $\tilde{Y}_k = (Y_k, \hat{Y}_1, \ldots, \hat{Y}_N)$ and the compressed channel outputs $(\hat{Y}_1, \ldots, \hat{Y}_N)$ are described to the destination nodes 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

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

for all subsets $\mathcal{S} \subseteq [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D}(\mathcal{S}) \neq \emptyset$ for some pmf $p(q) \prod_{k=1}^N p(x_k|q) p(\hat{y}_k|y_k, x_k, q)$, where $\mathcal{D}(\mathcal{S}) := \cup_{k \in \mathcal{S}} \mathcal{D}_k$.

The proof of this theorem is given in Section 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

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

for all subsets $\mathcal{T}, \mathcal{S} \subset [1:N]$ and $d \in \mathcal{D}(\mathcal{S})$ such that $\mathcal{S} \cap \mathcal{S}_d \subseteq \mathcal{T} \subseteq \mathcal{S}_d$ and $\mathcal{S}^c \cap \mathcal{D}(\mathcal{S}) \neq \emptyset$ for some pmf $p(q) \prod_{k=1}^N p(u_k, x_k|q) p(\hat{y}_k|y_k, u_k, q)$, 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 [24] 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 Section IV-B.

C. 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

$$Y^N = GX^N + Z^N \tag{12}$$

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 Section 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 bound for the Gaussian network (12), the rate tuple $(R_1 - 0.63N, \ldots, R_N - 0.63N)$ is in the inner bound in Theorem 1, 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 (Section V-B): Consider the Gaussian two-way relay channel

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

in which source nodes 1 and 2 wish to exchange messages reliably with the help of relay node 3 (multicast with $R_3 = 0$ and $\mathcal{D} = \{1, 2\}$). Various coding schemes for this channel have been investigated in [15], [16]. In Fig. 2, we compare noisy network coding with decode-forward, compress-forward, and amplify-forward studied in [15]. As shown in the figure, noisy network coding uniformly outperforms compress-forward. Furthermore, we can see that noisy network coding achieves uniformly within 1.5 bits from the cutset bound. In Appendix F, we show that the gap from the cutset bound for noisy network coding is within 1 bit for the individual rates and within 1.5 bits for the sum rate, while decode-forward, compress-forward, and amplify-forward have arbitrarily large gaps.

Interference Relay Channel (Section V-C): Consider the Gaussian interference relay channel with orthogonal receiver components in Fig. 3.

The channel outputs are

$$Y_j = g_{j1}X_1 + g_{j2}X_2 + Z_j, \quad j = 3, 4, 5$$

where g_{jk} is the channel gain from node k to node 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 Fig. 4, we compare noisy network coding (Theorems 2 and 3) to compress-forward (CF) and hash-forward (HF) in [18]. The curve representing noisy network coding depicts the maximum of achievable sum rates in Theorems 2 and 3. At high signal-to-noise

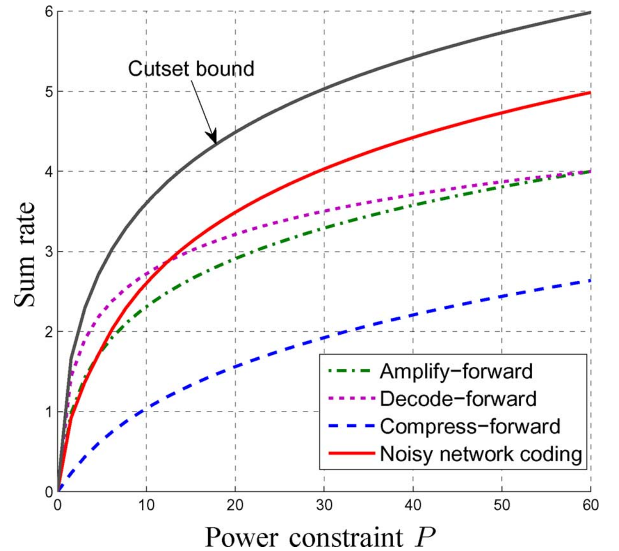

Fig. 2. Comparison of coding schemes for $g_{21} = g_{12} = 0.1$, $g_{31} = g_{23} = 0.5$, and $g_{32} = g_{13} = 2$.

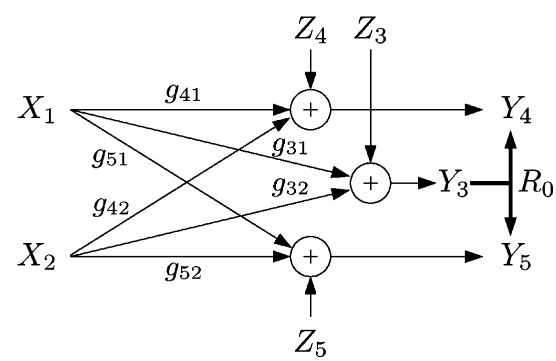

Fig. 3. Gaussian interference relay channel.

ratio (SNR), Theorem 2 provides further improvement, since decoding other messages is a better strategy when interference is strong. Note that, although not shown in the figure, Theorem 3 alone outperforms the other two schemes for all channel gains and power constraints. In Appendix G we give a detailed comparison of noisy network coding (Theorem 3) and the other two schemes.

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 [9], [13], 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 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

TABLE I
NOISY NETWORK CODING FOR THE RELAY CHANNEL

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

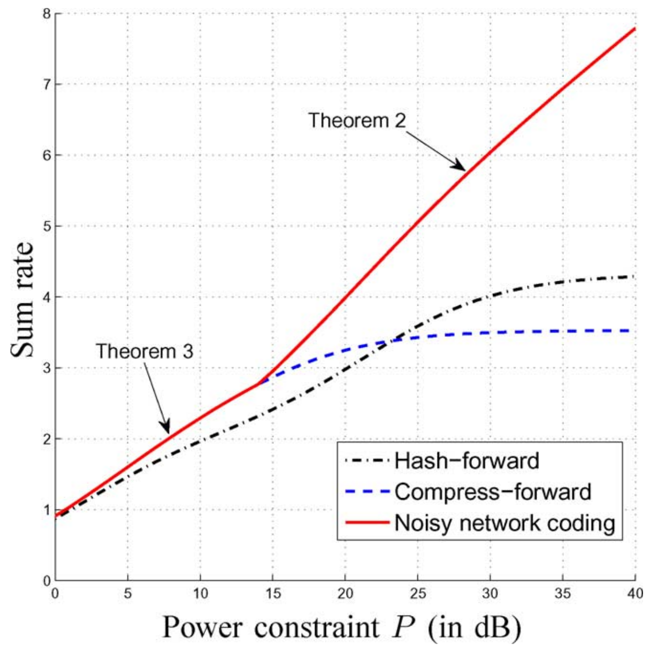

Fig. 4. Comparison of coding schemes for $g_{41} = g_{52} = 1, g_{51} = g_{42} = g_{31} = 0.5, g_{32} = 0.1, R_0 = 1$.

$m \in [1 : 2^{nbR}]$ using $(\mathbf{y}_{31}, \dots, \mathbf{y}_{3b})$ by joint typicality 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

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

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 $j \in [1 : b]$, 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 the unique message $\hat{m} \in [1 : 2^{nbR}]$ such that

$$(\mathbf{x}_{1j}(\hat{m}), \mathbf{x}_{2j}(\hat{l}_{j-1}), \hat{\mathbf{y}}_{2j}(\hat{l}_j|\hat{l}_{j-1}), \mathbf{y}_{3j}) \in \mathcal{T}_\epsilon^{(n)}$$

for all $j \in [1 : b]$ for some $\hat{l}_1, \hat{l}_2, \dots, \hat{l}_b$. If there is none or more than one such message, it declares an error.

Analysis of the Probability of Error: To bound the probability of error, assume without loss of generality that $M = 1$ and $L_1 = L_2 = \cdots = L_b = 1$. Then the decoder makes an error only if one or more of the following events occur:

$$\begin{aligned}\mathcal{E}_1 &= \big\{(\hat{\mathbf{Y}}_{2j}(l_j|1), \mathbf{X}_{2j}(1), \mathbf{Y}_{2j}) \notin \mathcal{T}_{\epsilon'}^{(n)} \\ &\qquad \text{for all } l_j \text{ for some } j \in [1 : b]\big\} \\ \mathcal{E}_2 &= \big\{(\mathbf{X}_{1j}(1), \mathbf{X}_{2j}(1), \hat{\mathbf{Y}}_{2j}(1|1), \mathbf{Y}_{3j}) \notin \mathcal{T}_\epsilon^{(n)} \\ &\qquad \text{for some } j \in [1 : b]\big\} \\ \mathcal{E}_3 &= \big\{(\mathbf{X}_{1j}(m), \mathbf{X}_{2j}(l_{j-1}), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)} \\ &\qquad \text{for all } j \text{ for some } l^b,\ m \neq 1\big\}.\end{aligned}$$

Thus, the probability of error is bounded as

$$\mathsf{P}(\mathcal{E}) \le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c) + \mathsf{P}(\mathcal{E}_3).$$

By the covering lemma [23, Lecture Note 3] and the union of events bound (over b blocks), $\mathsf{P}(\mathcal{E}_1)$ tends to zero as $n \to \infty$ if $R_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(\epsilon')$. By the conditional typicality lemma [23, Lecture Note 2] and the union of events bound, the second term $\mathsf{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c)$ tends to zero as $n \to \infty$. For the third term, define the events

$$\begin{aligned}&\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j) \\ &\quad = \{(\mathbf{X}_{1j}(m), \mathbf{X}_{2j}(l_{j-1}), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)}\}\end{aligned}$$

and consider

$$\begin{aligned}\mathsf{P}(\mathcal{E}_3) &= \mathsf{P}\big(\cup_{m\neq 1} \cup_{l^b} \cap_{j=1}^b \tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)\big) \\ &\le \sum_{m\neq 1} \sum_{l^b} \mathsf{P}\big(\cap_{j=1}^b \tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)\big) \\ &\overset{(a)}{=} \sum_{m\neq 1} \sum_{l^b} \prod_{j=1}^b \mathsf{P}(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \\ &\le \sum_{m\neq 1} \sum_{l^b} \prod_{j=2}^b \mathsf{P}(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j))\end{aligned}$$

where (a) follows since the codebook is generated independently for each block $j \in [1:b]$ 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 $(\hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j})$ and hence by the joint typicality lemma [23, Lecture Note 2]

$$\begin{aligned}&\mathsf{P}(\tilde{\mathcal{E}}_j(m,l_{j-1},l_j))\\&=\mathsf{P}\{(\mathbf{X}_{1j}(m),\mathbf{X}_{2j}(l_{j-1}),\hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}),\mathbf{Y}_{3j})\in\mathcal{T}_\epsilon^{(n)}\}\\&\le 2^{-n(I_1-\delta(\epsilon))}\end{aligned}\tag{14}$$

where $I_1 = I(X_1;\hat{Y}_2,Y_3|X_2)$. 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 $\mathbf{Y}_{3j}$. Hence, by Lemma 2 in Appendix A, which is an easy application of the joint typicality lemma

$$\mathsf{P}(\tilde{\mathcal{E}}_j(m,l_{j-1},l_j))\le 2^{-n(I_2-\delta(\epsilon))}\tag{15}$$

where $I_2 = I(X_1,X_2;Y_3) + I(\hat{Y}_2;X_1,Y_3|X_2)$. If the binary sequence l^{b-1} has k 1s, then by (14) and (15)

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

Therefore

$$\begin{aligned}&\sum_{m\neq1}\sum_{l^b}\prod_{j=2}^{b}\mathsf{P}(\tilde{\mathcal{E}}_j(m,l_{j-1},l_j))\\&=\sum_{m\neq1}\sum_{l_b}\sum_{l^{b-1}}\prod_{j=2}^{b}\mathsf{P}(\tilde{\mathcal{E}}_j(m,l_{j-1},l_j))\\&\le\sum_{m\neq1}\sum_{l_b}\sum_{k=0}^{b-1}\binom{b-1}{k}2^{n(b-1-k)\hat{R}_2}\\&\qquad\cdot 2^{-n(kI_1+(b-1-k)I_2-(b-1)\delta(\epsilon))}\\&=\sum_{m\neq1}\sum_{l_b}\sum_{k=0}^{b-1}\binom{b-1}{k}2^{-n\left(kI_1+(b-1-k)(I_2-\hat{R}_2)-(b-1)\delta(\epsilon)\right)}\\&\le\sum_{m\neq1}\sum_{l_b}\sum_{k=0}^{b-1}\binom{b-1}{k}2^{-n\left((b-1)(\min\{I_1,\,I_2-\hat{R}_2\}-\delta(\epsilon))\right)}\\&\le 2^{nbR}\cdot2^{n\hat{R}_2}\cdot2^{b}\cdot2^{-n(b-1)(\min\{I_1,\,I_2-R_2\}-\delta(\epsilon))}\end{aligned}$$

which tends to zero as $n \to \infty$ if

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

Finally, by eliminating $\hat{R}_2 > I(\hat{Y}_2;Y_2|X_2)+\delta(\epsilon')$, substituting I_1 and I_2, and taking $b \to \infty$, we have shown that the probability of error tends to zero as $n \to \infty$ if

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

This completes the proof of achievability.

We now describe the noisy network coding scheme for multimessage multicast over a general DMN $p(y^N|x^N)$. In this setting, each node is a source as well as a relay and hence it sends both a message and compression index. Furthermore, destination nodes decode a set of messages and hence the error events involve cutsets $\mathcal{T}$ of the nodes based on the messages (correctly decoded ones vs. the rest) in addition to cutsets $\mathcal{S}$ based on both the messages and the compression indices. These two types of cutsets are simplified at the last stage of the proof.

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 [23, Lecture Note 4].

Codebook Generation: Fix $\prod_{k=1}^{N} p(x_k)p(\hat{y}_k|y_k,x_k)$. For each block $j \in [1:b]$ and node $k \in [1:N]$, randomly and independently generate $2^{nbR_k} \times 2^{n\hat{R}_k}$ sequences $\mathbf{x}_{kj}(m_k,l_{k,j-1})$, $m_k \in [1:2^{nbR_k}]$, $l_{k,j-1} \in [1:2^{n\hat{R}_k}]$, 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

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

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 the 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

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

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 $(M_1,\ldots,M_N) = (1,\ldots,1) =: \mathbf{1}$ and $\mathbf{L}_1 =$

$\cdots = \mathbf{L}_b = \mathbf{1}$, where $\mathbf{L}_j = (L_{1j}, \ldots, L_{Nj})$. Then the decoder makes an error only if one of the following events occur:

$$\begin{aligned}
\mathcal{E}_1 = \big\{&(\hat{\mathbf{Y}}_{kj}(l_{kj}|1,1), \mathbf{X}_{kj}(1,1), \mathbf{Y}_{kj}) \notin \mathcal{T}_{\epsilon'}^{(n)} \\
&\text{for all } l_{kj} \text{ for some } j \in [1:b],\, k \in [1:N]\big\} \\
\mathcal{E}_2 = \big\{&(\mathbf{X}_{1j}(1,1), \ldots, \mathbf{X}_{Nj}(1,1), \\
&\hat{\mathbf{Y}}_{1j}(1|1,1), \ldots, \hat{\mathbf{Y}}_{Nj}(1|1,1), \mathbf{Y}_{dj}) \notin \mathcal{T}_{\epsilon}^{(n)} \\
&\text{for some } j \in [1:b]\big\} \\
\mathcal{E}_3 = \big\{&(\mathbf{X}_{1j}(m_1, l_{1,j-1}), \ldots, \mathbf{X}_{Nj}(m_N, l_{N,j-1}), \\
&\hat{\mathbf{Y}}_{1j}(l_{1j}|m_1, l_{1,j-1}), \ldots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|m_N, l_{N,j-1}), \\
&\mathbf{Y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for all } j \in [1:b] \text{ for some } \mathbf{m} \neq \mathbf{1} \\
&\text{with } m_d = 1 \text{ and } \mathbf{l}^b \text{ with } l_{dj} = 1, j \in [1:b]\big\}.
\end{aligned}$$

Thus, the probability of error is bounded as

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

By the covering lemma and the union of events bound, $\mathsf{P}(\mathcal{E}_1)$ tends to zero as $n \to \infty$ if $R_k > I(\hat{Y}_k; Y_k|X_k) + \delta(\epsilon'), k \in [1:N]$. By the Markov lemma [23, Lecture Note 13] and the union of events bound, the second term $\mathsf{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c)$ tends to zero as $n \to \infty$. For the third term, define the events

$$\begin{aligned}
&\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j) \\
&\quad = \big\{(\mathbf{X}_{1j}(m_1, l_{1,j-1}), \ldots, \mathbf{X}_{Nj}(m_N, l_{N,j-1}), \\
&\qquad \hat{\mathbf{Y}}_{1j}(l_{1j}|m_1, l_{1,j-1}), \ldots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|m_N, l_{N,j-1}), \\
&\qquad \mathbf{Y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)}\big\}.
\end{aligned}$$

Then

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_3) &= \mathsf{P}(\cup_{\mathbf{m}\neq\mathbf{1}} \cup_{\mathbf{l}^b} \cap_{j=1}^{b} \tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\leq \sum_{\mathbf{m}\neq\mathbf{1}} \sum_{\mathbf{l}^b} \mathsf{P}(\cap_{j=1}^{b} \tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\overset{(a)}{=} \sum_{\mathbf{m}\neq\mathbf{1}} \sum_{\mathbf{l}^b} \prod_{j=1}^{b} \mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\leq \sum_{\mathbf{m}\neq\mathbf{1}} \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j))
\end{aligned}$$

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

For each $\mathbf{m}$, $\mathbf{l}^b$, and $j \in [2:b]$, define $\mathcal{S}_j(\mathbf{m}, \mathbf{l}^b) = \{k \in [1:N] : m_k \neq 1 \text{or } l_{k,j-1} \neq 1\}$. 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 $\mathcal{T}(\mathbf{m}) = \{k \in [1:N] : m_k \neq 1\}$. Note that $\mathcal{T}(\mathbf{m}) \subseteq \mathcal{S}_j(\mathbf{m}, \mathbf{l}_{j-1})$ and $d \in \mathcal{S}_j^c(\mathbf{m}, \mathbf{l}_{j-1}) \subseteq \mathcal{T}^c(\mathbf{m})$ (recall our convention $m_d = 1$).

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 Lemma 2 and the fact that

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

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

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

where

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

Furthermore, by 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

$$\begin{aligned}
&\sum_{\mathbf{l}_{j-1}} 2^{-n(I_1(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) + I_2(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) - \delta(\epsilon))} \\
&\quad = \sum_{\substack{\mathcal{S}: d\in\mathcal{S}^c, \\ \mathcal{T}(\mathbf{m})\subseteq\mathcal{S}}} \sum_{\substack{\mathbf{l}_{j-1}: \\ \mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})=\mathcal{S}}} 2^{-n(I_1(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) + I_2(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) - \delta(\epsilon))} \\
&\quad \leq \sum_{\substack{\mathcal{S}: d\in\mathcal{S}^c, \\ \mathcal{T}(\mathbf{m})\subseteq\mathcal{S}}} 2^{-n(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k\in\mathcal{S}} \hat{R}_k - \delta(\epsilon))} \\
&\quad \leq 2^{N-1} 2^{-n(\min_{\mathcal{S}}(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k\in\mathcal{S}} \hat{R}_k - \delta(\epsilon)))}
\end{aligned}$$

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

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3) \\
&\leq \sum_{\mathbf{m}\neq\mathbf{1}} \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&= \sum_{\mathbf{m}\neq\mathbf{1}} \sum_{\mathbf{l}_b} \sum_{\mathbf{l}^{b-1}} \prod_{j=2}^{b} \mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\leq \sum_{\mathbf{m}\neq\mathbf{1}} \sum_{\mathbf{l}_b} \sum_{\mathbf{l}^{b-1}} \prod_{j=2}^{b} 2^{-n(I_1(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) + I_2(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) - \delta(\epsilon))} \\
&= \sum_{\mathbf{m}\neq\mathbf{1}} \sum_{\mathbf{l}_b} \prod_{j=2}^{b} \left(\sum_{\mathbf{l}_{j-1}} 2^{-n(I_1(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) + I_2(\mathcal{S}_j(\mathbf{m},\mathbf{l}_{j-1})) - \delta(\epsilon))} \right) \\
&\leq \sum_{\substack{\mathcal{T}\subset[1:N]: \\ \mathcal{T}\neq\emptyset, d\in\mathcal{T}^c}} 2^{\sum_{k\in\mathcal{T}} nbR_k} 2^{\sum_{k\neq d} n\hat{R}_k} 2^{(N-1)(b-1)} \\
&\qquad\qquad \cdot 2^{n\left(-(b-1)\min_{\mathcal{S}}(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k\in\mathcal{S}} \hat{R}_k - \delta(\epsilon))\right)}
\end{aligned} \tag{16}$$

where the minimum in (16) is over all $\mathcal{S} \subset [1:N]$ such that $\mathcal{T} \subseteq \mathcal{S}$ and $d \in \mathcal{S}^c$. Thus, (16) tends to zero as $n \to \infty$ if

$$R(\mathcal{T}) < \frac{b-1}{b}\left(\left(I_1(\mathcal{S})+I_2(\mathcal{S})-\sum_{k\in\mathcal{S}}\hat{R}_k\right)-\delta(\epsilon)\right)-\frac{1}{b}\sum_{k\neq d}\hat{R}_k$$

for all $\mathcal{T}, \mathcal{S} \subset [1:N]$ such that $\emptyset \neq \mathcal{T} \subseteq \mathcal{S}$ and $d \in \mathcal{S}^c$. 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

$$R(\mathcal{T}) < I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k\in\mathcal{S}} I(\hat{Y}_k; Y_k|X_k) - (N-1)\delta(\epsilon') - \delta(\epsilon)$$

for all $\mathcal{T}, \mathcal{S} \subset [1:N]$ such that $\emptyset \neq \mathcal{T} \subseteq \mathcal{S}$ and $d \in \mathcal{S}^c$. Furthermore, note that

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

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

$$\begin{aligned} R(\mathcal{T}) < &I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d|X(\mathcal{S}^c)) \\ &- I(\hat{Y}(\mathcal{S}); Y(\mathcal{S})|X^N, \hat{Y}(\mathcal{S}^c), Y_d) \\ &- (N-1)\delta(\epsilon') - \delta(\epsilon) \end{aligned} \tag{17}$$

for all $\mathcal{S}, \mathcal{T} \subset [1:N]$ such that $\emptyset \neq \mathcal{T} \subseteq \mathcal{S}$ and $d \in \mathcal{S}^c$. Since for every $\mathcal{S} \subset [1:N]$ 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 (17), the set of inequalities can be further simplified to

$$\begin{aligned} R(\mathcal{S}) < &I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d|X(\mathcal{S}^c)) \\ &- I(\hat{Y}(\mathcal{S}); Y(\mathcal{S})|X^N, \hat{Y}(\mathcal{S}^c), Y_d) \\ &- (N-1)\delta(\epsilon') - \delta(\epsilon) \end{aligned} \tag{18}$$

for all $\mathcal{S} \subset [1:N]$ 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 (18).

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 $\mathcal{S} \subset [1:N]$ 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 subsets $\mathcal{S} \subset [1:N]$ 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 Multi-Message Networks

A. Proof of Theorem 2 via Multicast Completion With Simultaneous Nonunique 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 the 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 $(\hat{l}_{1j}, \ldots, \hat{l}_{Nj})$ satisfying

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

for all $j \in [1:b]$, 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 $\hat{l}_{dj} = l_{dj}$, $j \in [1:b]$.

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 B.

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

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

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

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

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 nonunique 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 the unique index tuple $(\hat{m}_{kd} : k \in \mathcal{S}_d)$ such that there exists some $(\hat{l}_{1j}, \ldots, \hat{l}_{Nj})$ satisfying

$$\begin{aligned}((\mathbf{x}_{kj}(\hat{m}_{kd}|\hat{l}_{k,j-1}) : k \in \mathcal{S}_d), \mathbf{u}_{1j}(\hat{l}_{1,j-1}), \ldots, \mathbf{u}_{Nj}(\hat{l}_{N,j-1}),\\ \hat{\mathbf{y}}_{1j}(\hat{l}_{1j}|\hat{l}_{1,j-1}), \ldots, \hat{\mathbf{y}}_{Nj}(\hat{l}_{Nj}|\hat{l}_{N,j-1}), \mathbf{y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)}\end{aligned}$$

for all $j \in [1:b]$, 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 $\hat{l}_{dj} = l_{dj}$, $j \in [1:b]$.

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

V. Gaussian Networks

We consider the additive white Gaussian noise 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 $\mathcal{S} \subset [1:N]$, define a channel gain submatrix $G(\mathcal{S})$ such that

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

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

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

The cutset bound for the Gaussian multi-message multicast network can be further upper bounded by the use of the following lemma, the proof of which is delegated to Appendix D.

Lemma 1: Let A and B be $t\times t$ positive semidefinite matrices such that $\operatorname{rank}(A) \leq r$, $0 < r \leq t$, and $\operatorname{tr}(B) \leq t$. Suppose $\gamma \geq (e-1)r/t$ and

$$\alpha = \alpha(\gamma, r/t) = \begin{cases} e^{\gamma/e}, & \text{if } r/t \geq \gamma/e \\ (\gamma t/r)^{r/t}, & \text{otherwise.} \end{cases}$$

Then, $|I + AB| \leq |\alpha I + \alpha\gamma^{-1}A|$.

Now let $\gamma \geq e-1$ and consider the cutset bound

$$\begin{aligned} R(\mathcal{S}) &\leq I(X(\mathcal{S}); Y(\mathcal{S}^c)|X(\mathcal{S}^c)) \\ &= h(Y(\mathcal{S}^c)|X(\mathcal{S}^c)) - h(Y(\mathcal{S}^c)|X^N) \\ &= h\big(G(\mathcal{S})X(\mathcal{S}) + Z(\mathcal{S}^c)|X(\mathcal{S}^c)\big) - h(Y(\mathcal{S}^c)|X^N) \\ &\leq h\big(G(\mathcal{S})X(\mathcal{S}) + Z(\mathcal{S}^c)\big) - h(Y(\mathcal{S}^c)|X^N) \\ &\leq \frac{1}{2}\log(2\pi e)^{|\mathcal{S}^c|}\Big|I + G(\mathcal{S})K_{X(\mathcal{S})}G^T(\mathcal{S})\Big| \\ &\quad - \frac{|\mathcal{S}^c|}{2}\log(2\pi e) \\ &= \frac{1}{2}\log\Big|I + G(\mathcal{S})K_{X(\mathcal{S})}G^T(\mathcal{S})\Big| \\ &= \frac{1}{2}\log\Big|I + G^T(\mathcal{S})G(\mathcal{S})K_{X(\mathcal{S})}\Big| \\ &\overset{(a)}{\leq} \frac{1}{2}\log\Big|\alpha I + \alpha\gamma^{-1}G^T(\mathcal{S})G(\mathcal{S})P\Big| \\ &= \frac{|\mathcal{S}|}{2}\log(\alpha) + \frac{1}{2}\log\Big|I + \frac{P}{\gamma}G^T(\mathcal{S})G(\mathcal{S})\Big| \end{aligned} \tag{19}$$

where $K_{X(\mathcal{S})}$ is the covariance matrix of $X(\mathcal{S})$ and (a) follows by Lemma 1 with $\operatorname{tr}(K) \leq |\mathcal{S}|P$, $\alpha = \alpha(\gamma, r/|\mathcal{S}|)$, and $r = \min\{|\mathcal{S}|, |\mathcal{S}^c|\}$.

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

$$R(\mathcal{S}) < \frac{1}{2}\log\left|I + \frac{P}{\gamma}G(\mathcal{S})G^T(\mathcal{S})\right| - \frac{|\mathcal{S}|}{2}\log\left(\frac{\gamma}{\gamma-1}\right) \tag{20}$$

for all $\mathcal{S} \subset [1:N]$ with $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$, where $\gamma > 1$. To show this, first note that by the standard procedure [23, Lecture Note 3], Theorem 1 for the discrete memoryless network can be easily adapted for the Gaussian network with power constraint, which yields the inner bound in (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 variance $\gamma - 1$. Then for each $\mathcal{S} \subset [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$ and $d \in \mathcal{S}^c \cap \mathcal{D}$

$$\begin{aligned} I(\hat{Y}(\mathcal{S}); Y(\mathcal{S})|X^N, \hat{Y}(\mathcal{S}^c), Y_d) &\leq I(\hat{Y}(\mathcal{S}); Y(\mathcal{S})|X^N) \\ &= h(\hat{Y}(\mathcal{S})|X^N) - h(\hat{Y}(\mathcal{S})|Y(\mathcal{S}), X^N) \\ &= \frac{|\mathcal{S}|}{2}\log(2\pi e\gamma) - \frac{|\mathcal{S}|}{2}\log(2\pi e(\gamma-1)) \\ &= \frac{|\mathcal{S}|}{2}\log\left(\frac{\gamma}{\gamma-1}\right) \end{aligned}$$

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

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

Substituting these two bounds in Theorem 1 yields (20).

We now show that if a rate tuple $(R_1, \ldots, R_N)$ is in the cutset bound, then $(R_1 - \Delta^*, \ldots, R_N - \Delta^*)$ is in the inner bound in Theorem 1, where

$$\Delta^* \leq \min_{\gamma \geq e-1} \max_{\mathcal{S} \subseteq [1:N]} \frac{|S|}{2} \log \left(\frac{\alpha \gamma}{\gamma - 1} \right)$$

$\alpha = \alpha(\gamma, r/|\mathcal{S}|)$, and $r = \min\{|\mathcal{S}|, |\mathcal{S}^c|\}$. Let

$$\Delta(\gamma) = \max_{\mathcal{S} \subseteq [1:N]} \Delta_o(\mathcal{S}, \gamma) + \Delta_i(\mathcal{S}, \gamma)$$

where $\Delta_o(\mathcal{S}, \gamma) = \frac{|\mathcal{S}|}{2} \log(\alpha)$ and $\Delta_i(\mathcal{S}, \gamma) = \frac{|\mathcal{S}|}{2} \log \left(\frac{\gamma}{\gamma - 1} \right)$. Then, by (19), for $\gamma \geq e - 1$ and any rate tuple $(R_1, \ldots, R_N)$ in the cutset bound, the rate tuple $(R_1 - \Delta(\gamma), \ldots, R_N - \Delta(\gamma))$ must satisfy

$$\sum_{k \in \mathcal{S}} (R_k - \Delta(\gamma)) \leq \frac{1}{2} \log \left| I + \frac{P}{\gamma} G^T(\mathcal{S}) G(\mathcal{S}) \right| - (|\mathcal{S}| \Delta(\gamma) - \Delta_o(\mathcal{S}, \gamma))$$

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

$$|\mathcal{S}| \Delta(\gamma) - \Delta_o(\mathcal{S}, \gamma) \geq \Delta(\gamma) - \Delta_o(\mathcal{S}, \gamma) \geq \Delta_i(\mathcal{S}, \gamma)$$

we have

$$\sum_{k \in \mathcal{S}} (R_k - \Delta(\gamma)) \leq \frac{1}{2} \log \left| I + \frac{P}{\gamma} G(\mathcal{S})^T G(\mathcal{S}) \right| - \Delta_i(\mathcal{S}, \gamma)$$

for all $\mathcal{S} \subset [1:N]$ such that $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$. Thus, by (20), the rate tuple $(R_1 - \Delta(\gamma), \ldots, R_N - \Delta(\gamma))$ is achievable for every $\gamma \geq e - 1$.

Finally, we minimize $\Delta(\gamma)$ over γ. Consider

$$\begin{aligned}
\Delta^* &\leq \min_{\gamma \geq e-1} \max_{\mathcal{S} \subseteq [1:N]} \frac{|S|}{2} \log \left(\frac{\alpha \gamma}{\gamma - 1} \right) \\
&\leq \min_{\gamma \geq e-1} \max_{\mathcal{S} \subseteq [1:N]} \frac{N}{2} \Delta(|\mathcal{S}|/N, \gamma) \\
&\leq \min_{\gamma \geq e-1} \max_{\mu \in [0:1]} \frac{N}{2} \Delta(\mu, \gamma)
\end{aligned}$$

where, for $e - 1 \leq \gamma \leq e$

$$\Delta(\mu, \gamma) = \begin{cases} \mu \log \frac{\gamma e^{\gamma/e}}{\gamma - 1}, & \text{if } \mu \leq \frac{e}{e+\gamma} \\ \mu \log \frac{\gamma}{\gamma - 1} + (1 - \mu) \log \frac{\gamma \mu}{1 - \mu}, & \text{otherwise} \end{cases}$$

and for $\gamma > e$

$$\Delta(\mu, \gamma) = \begin{cases} \mu \log \frac{\gamma^2}{\gamma - 1}, & \text{if } \mu \leq \frac{1}{2} \\ \mu \log \frac{\gamma}{\gamma - 1} + (1 - \mu) \log \frac{\gamma \mu}{1 - \mu}, & \text{otherwise.} \end{cases}$$

Numerical evaluation of this bound yields $\Delta^* \leq 0.63N$, which completes the proof of Theorem 4.

B. Gaussian Two-Way Relay Channels

Recall the Gaussian two-way relay channel model (13) in Section II. Rankov and Wittneben [15] showed that the decode-forward (DF) coding scheme results in the inner bound on the capacity region that consists of all rate pairs (R_1, R_2) such that

$$\begin{aligned}
R_1 &< \min\{\mathsf{C}((1 - \rho_1^2) g_{31}^2 P), \\
&\qquad \mathsf{C}(g_{21}^2 P + g_{23}^2 \alpha P + 2\rho_1 g_{21} g_{23} \sqrt{\alpha} P)\} \\
R_2 &< \min\{\mathsf{C}((1 - \rho_2^2) g_{32}^2 P), \\
&\qquad \mathsf{C}(g_{12}^2 P + g_{13}^2 \bar{\alpha} P + 2\rho_2 g_{12} g_{13} \sqrt{\bar{\alpha}} P)\} \\
R_1 + R_2 &< \mathsf{C}((1 - \rho_1^2) g_{31}^2 + (1 - \rho_2^2) g_{32}^2 P)
\end{aligned}$$

for some $0 \leq \rho_1, \rho_2 \leq 1$ and $0 \leq \alpha \leq 1$, while 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 $\alpha \leq \sqrt{P/((g_{31}^2 + g_{32}^2) P + 1)}$, where $a_1 = 1 + (g_{21}^2 + g_{23}^2 g_{31}^2 \alpha^2) P / (g_{23}^2 \alpha^2 + 1)$, $a_2 = 1 + (g_{12}^2 + g_{13}^2 g_{32}^2 \alpha^2) P / (g_{13}^2 \alpha^2 + 1)$, $b_1 = 2 g_{21} g_{23} g_{31} \alpha P / (g_{23}^2 \alpha^2 + 1)$, and $b_2 = 2 g_{12} g_{13} g_{32} \alpha P / (g_{13}^2 \alpha^2 + 1)$. 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

$$\begin{aligned}
R_1 &< \mathsf{C}\left(\frac{g_{31}^2 P + (1 + \sigma^2) g_{21}^2 P}{1 + \sigma^2} \right) \\
R_2 &< \mathsf{C}\left(\frac{g_{32}^2 P + (1 + \sigma^2) g_{12}^2 P}{1 + \sigma^2} \right)
\end{aligned} \tag{21}$$

for some

$$\sigma^2 \geq \max\left\{ \frac{1 + g_{21}^2 P + g_{31}^2 P}{\min\{g_{23}^2, g_{13}^2\} P}, \frac{1 + g_{12}^2 P + g_{32}^2 P}{\min\{g_{23}^2, g_{13}^2\} P} \right\}. \tag{22}$$

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

$$\begin{aligned} R_1 &\le \min\{I(X_1;Y_2,\hat{Y}_3|X_2,X_3,Q), \\ &\qquad I(X_1,X_3;Y_2|X_2,Q) - I(Y_3;\hat{Y}_3|X_1,X_2,X_3,Y_2,Q)\} \\ R_2 &\le \min\{I(X_2;Y_1,\hat{Y}_3|X_1,X_3,Q), \\ &\qquad I(X_2,X_3;Y_1|X_1,Q) - I(Y_3;\hat{Y}_3|X_1,X_2,X_3,Y_1,Q)\} \end{aligned}$$

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

$$\begin{aligned} R_1 &< \mathsf{C}\left(\frac{g_{31}^2 P + (1+\sigma^2) g_{21}^2 P}{1+\sigma^2}\right) \\ R_1 &< \mathsf{C}(g_{21}^2 P + g_{23}^2 P) - \mathsf{C}(1/\sigma^2) \\ R_2 &< \mathsf{C}\left(\frac{g_{32}^2 P + (1+\sigma^2) g_{12}^2 P}{1+\sigma^2}\right) \\ R_2 &< \mathsf{C}(g_{12}^2 P + g_{13}^2 P) - \mathsf{C}(1/\sigma^2) \end{aligned} \tag{23}$$

for some $\sigma^2 > 0$. The sum rates of the above inner bounds are plotted in Fig. 2 in Section II.

C. Gaussian Interference Relay Channels

Recall the Gaussian interference relay channel model with orthogonal receiver components in Fig. 4. Djeumou, Belmaga, and Lasaulce [17], and Razaghi and Yu [18] showed that an extension of the original compress-forward (CF) coding scheme for the relay channel to the interference relay channel results in the inner bound on the capacity region that consists of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< \mathsf{C}\left(\frac{(g_{31}^2 + (1+\sigma^2) g_{41}^2)P + (g_{32}g_{41} - g_{42}g_{31})^2 P^2}{1+\sigma^2 + (g_{32}^2 + (1+\sigma^2) g_{42}^2)P}\right) \\ R_2 &< \mathsf{C}\left(\frac{(g_{32}^2 + (1+\sigma^2) g_{52}^2)P + (g_{31}g_{52} - g_{51}g_{32})^2 P^2}{1+\sigma^2 + (g_{31}^2 + (1+\sigma^2) g_{51}^2)P}\right) \end{aligned} \tag{24}$$

for some

$$\sigma^2 \ge \frac{1}{2^{2R_0}-1} \cdot \max\left\{ \frac{(g_{31}g_{42} - g_{32}g_{41})^2 P^2 + a_1}{(g_{41}^2 P + g_{42}^2 P + 1)}, \frac{(g_{31}g_{52} - g_{32}g_{51})^2 P^2 + a_2}{(g_{51}^2 P + g_{52}^2 P + 1)} \right\}$$

where $a_1 = (g_{31}^2 + g_{41}^2 + g_{32}^2 + g_{42}^2)P + 1$ and $a_2 = (g_{31}^2 + g_{51}^2 + g_{32}^2 + g_{52}^2)P + 1$.

Razaghi and Yu [18] generalized the hash-forward coding scheme [10], [25] 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 yields the inner bound on the capacity region that consists of the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< \mathsf{C}\left(\frac{g_{41}^2 P}{g_{42}^2 P + 1}\right) + R_0 - \mathsf{C}\left(\frac{(g_{32}^2 + g_{42}^2)P + 1}{(g_{42}^2 P + 1)\sigma^2}\right) \\ R_2 &< \mathsf{C}\left(\frac{g_{52}^2 P}{g_{51}^2 P + 1}\right) + R_0 - \mathsf{C}\left(\frac{(g_{31}^2 + g_{51}^2)P + 1}{(g_{51}^2 P + 1)\sigma^2}\right) \end{aligned} \tag{25}$$

for some $\sigma^2 > 0$ satisfying

$$\sigma^2 \le \frac{1}{2^{2R_0}-1} \cdot \min\left\{ \frac{(g_{31}g_{42} - g_{32}g_{41})^2 P^2 + a_1}{(g_{41}^2 P + g_{42}^2 P + 1)}, \frac{(g_{31}g_{52} - g_{32}g_{51})^2 P^2 + a_2}{(g_{51}^2 P + g_{52}^2 P + 1)} \right\}$$

where a_1 and a_2 are the same as above.

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

$$\begin{aligned} R_1 &< \mathsf{C}(g_{41}^2 P) + R_0 - \mathsf{C}(1/\sigma^2) \\ R_2 &< \mathsf{C}\left(\frac{(g_{31}^2 + (1+\sigma^2) g_{41}^2)P}{1+\sigma^2}\right) \\ R_2 &< \mathsf{C}(g_{52}^2 P) + R_0 - \mathsf{C}(1/\sigma^2) \\ R_2 &< \mathsf{C}\left(\frac{(g_{32}^2 + (1+\sigma^2) g_{52}^2)P}{1+\sigma^2}\right) \\ R_1 + R_2 &< \mathsf{C}((g_{41}^2 + g_{42}^2)P) + R_0 - \mathsf{C}(1/\sigma^2) \\ R_1 + R_2 &< \mathsf{C}\left(\frac{aP + (1+\sigma^2)(g_{41}^2 + g_{42}^2)P + b_1^2 P^2}{1+\sigma^2}\right) \\ R_1 + R_2 &< \mathsf{C}((g_{51}^2 + g_{52}^2)P) + R_0 - \mathsf{C}(1/\sigma^2) \\ R_1 + R_2 &< \mathsf{C}\left(\frac{aP + (1+\sigma^2)(g_{52}^2 + g_{51}^2)P + b_2^2 P^2}{1+\sigma^2}\right) \end{aligned}$$

where $a = g_{31}^2 + g_{32}^2$, $b_1 = g_{31}g_{42} - g_{32}g_{41}$, and $b_2 = g_{32}g_{51} - g_{31}g_{52}$, 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

$$\begin{aligned} R_1 &< \mathsf{C}\left(\frac{g_{41}^2 P}{g_{42}^2 P + 1}\right) + R_0 - \mathsf{C}\left(\frac{(g_{32}^2 + g_{42}^2)P + 1}{(g_{42}^2 P + 1)\sigma^2}\right) \\ R_1 &< \mathsf{C}\left(\frac{(g_{31}^2 + (1+\sigma^2) g_{41}^2)P + (g_{32}g_{41} - g_{42}g_{31})^2 P^2}{1+\sigma^2 + (g_{32}^2 + (1+\sigma^2) g_{42}^2)P}\right) \\ R_2 &< \mathsf{C}\left(\frac{g_{52}^2 P}{g_{51}^2 P + 1}\right) + R_0 - \mathsf{C}\left(\frac{(g_{31}^2 + g_{51}^2)P + 1}{(g_{51}^2 P + 1)\sigma^2}\right) \\ R_2 &< \mathsf{C}\left(\frac{(g_{32}^2 + (1+\sigma^2) g_{52}^2)P + (g_{31}g_{52} - g_{51}g_{32})^2 P^2}{1+\sigma^2 + (g_{31}^2 + (1+\sigma^2) g_{51}^2)P}\right) \end{aligned} \tag{26}$$

for some $\sigma^2 > 0$. The sum rates of these inner bounds are plotted in Fig. 4. In Appendix G, we show that the inner bound in (26) is tighter than both compress-forward and hash-forward inner bounds in (24) and (25).

VI. Concluding Remarks

We presented a new noisy network coding scheme and used it to establish inner bounds on the capacity region of general multi-message 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, a source node sends the same message over multiple blocks, second, the relays do not use Wyner-Ziv coding (no binning indices to decode), and third, simultaneous nonunique decoding over all blocks is used without requiring the compression indices to be decoded uniquely.

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 Gaussian 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 [9] and its extensions to networks [13], [26]. Further improvement can be obtained by only partially decoding of messages at the relays [9], and by combining noisy network coding with partial decode-forward to obtain the type of hybrid schemes in [9] and [13]. Another direction for improvement is to incorporate more sophisticated compression strategies at relays. For instance, as shown in [27], each relay node can compress its observation into multiple indices to leverage asymmetry in link quality and side information. On the receiver side, noisy network coding for multiple messages can be potentially improved by using more sophisticated interference coding schemes, such as interference alignment [28] and Han–Kobayashi superposition coding [29].

Appendix A
An Application of the Joint Typicality Lemma

Lemma 2: Let $(X^N, Y^N, Z) \sim p(x^N, y^N, z)$. Let $\tilde{\mathbf{Z}}$ be distributed according to an arbitrary pmf $p(\tilde{\mathbf{z}})$ and

$$(\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_N) \sim \prod_{i=1}^{n} p_{X^N}(\tilde{x}_{1i}, \ldots, \tilde{x}_{Ni}) \prod_{k=1}^{N} \prod_{i=1}^{n} p_{Y_k|X_k}(\tilde{y}_{ki}|\tilde{x}_{ki})$$

independent of $\tilde{\mathbf{Z}}$. Then there exists $\delta(\epsilon)$ that tends to zero as $\epsilon \to 0$ such that

$$\mathsf{P}\{(\tilde{\mathbf{Z}}, \tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_N) \in \mathcal{T}_\epsilon^{(n)}\} \le 2^{-n(I(Z;X^N) + \sum_{k=1}^N I(Y_k; X^N, Y^{k-1}, Z|X_k) - \delta(\epsilon))}.$$

Proof: By the joint typicality lemma [23, Lecture Note 2]

$$\begin{aligned}
&\mathsf{P}\{(\tilde{\mathbf{Z}}, \tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_N, \tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_N) \in \mathcal{T}_\epsilon^{(n)}\} \\
&= \sum_{(\tilde{\mathbf{z}}, \tilde{\mathbf{x}}^N, \tilde{\mathbf{y}}^N) \in \mathcal{T}_\epsilon^{(n)}(Z, X^N, Y^N)} p(\tilde{\mathbf{z}}) \cdot p(\tilde{\mathbf{x}}^N) \cdot \prod_{k=1}^{N} p(\tilde{\mathbf{y}}_k|\tilde{\mathbf{x}}_k) \\
&= \sum_{\tilde{\mathbf{z}} \in \mathcal{T}_\epsilon^{(n)}(Z)} p(\tilde{\mathbf{z}}) \sum_{\tilde{\mathbf{x}}^N \in \mathcal{T}_\epsilon^{(n)}(X^N|\tilde{\mathbf{z}})} p(\tilde{\mathbf{x}}^N) \\
&\quad \cdot \prod_{k=1}^{N} \sum_{\tilde{\mathbf{y}}_k \in \mathcal{T}_\epsilon^{(n)}(Y_k|\tilde{\mathbf{x}}^N, \tilde{\mathbf{z}}, \tilde{\mathbf{y}}^{k-1})} p(\tilde{\mathbf{y}}_k|\tilde{\mathbf{x}}_k) \\
&\le \sum_{\tilde{\mathbf{z}} \in \mathcal{T}_\epsilon^{(n)}(Z)} p(\tilde{\mathbf{z}}) 2^{-n(I(Z;X^N) - \delta_1(\epsilon))} \\
&\qquad \cdot 2^{-n(\sum_{k=1}^N I(Y_k; X^N, Z, Y^{k-1}|X_k) - \delta_2(\epsilon))} \\
&\le 2^{-n(I(Z;X^N) + \sum_{k=1}^N I(Y_k; X^N, Z, Y^{k-1}|X_k) - \delta(\epsilon))}.
\end{aligned}$$

■

Appendix B
Error Probability Analysis for Theorem 2

The analysis follows the same steps of the multicast case except that $\mathbf{m} = (m_k : k \in \mathcal{S}_d)$ with $m_d = 1$ in error event $\mathcal{E}_3$. Thus

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3) \\
&\le \sum_{\mathbf{m}} \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\le \sum_{\substack{\mathcal{T} \subset [1:N]: \\ \mathcal{T} \cap \mathcal{S}_d \ne \emptyset, d \in \mathcal{T}^c}} 2^{\sum_{k \in \mathcal{T}} nbR_k} 2^{\sum_{k \ne d} n\hat{R}_k} 2^{(N-1)(b-1)} \\
&\quad \cdot 2^{n\left(-(b-1)\min_{\mathcal{S}}(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon))\right)}
\end{aligned} \tag{27}$$

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

$$R(\mathcal{T}) < \frac{b-1}{b}\left(\left(I_1(\mathcal{S}) + I_2(\mathcal{S}) - \sum_{k \in \mathcal{S}} \hat{R}_k\right) - \delta(\epsilon)\right) - \frac{1}{b}\sum_{k \ne d} \hat{R}_k$$

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

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

for all $\mathcal{S} \subset [1:N]$ such that $\mathcal{S} \cap \mathcal{S}_d \ne \emptyset$ and $d \in \mathcal{S}^c$. The rest of the proof is identical to the multicast case.

Appendix C
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 $(M_1, \ldots, M_N) = \mathbf{1}$ and

$\mathbf{L}_1 = \cdots = \mathbf{L}_b = \mathbf{1}$, where $\mathbf{L}_j = (L_{1j}, \ldots, L_{Nj})$. Then the decoder makes an error only if one or more of the following events occur:

$$\begin{aligned}
\mathcal{E}_1 = \{&(\hat{\mathbf{Y}}_{kj}(l_{kj}|1), \mathbf{U}_{kj}(1), \mathbf{Y}_{kj}) \notin \mathcal{T}_{\epsilon'}^{(n)} \\
&\text{for all } l_{kj} \text{for some } j \in [1:b], k \in [1:N]\} \\
\mathcal{E}_2 = \{&((\mathbf{X}_{kj}(1|1) : k \in \mathcal{S}_d), \mathbf{U}_{1j}(1), \ldots, \mathbf{U}_{Nj}(1), \\
&\hat{\mathbf{Y}}_{1j}(1|1), \ldots, \hat{\mathbf{Y}}_{Nj}(1|1), \mathbf{Y}_{dj}) \notin \mathcal{T}_{\epsilon}^{(n)} \\
&\text{for some } j \in [1:b]\} \\
\mathcal{E}_3 = \{&((\mathbf{X}_{kj}(m_k|l_{k,j-1}) : k \in \mathcal{S}_d), \mathbf{U}_{1j}(l_{1,j-1}), \ldots, \\
&\mathbf{U}_{Nj}(l_{N,j-1}), \hat{\mathbf{Y}}_{1j}(l_{1j}|l_{1,j-1}), \ldots, \hat{\mathbf{Y}}_{Nj}(l_{Nj}|l_{N,j-1}), \\
&\mathbf{Y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for all } j \in [1:b] \text{for some } \mathbf{l}^b, \mathbf{m} \neq \mathbf{1}\}.
\end{aligned}$$

Here, $\mathbf{m} = (m_k : k \in \mathcal{S}_d)$ with $m_d = 1$, and $\mathbf{l}_j = (l_{1j}, \ldots, l_{Nj})$ with $l_{dj} = 1$ for $j \in [1:b]$. Then the probability of error is upper bounded as

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

By the covering lemma and the union of events bound, $\mathsf{P}(\mathcal{E}_1) \to 0$ 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 Markov lemma and the union of events bound, the second term $\mathsf{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c)$ tends to zero as $n \to \infty$. For the third term, define the events

$$\begin{aligned}
&\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j) \\
&= \{((\mathbf{X}_{kj}(m_k|l_{k,j-1}) : k \in \mathcal{S}_d), \\
&\quad \mathbf{U}_{1j}(l_{1,j-1}), \ldots, \mathbf{U}_{Nj}(l_{N,j-1}), \hat{\mathbf{Y}}_{1j}(l_{1j}|l_{1,j-1}), \ldots, \\
&\quad \hat{\mathbf{Y}}_{Nj}(l_{Nj}|l_{N,j-1}), \mathbf{Y}_{dj}) \in \mathcal{T}_{\epsilon}^{(n)}\}
\end{aligned}$$

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

$$\mathsf{P}(\mathcal{E}_3) \leq \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)).$$

For each $\mathbf{l}^b$ and $j \in [2:b]$, define $\mathcal{S}_j(\mathbf{l}^b) = \{k \in [1:N] : l_{k,j-1} \neq 1\}$ and $\mathcal{T}(\mathbf{m}) = \{k \in \mathcal{S}_d : m_k \neq 1\}$. 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 Lemma 2

$$\begin{aligned}
&\mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\qquad \leq 2^{-n(I_1(\mathcal{S}(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) + I_2(\mathcal{S}(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) - \delta(\epsilon))}
\end{aligned}$$

where

$$\begin{aligned}
I_1(\mathcal{S}, \mathcal{T}) &= I(X((\mathcal{S} \cup \mathcal{T}) \cap \mathcal{S}_d), U(\mathcal{S}); \hat{Y}(\mathcal{S}^c), \\
&\qquad Y_d | X((\mathcal{S}^c \cap \mathcal{T}^c) \cap \mathcal{S}_d), U(\mathcal{S}^c)) \\
I_2(\mathcal{S}, \mathcal{T}) &= \sum_{k \in \mathcal{S}} I(\hat{Y}_k; \hat{Y}(\mathcal{S}^c \cup \{k' \in \mathcal{S} : k' < k\}), \\
&\qquad Y_d, X(\mathcal{S}_d), U^N | U_k).
\end{aligned}$$

Furthermore, by 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

$$\begin{aligned}
&\sum_{\mathbf{l}_{j-1}} 2^{-n(I_1(\mathcal{S}_j(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) + I_2(\mathcal{S}_j(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) - \delta(\epsilon))} \\
&\quad \leq \sum_{\substack{\mathcal{S} \subset [1:N]: \\ d \in \mathcal{S}^c}} 2^{-n(I_1(\mathcal{S}, \mathcal{T}(\mathbf{m})) + I_2(\mathcal{S}, \mathcal{T}(\mathbf{m})) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon))} \\
&\quad \leq 2^{N-1} 2^{-n(\min_{\mathcal{S}} (I_1(\mathcal{S}, \mathcal{T}(\mathbf{m})) + I_2(\mathcal{S}, \mathcal{T}(\mathbf{m})) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon)))}.
\end{aligned}$$

Denote $I_3(\mathcal{S}, \mathcal{T}) = I_1(\mathcal{S}, \mathcal{T}) + I_2(\mathcal{S}, \mathcal{T})$. Then

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3) \\
&= \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}^b} \prod_{j=2}^{b} \mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&= \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}_b} \sum_{\mathbf{l}^{b-1}} \prod_{j=2}^{b} \mathsf{P}(\tilde{\mathcal{E}}_j(\mathbf{m}, \mathbf{l}_{j-1}, \mathbf{l}_j)) \\
&\leq \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}_b} \sum_{\mathbf{l}^{b-1}} \prod_{j=2}^{b} 2^{-n(I_3(\mathcal{S}_j(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) - \delta(\epsilon))} \\
&= \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}_b} \prod_{j=2}^{b} \left(\sum_{\mathbf{l}_{j-1}} 2^{-n(I_3(\mathcal{S}_j(\mathbf{l}_{j-1}), \mathcal{T}(\mathbf{m})) - \delta(\epsilon))} \right) \\
&\leq \sum_{\mathbf{m} \neq \mathbf{1}} \sum_{\mathbf{l}_b} \prod_{j=2}^{b} \left(2^{N-1} 2^{-n(\min_{\mathcal{S}} (I_3(\mathcal{S}, \mathcal{T}(\mathbf{m})) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon)))} \right) \\
&\leq \sum_{\substack{\mathcal{T} \subseteq \mathcal{S}_d: \\ \mathcal{T} \neq \emptyset, d \in \mathcal{T}^c}} 2^{\sum_{k \in \mathcal{T}} nbR_k} 2^{\sum_{k \neq d} n\hat{R}_k} 2^{(N-1)(b-1)} \\
&\qquad \cdot 2^{n\left(-(b-1) \min_{\mathcal{S}} (I_3(\mathcal{S}, \mathcal{T}) - \sum_{k \in \mathcal{S}} \hat{R}_k - \delta(\epsilon))\right)}
\end{aligned}$$

where the minimum is over $\mathcal{S} \subset [1:N]$ such that $d \in \mathcal{S}^c$. Hence, $\mathsf{P}(\mathcal{E}_3)$ tends to zero as $n \to \infty$ if

$$\begin{aligned}
R(\mathcal{T}) < \frac{b-1}{b} &\left(\left(I_1(\mathcal{S}, \mathcal{T}) + I_2(\mathcal{S}, \mathcal{T}) - \sum_{k \in \mathcal{S}} \hat{R}_k \right) - \delta(\epsilon) \right) \\
&- \frac{1}{b} \sum_{k \neq d} \hat{R}_k
\end{aligned}$$

for all $\mathcal{S} \subset [1:N]$ and $\mathcal{T} \subseteq \mathcal{S}_d$ such that $d \in \mathcal{S}^c \cap \mathcal{T}^c$. 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

$$\begin{aligned}
R(\mathcal{T}) < I_1(\mathcal{S}, \mathcal{T}) + I_2(\mathcal{S}, \mathcal{T}) - \sum_{k \in \mathcal{S}} I(\hat{Y}_k; Y_k|U_k) \\
-(N-1)\delta(\epsilon') - \delta(\epsilon
\end{aligned}$$

for all $\mathcal{S} \subset [1:N]$ and $\mathcal{T} \subseteq \mathcal{S}_d$ such that $d \in \mathcal{S}^c \cap \mathcal{T}^c$. Furthermore, note that

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

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

$$R(\mathcal{T}) < I_1(\mathcal{S},\mathcal{T}) - I(\hat{Y}(\mathcal{S});Y(\mathcal{S})|X(\mathcal{S}_d),U^N,\hat{Y}(\mathcal{S}^c),Y_d) - (N-1)\delta(\epsilon') - \delta(\epsilon) \quad (28)$$

for all $\mathcal{S} \subset [1:N]$ and $\mathcal{T} \subseteq \mathcal{S}_d$ such that $d \in \mathcal{S}^c \cap \mathcal{T}^c$. Since for every $\mathcal{S} \subset [1:N]$, $d\in\mathcal{S}^c$ 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 (28), the set of inequalities can be further simplified to

$$\begin{aligned} R(\mathcal{T}) < {} & I(X(\mathcal{T}),U(\mathcal{S});\hat{Y}(\mathcal{S}^c),Y_d|X(\mathcal{T}^c),U(\mathcal{S}^c)) \\ & - I(Y(\mathcal{S});\hat{Y}(\mathcal{S})|X(\mathcal{S}_d),U^N,\hat{Y}(\mathcal{S}^c),Y_d) \\ & - (N-1)\delta(\epsilon') - \delta(\epsilon) \end{aligned} \quad (29)$$

for all $\mathcal{T},\mathcal{S} \subset [1:N]$ such that $\mathcal{S}\cap\mathcal{S}_d \subseteq \mathcal{T} \subseteq \mathcal{S}_d$ and $d\in\mathcal{S}^c$, 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 (29). The rest of the proof is identical to the multicast case.

Appendix D
Proof of Lemma 1

For $j \in [1:t]$ and $a_1 \ge a_2 \ge \cdots \ge a_t \ge 0$, let

$$b_j(a_1,\ldots,a_t) = \begin{cases} (\nu - a_j^{-1})^+, & \text{if } a_j > 0 \\ 0, & \text{otherwise} \end{cases}$$

where ν is chosen such that $\sum_{j=1}^t b_j(a_1,\ldots,a_t) = t$. Then, by diagonalizing A and water filling under the constraint $\mathrm{tr}(B) \le t$

$$\begin{aligned} \frac{|I+AB|}{|I+\gamma^{-1}A|} &\le \prod_{j=1}^t \frac{1+\lambda_j b_j^*}{1+\lambda_j\gamma^{-1}} \\ &\le \max_{\kappa\in[0:r]} \sup_{(a_1,\ldots,a_t)\in\mathcal{A}_\kappa} \prod_{j=1}^t \frac{1+a_jb_j}{1+a_j\gamma^{-1}} \end{aligned} \quad (30)$$

where $\lambda_1 \ge \cdots \ge \lambda_t \ge 0$ are the eigenvalues of A, $b_j^* = b_j(\lambda_1,\ldots,\lambda_t)$, $j\in[1:t]$, $b_j = b_j(a_1,\ldots,a_t)$, $j\in[1:t]$, and

$$\begin{aligned} \mathcal{A}_\kappa = \{(a_1,\ldots,a_t) : a_1 \ge \cdots \ge a_t \ge 0 \text{ such that} \\ |\{j\in[1:t] : b_j(a_1,\ldots,a_t) > 0\}| = \kappa\}. \end{aligned}$$

We further upper bound (30) as follows:

$$\begin{aligned} & \max_{\kappa\in[0:r]} \sup_{(a_1,\ldots,a_t)\in\mathcal{A}_\kappa} \prod_{j=1}^t \frac{1+a_jb_j}{1+a_j\gamma^{-1}} \\ & \overset{(a)}{=} \max_{\kappa\in[0:r]} \sup_{(a_1,\ldots,a_t)\in\mathcal{A}_\kappa} \prod_{j=1}^\kappa \frac{1+a_jb_j}{1+a_j\gamma^{-1}} \cdot \prod_{j=\kappa+1}^t \frac{1}{1+a_j\gamma^{-1}} \\ & \overset{(b)}{\le} \max_{\kappa\in[0:r]} \sup_{(a_1,\ldots,a_t)\in\mathcal{A}_\kappa} \prod_{j=1}^\kappa \frac{1+a_jb_j}{1+a_j\gamma^{-1}} \\ & \overset{(c)}{=} \max_{\kappa\in[0:r]} \sup_{(a_1,\ldots,a_t)\in\mathcal{A}_\kappa} \prod_{j=1}^\kappa \frac{a_j\nu}{1+a_j\gamma^{-1}} \\ & \overset{(d)}{=} \max_{\kappa\in[0:r]} \sup_{(a_1,\ldots,a_t)\in\mathcal{A}_\kappa} \gamma^\kappa \prod_{j=1}^\kappa \frac{\frac{t}{\kappa}+\frac{1}{\kappa}\sum_{l=1}^\kappa a_l^{-1}}{1+\gamma a_j^{-1}} \\ & = \max_{\kappa\in[0:r]} \sup_{(a_1,\ldots,a_t)\in\mathcal{A}_\kappa} \left(\frac{t\gamma}{\kappa}\right)^\kappa \prod_{j=1}^\kappa \frac{1+\frac{1}{\kappa}\sum_{l=1}^\kappa \frac{\kappa}{t}a_l^{-1}}{1+\left(\frac{t\gamma}{\kappa}\right)\left(\frac{\kappa}{t}a_j^{-1}\right)} \\ & \overset{(e)}{\le} \max\left\{1, \max_{\kappa\in[1:r]} \left(\frac{t\gamma}{\kappa}\right)^\kappa\right\} \\ & \overset{(f)}{\le} \sup_{\beta\in(0,r/t]} (\gamma/\beta)^{\beta t} \\ & \overset{(g)}{=} \begin{cases} e^{\gamma t/e}, & \text{if } r/t \ge \gamma/e \\ (\gamma t/r)^r, & \text{otherwise} \end{cases} \end{aligned}$$

where (a) follows since $b_j = 0$ for $j > \kappa$, (b) follows since $a_j \ge 0$, (c) follows since $b_j = \nu - a_j^{-1}$ for $j \le \kappa$, (d) follows since $\sum_{j=1}^\kappa (\nu - a_j^{-1}) = t$, (f) follows by relaxing $\frac{\kappa}{t}$ by $\beta \in (0,r/t]$, and (g) follows since $(\gamma/\beta)^\beta$ is maximized when $\beta = \min\{r/t,\gamma/e\}$. Finally, to justify step (e) we use the following.

Lemma 3: Suppose $0 \le w_j \le 1 + (1/\kappa)\sum_{l=1}^\kappa w_l$ for $j\in[1:\kappa]$. If $\gamma \ge e-1$, then

$$\prod_{j=1}^\kappa \frac{1+(1/\kappa)\sum_{l=1}^\kappa w_l}{1+\gamma w_j} \le 1.$$

Proof: Let $\psi(w_1,\ldots,w_\kappa) = \prod_{j=1}^\kappa \frac{1+(1/\kappa)\sum_{l=1}^\kappa w_l}{1+\gamma w_j}$. Note that it suffices to prove the lemma for $\gamma = e-1$, which we assume here. For every $s \ge 0$, ψ is convex on the simplex defined by $\sum_{j=1}^\kappa w_j = s$ and $w_j \ge 0$, $j\in[1:\kappa]$, since $\nabla^2 \ln\psi$ is positive semidefinite on the simplex. Therefore, ψ restricted to the intersection of $s = \sum_{j=1}^\kappa w_j$ and the polytope $\{(w_1,\ldots,w_\kappa) : 0 \le w_j \le 1+(1/\kappa)\sum_{l=1}^\kappa w_l,\ j\in[1:\kappa]\}$ is maximized at one of the corner points of the intersection. By symmetry, we only need to consider $w_1 \le w_2 \le \cdots \le w_\kappa$. Then, the corresponding corner points of the intersection have the form $(0,\ldots,0,w_r,w_{r+1},\ldots,w_\kappa)$ for some $1 \le r \le \kappa$, where $w_j = 1 + s/\kappa$ for $j\in[r+1:\kappa]$ and $0 \le w_r = rs/\kappa - (\kappa - r) < 1 + s/\kappa$. Hence

$$\psi(0,\ldots,0,w_r,\ldots,w_\kappa) = \frac{(1+s/\kappa)^\kappa}{(1+\gamma(rs/\kappa-(\kappa-r)))(1+\gamma(1+s/\kappa))^{\kappa-r}} \quad (31)$$

where the range of s is given by

$$\begin{cases} \frac{(\kappa-r)\kappa}{r} \le s < \frac{(\kappa-r+1)\kappa}{r-1}, & \text{if } 1 < r \le \kappa \\ (\kappa-1)\kappa \le s, & \text{if } r = 1. \end{cases}$$

Let $\phi(s,\kappa,r)$ denote the right-hand side (RHS) of (31). If $r = 1$ or $\kappa = 1$, then it is easy to verify that $\phi(s,\kappa,r)$ is strictly decreasing in s and is thus maximized at $s = (\kappa-1)\kappa$, which corresponds to $w_r = 0$. For the case $1 < r \le \kappa$, let s be parametrized by $\tau \in [0,1)$ such that $s(\tau) = \frac{(\kappa-r+\tau)\kappa}{r-\tau}$. Then

$$\phi(s(\tau),\kappa,r) = \frac{\kappa^\kappa}{(r-\tau)^{r-1}(r-\tau+\gamma\kappa\tau)(r-\tau+\gamma\kappa)^{\kappa-r}}.$$

Hence

$$\begin{aligned}
&\frac{\phi(s(0),\kappa,r)}{\phi(s(\tau),\kappa,r)} \\
&\quad = \left(1-\frac{\tau}{r}\right)^{r-1}\left(1+\frac{\gamma\kappa-1}{r}\tau\right)\left(1-\frac{\tau}{r+\gamma\kappa}\right)^{\kappa-r} \\
&\quad \overset{(a)}{\ge} \left(1-\frac{\tau}{r}\right)^{r-1}\left(1+\frac{\gamma\kappa-1}{r}\tau\right)\left(1-\frac{\tau(\kappa-r)}{r+\gamma\kappa}\right) \\
&\quad \overset{(b)}{\ge} \left(1-\frac{\tau}{r}\right)^{r-1}\left(1+\frac{\gamma r-1}{r}\tau\right) \\
&\quad \overset{(c)}{\ge} 1
\end{aligned}$$

where (a) follows since $(1-x)^t \ge 1-tx$ for $0 \le x \le 1$ and $t \ge 1$, (b) follows since the RHS of (a) is monotonically increasing in κ for given r and τ, and is thus minimized at $\kappa = r$, and (c) follows since $(1-x)^t(1+((e-1)(t+1)-1)x) \ge 1$ for $0 \le x \le 1/(t+1)$ and $t \ge 1$. This implies that $\phi(s,\kappa,r)$ is maximized when $s = \frac{(\kappa-r)\kappa}{r}$, which again corresponds to $w_r = 0$. Therefore, for all $1 \le r \le \kappa$, it suffices to assume $w_r = 0$. Finally

$$\begin{aligned}
\psi\left(0,\ldots,0,\frac{\kappa}{r},\ldots,\frac{\kappa}{r}\right) &= \frac{(\kappa/r)^\kappa}{(1+\gamma(\kappa/r))^{\kappa-r}} \\
&= \left[p^{-p}(e-(1-p))^{-(1-p)}\right]^\kappa \\
&\overset{(a)}{\le} 1
\end{aligned}$$

where $p = r/\kappa$ and (a) holds for all $p \in [0,1]$. This completes the proof of Lemma 3, which, in turn, completes the proof of Lemma 1. ■

Appendix E
Comparison to a Previous Extension of Compress-Forward to Multicast Networks

For a single-message multicast network with source node 1 and destination nodes $\mathcal{D} \subseteq [2:N]$, a hybrid scheme proposed by Kramer, Gastpar, and Gupta [13, Theorem 3] achieves the capacity lower bound

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

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

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

for all $\mathcal{T} \subset [2:N]$, all partitions $\{\mathcal{K}_t\}_{t=1}^T$ of $\mathcal{T}$, 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 $[2:N]$. This hybrid coding scheme uses an extension of the original compress-forward scheme for the relay channel as well as decoding of the compression indices at the relays.

In the following subsections, we compare noisy network coding with the hybrid scheme. First we show that noisy network coding always outperforms the compress-forward part of the scheme (without compression index decoding at the relays), which corresponds to $U_k = \emptyset$ in (32). Then we consider a layered network and show that noisy network coding achieves a positive rate while the hybrid scheme in (32) achieves zero rate.

1) Comparison to the Compress-Forward Part: The compress-forward part of (32) without decoding yields the capacity lower bound

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

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

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

for all $\mathcal{T} \subseteq [2:N]$ and $\mathcal{T}^c = [2:N]\backslash\mathcal{T}$. This bound is identical to (32) and (33) with $U_j = \emptyset$, $j \in [2:N]$. By similar steps to those in [14, Appendix C] and some algebra, lower bound (34) can be upper bounded as

$$\begin{aligned}
R^* &\le \max \min_{d\in\mathcal{D}} \min_{\mathcal{T}\subseteq[2:N]} I(X_1;\hat{Y}_2^N,Y_d|X_2^N) \\
&\qquad + I(X(\mathcal{T});Y_d|X(\mathcal{T}^c),X_d) \\
&\qquad - I(\hat{Y}(\mathcal{T});Y(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y_d) \\
&\qquad - \sum_{k\in\mathcal{T}} I(\hat{Y}_k;X_2^N|X_k) \\
&= \max \min_{d\in\mathcal{D}} \min_{\mathcal{T}\subseteq[2:N]} I(X_1,X(\mathcal{T});\hat{Y}(\mathcal{T}^c),Y_d|X(\mathcal{T}^c),X_d) \\
&\qquad - I(\hat{Y}(\mathcal{T});Y(\mathcal{T})|X^N,\hat{Y}(\mathcal{T}),Y_d) \\
&\qquad - I(X(\mathcal{T});\hat{Y}(\mathcal{T}^c)|Y_d,X(\mathcal{T}^c),X_d) \\
&\qquad - \sum_{k\in\mathcal{T}} I(\hat{Y}_k;X_2^N|X_k)
\end{aligned} \tag{35}$$

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 from

$$\begin{aligned}
&I(X_1;\hat{Y}_2^N,Y_d|X_2^N)+I(X(\mathcal{T});Y_d|X(\mathcal{T}^c),X_d)\\
&\quad -I(\hat{Y}(\mathcal{T});Y(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y_d) &(36)\\
&=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)\\
&\quad +I(X(\mathcal{T});Y_d|X(\mathcal{T}^c),X_d)\\
&\quad -I(\hat{Y}(\mathcal{T});Y(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y_d) &(37)\\
&=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)\\
&\quad +I(X(\mathcal{T});\hat{Y}(\mathcal{T}^c),Y_d|X(\mathcal{T}^c),X_d)\\
&\quad -I(X(\mathcal{T});\hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c),Y_d,X_d)\\
&\quad -I(\hat{Y}(\mathcal{T});Y(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y_d) &(38)\\
&=I(X_1,X(\mathcal{T});\hat{Y}(\mathcal{T}^c),Y_d|X(\mathcal{T}^c),X_d)\\
&\quad +I(X_1;\hat{Y}(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y_d)\\
&\quad -I(X(\mathcal{T});\hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c),Y_d,X_d)\\
&\quad -I(\hat{Y}(\mathcal{T});Y(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y_d) &(39)\\
&=I(X_1,X(\mathcal{T});\hat{Y}(\mathcal{T}^c),Y_d|X(\mathcal{T}^c),X_d)\\
&\quad +I(X_1,Y(\mathcal{T});\hat{Y}(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y_d)\\
&\quad -I(Y(\mathcal{T});\hat{Y}(\mathcal{T})|X_1,X_2^N,\hat{Y}(\mathcal{T}^c),Y_d)\\
&\quad -I(X(\mathcal{T});\hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c),Y_d,X_d)\\
&\quad -I(\hat{Y}(\mathcal{T});Y(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y_d) &(40)\\
&=I(X_1,X(\mathcal{T});\hat{Y}(\mathcal{T}^c),Y_d|X(\mathcal{T}^c),X_d)\\
&\quad +I(X_1;\hat{Y}(\mathcal{T})|X_2^N,\hat{Y}(\mathcal{T}^c),Y(\mathcal{T}),Y_d)\\
&\quad -I(Y(\mathcal{T});\hat{Y}(\mathcal{T})|X_1,X_2^N,\hat{Y}(\mathcal{T}^c),Y_d)\\
&\quad -I(X(\mathcal{T});\hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c),Y_d,X_d) &(41)\\
&=I(X_1,X(\mathcal{T});\hat{Y}(\mathcal{T}^c),Y_d|X(\mathcal{T}^c),X_d)\\
&\quad -I(Y(\mathcal{T});\hat{Y}(\mathcal{T})|X_1,X_2^N,\hat{Y}(\mathcal{T}^c),Y_d)\\
&\quad -I(X(\mathcal{T});\hat{Y}(\mathcal{T}^c)|X(\mathcal{T}^c),Y_d,X_d) &(42)
\end{aligned}$$

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, the inner bound in Theorem 1 can be specialized to the single-message case by setting $Q=\emptyset$ and $R_2=\cdots=R_N=0$ to yield

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

where the maximum is over all pmfs $\prod_{k=1}^N p(x_k)p(\hat{y}_k|y_k,x_k)$ and $\mathcal{T}^c=[2:N]\backslash\mathcal{T}$. 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 $\mathcal{T}\subseteq[2:N]$.

2) Comparison to the General Hybrid Scheme: Consider an N-node DMN with single source node 1 and single destination node N. Assume that $Y_1=X_N=\emptyset$. Suppose the network is *layered*, that is, the channel pmf has the form

$$p(y_2,\ldots,y_N|x_1,\ldots,x_{N-1})=\prod_{i=2}^L p(y(\mathcal{L}_i)|x(\mathcal{L}_{i-1}))$$

where $\mathcal{L}_1,\ldots,\mathcal{L}_L$ partitions $[1:N]$ such that $\mathcal{L}_1=\{1\}$ and $\mathcal{L}_L=\{N\}$. In the following, we assume that $L\ge4$ and show that the rate achievable by the Kramer-Gastpar–Gupta hybrid scheme in (32) is zero for this case.

First, note that the constraint in (33) corresponding to $\mathcal{T}=\mathcal{K}_1=\mathcal{L}_2$, and $r(1)=N$ is given by

$$\begin{aligned}
&I(\hat{Y}(\mathcal{L}_2);Y(\mathcal{L}_2)|U_2^{N-1},X_2^{N-1},\hat{Y}(\mathcal{L}_2^c),Y_N)\\
&\le I(X(\mathcal{L}_2);Y_N|U(\mathcal{L}_2),X(\mathcal{L}_2^c))+I(U(\mathcal{L}_2);Y_N|U(\mathcal{L}_2^c))\\
&\quad -\sum_{k\in\mathcal{L}_2}I(\hat{Y}_k;X_2^{N-1}|U_2^{N-1},X_k). &(43)
\end{aligned}$$

Now the RHS of (43) is zero since

$$\begin{aligned}
I(X(\mathcal{L}_2);Y_N|U(\mathcal{L}_2),X(\mathcal{L}_2^c))&=I(X(\mathcal{L}_2);Y_N|X(\mathcal{L}_{L-1}))\\
&=0
\end{aligned}$$

which follows by the Markov relationship $Y_N\to X(\mathcal{L}_{L-1})\to(X(\mathcal{L}_{L-1}^c),U(\mathcal{L}_2))$ for $L\ge4$, and

$$I(U(\mathcal{L}_2);Y_N|U(\mathcal{L}_2^c))=I(U(\mathcal{L}_2);Y_N|U(\mathcal{L}_{L-1}))=0$$

which follows by $Y_N\to U(\mathcal{L}_{L-1})\to U(\mathcal{L}_{L-1}^c)$. Thus, (43) implies that a feasible pmf for the hybrid scheme must satisfy

$$I(\hat{Y}(\mathcal{L}_2);Y(\mathcal{L}_2)|U_2^{N-1},X_2^{N-1},\hat{Y}(\mathcal{L}_2^c),Y_N)=0. \tag{44}$$

By (32), an achievable rate of the hybrid scheme satisfies

$$\begin{aligned}
R&\le I(X_1;\hat{Y}_2^{N-1},Y_N|U_2^{N-1},X_2^{N-1})\\
&=I(X_1;\hat{Y}(\mathcal{L}_2)|U_2^{N-1},X_2^{N-1},\hat{Y}(\mathcal{L}_2^c),Y_N)\\
&\quad +I(X_1;\hat{Y}(\mathcal{L}_2^c),Y_N|U_2^{N-1},X_2^{N-1})\\
&\le I(Y(\mathcal{L}_2),X_1;\hat{Y}(\mathcal{L}_2)|U_2^{N-1},X_2^{N-1},\hat{Y}(\mathcal{L}_2^c),Y_N)\\
&\quad +I(X_1;\hat{Y}(\mathcal{L}_2^c),Y_N|U_2^{N-1},X_2^{N-1})\\
&=I(Y(\mathcal{L}_2);\hat{Y}(\mathcal{L}_2)|U_2^{N-1},X_2^{N-1},\hat{Y}(\mathcal{L}_2^c),Y_N)\\
&\quad +I(X_1;\hat{Y}(\mathcal{L}_2)|Y(\mathcal{L}_2),U_2^{N-1},X_2^{N-1},\hat{Y}(\mathcal{L}_2^c),Y_N)\\
&\quad +I(X_1;\hat{Y}(\mathcal{L}_2^c),Y_N|U_2^{N-1},X_2^{N-1})\\
&\overset{(a)}{=}I(X_1;\hat{Y}(\mathcal{L}_2^c),Y_N|U_2^{N-1},X_2^{N-1})\\
&\le I(X_1;Y(\mathcal{L}_2^c),\hat{Y}(\mathcal{L}_2^c),Y_N|U_2^{N-1},X_2^{N-1})\\
&=I(X_1;Y(\mathcal{L}_2^c),Y_N|U_2^{N-1},X_2^{N-1})\\
&\overset{(b)}{=}0
\end{aligned}$$

where (a) follows by (44) and the fact that

$$\begin{aligned}\hat{Y}(\mathcal{L}_2) &\to (Y(\mathcal{L}_2), U(\mathcal{L}_2), X(\mathcal{L}_2))\\ &\to (X_1, X(\mathcal{L}_2^c), U(\mathcal{L}_2^c), \hat{Y}(\mathcal{L}_2^c), Y_N)\end{aligned}$$

and (b) follows since $(Y(\mathcal{L}_2^c), Y_N) \to (U_2^{N-1}, X_2^{N-1}) \to X_1$. Thus the hybrid scheme achieves zero rate for every layered network with 4 or more layers.

In comparison, it is easy to find a layered-network example for which noisy network coding achieves a positive rate (and the capacity itself). For instance, consider a 4-node layered network $Y_2 = X_1, Y_3 = X_2, Y_4 = X_3$, where X_1, X_2, X_3 are binary, $\mathcal{D} = \{4\}$, and $R_2 = R_3 = R_4 = 0$. The capacity of this network is 1 bit/transmission, and by (5) it is achieved by noisy network coding.

APPENDIX F
COMPARISON TO PREVIOUS SCHEMES FOR THE TWO-WAY RELAY CHANNEL

1) Comparison to Compress-Forward: We show that the compress-forward inner bound is included in the noisy network coding inner bound for the Gaussian two-way relay channel. Fix the channel gains and power constraint. Denote

$$\begin{aligned}R_{11}(\sigma^2) &= \mathsf{C}\left(\frac{g_{31}^2 P + (1+\sigma^2) g_{21}^2 P}{1+\sigma^2}\right)\\ R_{12}(\sigma^2) &= \mathsf{C}(g_{21}^2 P + g_{23}^2 P) - \mathsf{C}(1/\sigma^2)\\ R_{21}(\sigma^2) &= \mathsf{C}\left(\frac{g_{32}^2 P + (1+\sigma^2) g_{12}^2 P}{1+\sigma^2}\right)\\ R_{22}(\sigma^2) &= \mathsf{C}(g_{12}^2 P + g_{13}^2 P) - \mathsf{C}(1/\sigma^2).\end{aligned}$$

Then, by using the above notation, we can write the noisy network coding inner bound (23) by

$$\begin{aligned}R_1 &< \min\{R_{11}(\sigma^2), R_{12}(\sigma^2)\}\\ R_2 &< \min\{R_{21}(\sigma^2), R_{22}(\sigma^2)\}\end{aligned}$$

for some $\sigma^2 > 0$. On the other hand, the compress-forward inner bound (21) can be written by

$$\begin{aligned}R_1 &< R_{11}(\sigma^2)\\ R_2 &< R_{21}(\sigma^2)\end{aligned}$$

for some $\sigma^2 \ge \max\{\sigma_1^2, \sigma_2^2\}$, where

$$\begin{aligned}\sigma_1^2 &= \frac{1 + g_{21}^2 P + g_{31}^2 P}{\min\{g_{23}^2, g_{13}^2\} P}\\ \sigma_2^2 &= \frac{1 + g_{12}^2 P + g_{32}^2 P}{\min\{g_{23}^2, g_{13}^2\} P}.\end{aligned}$$

Note that $R_{k1}(\sigma^2)$, $k = 1, 2$, is nonincreasing and $R_{k2}(\sigma^2)$, $k = 1, 2$, is nondecreasing since

$$\begin{aligned}\frac{\partial R_{k1}(\sigma^2)}{\partial \sigma^2} &\le 0\\ \frac{\partial R_{k2}(\sigma^2)}{\partial \sigma^2} &\ge 0.\end{aligned}$$

Furthermore $R_{11}(\sigma_{e1}^2) = R_{12}(\sigma_{e1}^2)$ and $R_{21}(\sigma_{e2}^2) = R_{22}(\sigma_{e2}^2)$, where $\sigma_{e1}^2 = (1 + g_{21}^2 P + g_{31}^2 P)/(g_{23}^2 P)$ and $\sigma_{e2}^2 = (1 + g_{12}^2 P + g_{32}^2 P)/(g_{13}^2 P)$. Hence, $R_{11}(\sigma^2) \le R_{12}(\sigma^2)$ and $R_{21}(\sigma^2) \le R_{22}(\sigma^2)$ for $\sigma^2 \ge \max\{\sigma_{e1}^2, \sigma_{e2}^2\}$. Since $\max\{\sigma_1^2, \sigma_2^2\} \ge \max\{\sigma_{e1}^2, \sigma_{e2}^2\}$, the noisy network coding inner bound (23) is equal to the compress-forward inner bound (21) for $\sigma^2 \ge \max\{\sigma_1^2, \sigma_2^2\}$, i.e., the range of σ^2 that satisfies the compress-forward constraint, and the noisy network coding inner bound is tighter than the compress-forward inner bound for $\sigma^2 < \max\{\sigma_1^2, \sigma_2^2\}$, in which case compress-forward achieves zero rate.

2) Comparison of Gap From Capacity: We first show that noisy network coding achieves within 1 bit for individual rates and 1.5 bit for the sum rate from the cutset bound. We begin by presenting some properties of the cutset bound and the noisy network coding inner bound.

Consider a looser version of the cutset bound [30] for the two-way relay channel

$$\begin{aligned}R_1 \le C_{u1} = \max_{-1\le\rho_1\le 1} \min\{&\mathsf{C}\left((1-\rho_1^2)(g_{31}^2 P + g_{21}^2 P)\right),\\ &\mathsf{C}(g_{23}^2 P + g_{21}^2 P + 2 g_{21} g_{23} \rho_1 P)\}\\ R_2 \le C_{u2} = \max_{-1\le\rho_2\le 1} \min\{&\mathsf{C}\left((1-\rho_2^2)(g_{32}^2 P + g_{12}^2 P)\right),\\ &\mathsf{C}(g_{13}^2 P + g_{12}^2 P + 2 g_{12} g_{13} \rho_2 P)\}.\end{aligned}$$

Note that if $g_{31} \le g_{23}$ then

$$C_{u1} = \mathsf{C}\left(g_{31}^2 P + g_{21}^2 P\right) \tag{45}$$

and if $g_{31} \ge g_{23}$ then C_{u1} satisfies

$$\begin{aligned}C_{u1} &= \mathsf{C}\left(g_{21}^2 P + g_{23}^2 P + \frac{2 a_1 P}{g_{21}^2 + g_{31}^2}\right)\\ &\le \mathsf{C}\left(g_{21}^2 P + g_{23}^2 P\right) + 1/2\end{aligned} \tag{46}$$

where $a_1 = g_{21} g_{23} g_{31} \sqrt{g_{21}^2 - g_{23}^2 + g_{31}^2} - g_{21}^2 g_{23}^2$. Similarly, if $g_{32} \le g_{13}$ then

$$C_{u2} = \mathsf{C}\left(g_{32}^2 P + g_{12}^2 P\right) \tag{47}$$

and if $g_{32} \ge g_{13}$ then C_{u2} satisfies

$$\begin{aligned}C_{u2} &= \mathsf{C}\left(g_{12}^2 P + g_{13}^2 P + \frac{2 a_2 P}{g_{12}^2 + g_{32}^2}\right)\\ &\le \mathsf{C}\left(g_{12}^2 P + g_{13}^2 P\right) + 1/2\end{aligned} \tag{48}$$

where $a_2 = g_{12} g_{13} g_{32} \sqrt{g_{12}^2 - g_{13}^2 + g_{32}^2} - g_{12}^2 g_{13}^2$. On the other hand, by the choice of $\sigma^2 = 1$, noisy network coding inner bound is given by the set of rate pairs (R_1, R_2) that satisfies

$$R_1 < \min\{\mathsf{C}(g_{31}^2 P + g_{21}^2 P), \mathsf{C}(g_{23}^2 P + g_{21}^2 P)\} - 1/2 \tag{49}$$
$$R_2 < \min\{\mathsf{C}(g_{32}^2 P + g_{12}^2 P), \mathsf{C}(g_{13}^2 P + g_{12}^2 P)\} - 1/2. \tag{50}$$

Hence, by (45)–(50), the individual rates of the noisy network coding inner bound are within 1 bit of those in the cutset bound.

Next we bound the gap for the sum rate. Consider the cutset bound on the sum rate

$$C_u = C_{u1} + C_{u2}$$

and the sum rate of the noisy network coding inner bound

$$C_l = \max_{\sigma^2>0}(\min\{R_{11}(\sigma^2), R_{12}(\sigma^2)\} + \min\{R_{21}(\sigma^2), R_{22}(\sigma^2)\}).$$

We consider the following cases:

1) $g_{31} \le g_{23}$ and $g_{32} \le g_{13}$: By (45) and (47), $C_u = \mathsf{C}(g_{31}^2P+g_{21}^2P)+\mathsf{C}(g_{32}^2P+g_{12}^2P)$. On the other hand, by (49) and (50), $C_l \ge \mathsf{C}(g_{31}^2P+g_{21}^2P)+\mathsf{C}(g_{32}^2P+g_{12}^2P)-1$. Hence, the gap is upper bounded by 1 bit.
2) $g_{31} \ge g_{23}$ and $g_{32} \ge g_{13}$: First suppose $\sigma_{e1}^2 \le \sigma_{e2}^2$. By taking $\sigma^2 = \sigma_{e1}^2$ for the noisy network coding inner bound

$$\begin{aligned} C_u - C_l &= C_{u1} + C_{u2} - R_{11}(\sigma_{e1}^2) - R_{22}(\sigma_{e1}^2) \\ &= \left(C_{u1} - R_{11}(\sigma_{e1}^2)\right) + \left(C_{u2} - R_{22}(\sigma_{e1}^2)\right). \end{aligned}$$

Note that C_{u1} is the cutset upper bound and $R_{11}(\sigma_{e1}^2)$ is the compress-forward rate for the 3 node one-way relay channel when node 1 is the source and node 2 is the destination. Since compress-forward achieves uniformly within 1/2 bit for the 3-node one-way relay channel [31], the first term is bounded as

$$C_{u1} - R_{11}(\sigma_{e1}^2) \le 1/2.$$

To upper bound the second term, first note that

$$\begin{aligned} \mathsf{C}(1/\sigma_{e1}^2) &= \mathsf{C}\left(\frac{g_{23}^2P}{1+g_{21}^2P+g_{31}^2P}\right) \\ &= \frac{1}{2}\log\left(1+\frac{g_{23}^2P}{1+g_{21}^2P+g_{31}^2P}\right) \le 1/2 \end{aligned}$$

since $g_{31} \ge g_{23}$ by the standing assumption. Hence

$$\begin{aligned} R_{22}(\sigma_{e1}^2) &= \mathsf{C}(g_{13}^2P+g_{12}^2P) - \mathsf{C}(1/\sigma_{e1}^2) \\ &\ge \mathsf{C}(g_{13}^2P+g_{12}^2P) - 1/2. \end{aligned} \tag{51}$$

Combining (48) and (51), we have

$$C_{u2} - R_{22}(\sigma_{e1}^2) \le 1$$

and the sum rate gap is within 1.5 bits.

Next, suppose $\sigma_{e1}^2 \ge \sigma_{e1}^2$. By symmetry, the gap is again within 1.5 bits by taking $\sigma^2 = \sigma_{e2}^2$ in the noisy network coding inner bound.
3) $g_{13} \le g_{32}$ and $g_{23} \ge g_{31}$: Take $\sigma^2 = 1$ for the noisy network coding inner bound. Then, by (45), (48), (49), and (50), it can be shown that the sum rate gap is within 1.5 bits.
4) $g_{13} \ge g_{32}$ and $g_{23} \le g_{31}$: By switching the roles of users 1 and 2 in the third case above, the gap is again within 1.5 bits for this case.

Since we have covered all possible cases, the sum rate gap is within 1.5 bits, regardless of the channel gains and power constraint.

We now show that decode-forward, amplify-forward, and compress-forward have arbitrarily large gap from the cutset bound.

1) Decode-forward: Suppose $g_{12} = g_{21} = 0$, $g_{31} = g_{32} = g_{13} = g_{23} = 1$. Then, the sum rate of the cutset bound is $2\mathsf{C}(P)$. On the other hand, the decode-forward sum rate is bounded as

$$R_1 + R_2 < \mathsf{C}(2P).$$

Hence, the sum rate gap for decode-forward becomes arbitrarily large as $P \to \infty$.
2) Amplify-forward and compress-forward: Suppose $g_{12} = g_{21} = 0$, $g_{31} = g_{13} = g_{23} = 1$, $g_{32} = \sqrt{P}$ and $P \ge 1$. Then, the sum rate of the cutset bound is $2\mathsf{C}(P)$. On the other hand, the amplify-forward inner bound is simplified to the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< \mathsf{C}\left(\frac{P^2}{1+2P+P^2}\right) \\ R_2 &< \mathsf{C}\left(\frac{P^3}{1+2P+P^2}\right). \end{aligned}$$

Similarly, the compress-forward inner bound is simplified to the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< \mathsf{C}\left(\frac{P^2}{1+P+P^2}\right) \\ R_2 &< \mathsf{C}\left(\frac{P^3}{1+P+P^2}\right). \end{aligned}$$

Compared to the cutset bound, these bounds have arbitrarily large gap as $P \to \infty$.

Appendix G
Comparison to Previous Schemes for the Interference Relay Channel

We show that the inner bound in Theorem 3 is tighter than both hash-forward and compress-forward inner bounds in [17], [18]. Fix the channel gains and power constraint P. Define

$$\begin{aligned} R_{11}(\sigma^2) &= \mathsf{C}\left(\frac{g_{41}^2P}{g_{42}^2P+1}\right) + R_0 - \mathsf{C}\left(\frac{(g_{32}^2+g_{42}^2)P+1}{(g_{42}^2P+1)\sigma^2}\right) \\ R_{21}(\sigma^2) &= \mathsf{C}\left(\frac{g_{52}^2P}{g_{51}^2P+1}\right) + R_0 - \mathsf{C}\left(\frac{(g_{31}^2+g_{51}^2)P+1}{(g_{51}^2P+1)\sigma^2}\right) \\ R_{12}(\sigma^2) &= \mathsf{C}\left(\frac{(g_{31}^2+(1+\sigma^2)g_{41}^2)P+a_1^2P^2}{1+\sigma^2+(g_{32}^2+(1+\sigma^2)g_{42}^2)P}\right) \\ R_{22}(\sigma^2) &= \mathsf{C}\left(\frac{(g_{32}^2+(1+\sigma^2)g_{52}^2)P+a_2^2P^2}{1+\sigma^2+(g_{31}^2+(1+\sigma^2)g_{51}^2)P}\right) \end{aligned}$$

where $a_1 = g_{32}g_{41} - g_{42}g_{31}$ and $a_2 = g_{31}g_{52} - g_{51}g_{32}$. Then, the hash-forward inner bound is given by

$$\begin{aligned} R_1 &< R_{11}(\sigma^2) \\ R_2 &< R_{21}(\sigma^2) \end{aligned}$$

for some $0 < \sigma^2 \le \min\{\sigma_1^2, \sigma_2^2\}$, where

$$\sigma_1^2 = \frac{1}{2^{2R_0}-1}\frac{(g_{31}g_{42}-g_{32}g_{41})^2P^2+a_1}{(g_{41}^2P+g_{42}^2P+1)}$$
$$\sigma_2^2 = \frac{1}{2^{2R_0}-1}\frac{(g_{31}g_{52}-g_{32}g_{51})^2P^2+a_2}{(g_{51}^2P+g_{52}^2P+1)}$$

and

$$a_1 = (g_{31}^2+g_{41}^2)P+(g_{32}^2+g_{42}^2)P+1$$
$$a_2 = (g_{31}^2+g_{51}^2)P+(g_{32}^2+g_{52}^2)P+1.$$

Using the above notation, compress-forward inner bound can be written as

$$R_1 < R_{12}(\sigma^2)$$
$$R_2 < R_{22}(\sigma^2)$$

for some $\sigma^2 \ge \max\{\sigma_1^2, \sigma_2^2\}$. Similarly, the inner bound in Theorem 3 can be written as

$$R_1 < \min\{R_{11}(\sigma^2), R_{12}(\sigma^2)\}$$
$$R_2 < \min\{R_{21}(\sigma^2), R_{22}(\sigma^2)\}$$

for some $\sigma^2 > 0$. Note that $R_{k2}(\sigma^2)$, $k=1,2$, is nonincreasing in σ^2 and $R_{k1}(\sigma^2)$, $k=1,2$, is nondecreasing in σ^2, since

$$\frac{\partial R_{k2}(\sigma^2)}{\partial \sigma^2} \le 0 \tag{52}$$
$$\frac{\partial R_{k1}(\sigma^2)}{\partial \sigma^2} \ge 0. \tag{53}$$

From (52), the sum rate of the compress-forward scheme is maximized by $\sigma^2 = \max\{\sigma_1^2, \sigma_2^2\}$, and from (53), the sum rate of the hash-forward scheme is maximized by $\sigma^2 = \min\{\sigma_1^2, \sigma_2^2\}$. Furthermore, it can be easily shown that $R_{12}(\sigma_1^2) = R_{11}(\sigma_1^2)$ and $R_{22}(\sigma_2^2) = R_{21}(\sigma_2^2)$. Hence, $R_{k1}(\sigma^2) \le R_{k2}(\sigma^2)$ in $\sigma^2 < \sigma_k^2$, $k=1,2$, and $R_{k1}(\sigma^2) \ge R_{k2}(\sigma^2)$ in $\sigma^2 > \sigma_k^2$, $k=1,2$. Therefore, the noisy network coding inner bound is equal to the compress-forward inner bound for $\sigma^2 \ge \max\{\sigma_1^2, \sigma_2^2\}$, i.e., the range of σ^2 that satisfies the compress-forward constraint, and is equal to the hash-forward inner bound for $\sigma^2 \le \min\{\sigma_1^2, \sigma_2^2\}$, i.e., the range of σ^2 that satisfies the hash-forward constraint. Finally, when $\min\{\sigma_1^2, \sigma_2^2\} < \sigma^2 < \max\{\sigma_1^2, \sigma_2^2\}$, the noisy network coding inner bound is tighter than both bounds.

ACKNOWLEDGMENT

The authors would like to thank H.-I. Su for his comments on earlier drafts of the paper. They would also like to thank the anonymous reviewers for many valuable comments that helped improve the presentation of this paper.

REFERENCES

[1] A. El Gamal, "On information flow in relay networks," in *Proc. IEEE Nat. Telecom Conf.*, 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] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] B. Rankov and A. Wittneben, "Achievable rate region for the two-way relay channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 2006.

[16] 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, Jun. 2007, pp. 1101–1105.

[17] B. Djeumou, E. V. Belmega, and S. Lasaulce, Interference Relay Channels—Part I: Transmission Rates 2009 [Online]. Available: http://arxiv.org/abs/0904.2585/

[18] P. Razaghi and W. Yu, "Universal relaying for the interference channel," in *Proc. UCSD Inf. Theory Appl. Workshop*, La Jolla, CA, 2010.

[19] S. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, 2011, to be published.

[20] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han–Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, Jul. 2008.

[21] C. Nair and A. El Gamal, "The capacity region of a class of three-receiver broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4479–4493, Oct. 2009.

[22] E. Perron, "Information-Theoretic Secrecy for Wireless Networks," Ph.D., École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.

[23] A. El Gamal and Y.-H. Kim, Lecture Notes on Network Information Theory 2010 [Online]. Available: http://arxiv.org/abs/1001.3404/

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

[25] T. M. Cover and Y.-H. Kim, "Capacity of a class of deterministic relay channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 591–595.

[26] 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.

[27] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Layered noisy network coding," in *Proc. IEEE Wireless Network Coding Workshop*, Boston, MA, Aug. 2010.

[28] 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.

[29] 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.

[30] S. Avestimehr, A. Sezgin, and D. Tse, "Approximate capacity of the two-way relay channel: A deterministic approach," in *Proc. 46th Annual Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 2008.

[31] W. Chang, S.-Y. Chung, and Y. H. Lee, Gaussian Relay Channel Capacity to Within a Fixed Number of Bits 2010 [Online]. Available: http://arxiv.org/abs/1011.5065/

Sung Hoon Lim (S'08) received the B.S. degree with honors in electrical and computer engineering from Korea University in 2005, and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST) in 2007.

He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering from KAIST. His research interests are in information theory, communication systems, and data compression.

Young-Han Kim (S'99–M'06) received the B.S. degree with honors in electrical engineering from Seoul National University, Seoul, Korea, in 1996. He received the M.S. degrees in electrical engineering and statistics and the Ph.D. degree in electrical engineering, both from Stanford University, Stanford, CA, in 2001, 2006, and 2006, respectively.

In July 2006, he joined the University of California, San Diego, where he is an Assistant Professor of Electrical and Computer Engineering. His research interests are in statistical signal processing and information theory, with applications in communication, control, computation, networking, data compression, and learning.

Dr. Kim is a recipient of the 2008 NSF Faculty Early Career Development (CAREER) Award and the 2009 U.S.-Israel Binational Science Foundation Bergmann Memorial Award.

Abbas El Gamal (S'71–M'73–SM'83–F'00) received the B.Sc. (honors) degree in electrical engineering from Cairo University in 1972 and the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1977 and 1978, respectively.

From 1978 to 1980, he was an Assistant Professor of Electrical Engineering at the University of Southern California (USC), Los Angeles. He has been on the Stanford University faculty since 1981, where he is currently the Hitachi America Professor in the School of Engineering. His research interest and contributions are in the areas of network information theory, digital imaging, and integrated circuit design. He has authored or coauthored more than 200 papers and 30 patents in these areas.

Dr. El Gamal is on the Board of Governors of the IEEE Information Theory Society.

Sae-Young Chung (S'89–M'00–SM'07) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1990 and 1992, respectively. He received the Ph.D. degree in electrical engineering and computer science at the Massachusetts Institute of Technology (MIT), Cambridge, in 2000.

From June to August 1998 and from June to August 1999, he was with Lucent Technologies. From September 2000 to December 2004, he was with Airvana, Inc., where he conducted research on the third-generation wireless communications. Since January 2005, he has been with the Korea Advanced Institute of Science and Technology (KAIST), where he is now an Associate Professor in the Department of Electrical Engineering. His research interests include network information theory, coding theory, and their applications to wireless communications.

Dr. Chung is currently serving as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He served as a Guest Editor of *JCN*, EURASIP *Journal*, and *Telecommunications Review*. He is serving as a TPC Co-Chair of the 2014 IEEE International Symposium on Information Theory (ISIT) to be held in Hawaii. He also served as a TPC Co-Chair of WiOpt 2009.