Noisy Network Coding

Sung Hoon Lim
Department of EE
KAIST
Daejeon, Korea
sunghlim@kaist.ac.kr

Young-Han Kim
Department of ECE
UCSD
La Jolla, CA 92093, USA
yhk@ucsd.edu

Abbas El Gamal
Department of EE
Stanford University
Standord, CA 94305, USA
abbas@ee.stanford.edu

Sae-Young Chung
Department of EE
KAIST
Daejeon, Korea
sychung@ee.kaist.ac.kr

***Abstract*—A new coding scheme for multicasting multiple sources over a general noisy network is presented. The scheme naturally extends both network coding over noiseless networks by Ahlswede, Cai, Li, and Yeung, and compress–forward coding for the relay channel by Cover–El Gamal to general discrete memoryless and Gaussian networks. The scheme also recovers as special cases the results on coding for wireless relay networks and deterministic networks by Avestimehr, Diggavi, and Tse, and coding for wireless erasure networks by Dana, Gowaikar, Palanki, Hassibi, and Effros. The key idea is to use block Markov message repetition coding and simultaneous decoding. Instead of sending multiple independent messages over several blocks and decoding them sequentially as in previous relaying schemes, the same message is sent multiple times using independent codebooks and the decoder performs joint typicality decoding on the received signals from all the blocks without explicitly decoding the compression indices. New results on semideterministic relay networks and Gaussian networks demonstrate the potential of noisy network coding as a robust and scalable scheme for communication over wireless networks.**

I. Introduction

Consider a discrete memoryless multi-source multicast network with N sender–receiver pairs. Sender X_k wishes to send a message M_k to the set of receivers $Y(\mathcal{D})$, while acting as a relay for messages from other sources. The information capacity region of this network is not known in general. In the seminal paper on network coding [1], Ahlswede, Cai, Li, and Yeung established the capacity for the single source multicast case when the network is *noiseless*, that is, when it can be represented by a directed graph $(\mathcal{N}, \mathcal{E})$ with capacity limited links. They showed that capacity coincides with the cutset bound, generalizing the max-flow min-cut theorem [2], [3] to multiple destinations. Each relay in network coding sends a function of its incoming signals over each outgoing link instead of simply forwarding incoming signals. Dana, Gowaikar, Palanki, Hassibi, and Effros [4] showed that network coding is also optimal for noiseless multi-source multicast networks.

Subsequently, Ratnakar and Kramer [5] extended network coding to characterize the multicast capacity for single-source deterministic networks with broadcast but no interference at the receivers. Avestimehr, Diggavi, and Tse [6] further extended this result to deterministic networks with broadcast *and* interference to obtain the capacity lower bound

$$C \geq \max_{\prod_{k=1}^{N} p(x_k)} \min_{d \in \mathcal{D}} \min_{\mathcal{S}: 1 \in \mathcal{S}, d \in \mathcal{S}^c} H(Y(\mathcal{S}^c) | X(\mathcal{S}^c)). \quad (1)$$

They showed that this lower bound coincides with the cutset bound in the special case of linear finite-field networks, where the channel output is a linear function of input signals over a finite field.

In addition to signal interactions, randomness in the channel can be modeled by introducing erasures in the network. Dana, Gowaikar, Palanki, Hassibi, and Effros [4] considered erasure networks with broadcast and no interference, where the input signals are randomly erased. Smith and Vishwanath [7] considered erasure networks without broadcast, where the interference is modeled as a linear finite-field sum of incoming signals that are not erased. In particular, these papers establish capacity when the destination node has perfect knowledge of the erasure information of the entire network.

In an earlier and seemingly unrelated line of investigation, van der Meulen [8] introduced the relay channel $p(y_2, y_3 | x_1, x_2)$ with one source X_1, one destination Y_3, and one relay with sender–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 over the past 40 years. In their seminal paper on the relay channel [9], Cover and El Gamal established the cutset bound on capacity

$$C \leq \max_{p(x_1, x_2)} \min \{ I(X_1, X_2; Y_3), I(X_1; Y_2, Y_3 | X_2) \}. \quad (2)$$

This bound in essence generalizes the converse of the max-flow min-cut theorem to a noisy channel. In the same paper, they proposed two distinct coding schemes. At one extreme, the relay decodes the intended message from the source and forwards it to the destination. As Cover and El Gamal showed, this simple multi-hop coding scheme can be improved via coherent cooperation between the source and relay, the use of more sophisticated decoding at the destination, and partial decoding of the message at the relay. The resulting decode–forward scheme was shown to be optimal for the classes of degraded [9] and semi-deterministic [10] relay channels, and when the relay channel has orthogonal sender components [11]. The decode–forward coding scheme was extended to multiple relays by Aref [12] (see also El Gamal [13], Xie and Kumar [14], and Kramer, Gastpar, and Gupta [15]), who established the capacity of the physically degraded relay network and the deterministic network with broadcast but no interference. Although the later results by Ratnakar and

Kramer [5], and Avestimehr, Diggavi, and Tse [6] can be considered as extensions of Aref's result on the deterministic network, they used generalized network coding schemes to establish the multicast capacity.

At the other extreme, the relay compresses its noisy observation of the source signal and forwards the compressed description to the destination. This compress–forward coding scheme, again proposed in [9], can be viewed as a generalization of an analog-to-digital interface for forwarding relay observations. Due to its simplicity (and in some sense, less intelligence), the relay operation in the compress–forward coding scheme is more robust to end-to-end operations at the source and destination. Despite its simplicity, compress–forward was shown to be optimal for classes of deterministic [16] and modulo-sum [17] relay channels.

The Cover–El Gamal compress–forward lower bound on capacity has the form

$$C \geq \max I(X_1; \hat{Y}_2, Y_3 | X_2), \tag{3}$$

where the maximum is over $p(x_1)p(x_2)p(\hat{y}_2|y_2, x_2)$ subject to $I(X_2; Y_3) \geq I(Y_2; \hat{Y}_2 | X_2, Y_3)$. This lower bound is achieved via a block Markov coding scheme, whereby 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. The receiver sequentially decodes the compression index and then uses it to decode the message sent in the previous block. Kramer, Gastpar, and Gupta extended this form of the compress–forward lower bound to general networks to obtain the following lower bound on the single-source multicast capacity [15, Theorem 3 with $U = \emptyset$]:

$$C \geq \max \min_{d \in \mathcal{D}} I(X_1; \hat{Y}_2^N, Y_d | X_2^N), \tag{4}$$

where the maximum is over $\prod_{k=1}^N p(x_k)p(\hat{y}_k|y_k, x_k)$ satisfying $I(X(\mathcal{T}); Y_d | X(\mathcal{T}^c), X_d) \geq I(Y(\mathcal{T}); \hat{Y}(\mathcal{T}) | X_2^N, \hat{Y}(\mathcal{T}^c), Y_d) + \sum_{k \in \mathcal{T}} I(X_2^N; \hat{Y}_k | X_k)$ for all $d \in \mathcal{D}$ and $\mathcal{T} \subseteq [2:N]$ (here $\mathcal{T}^c = [2:N] \backslash \mathcal{T}$).

Around the same time, El Gamal, Mohseni, and Zahedi [18] showed that the characterization of the compress–forward lower bound (3) is equivalent to

$$C \geq \max_{\substack{p(x_1)p(x_2) \\ p(\hat{y}_2|y_2,x_2)}} \min\{I(X_1; \hat{Y}_2, Y_3 | X_2), \; I(X_1, X_2; Y_3) - I(Y_2; \hat{Y}_2 | X_1, X_2, Y_3)\}. \tag{5}$$

This form of the compress–forward lower bound closely resembles the cutset bound (2), except that in the first term of the cutset bound Y_2 is replaced by $\hat{Y}_2$, in the second term we have a negative compression penalty, and the maximization is over independent (X_1, X_2). Recently, El Gamal and Kim [19] proved the achievability of this alternative characterization directly using the same codebook generation and encoding steps as in the achievability proof of (3) but with simultaneous decoding of the compression index and the message.

In this paper, we present a new coding scheme that extends the alternative characterization (3) to general noisy multi-source multicast networks. When applied to the single-source multicast network case, our coding scheme gives the lower bound on capacity

$$C \geq \max \min_{d \in \mathcal{D}} \min_{\substack{\mathcal{S} \subseteq [1:N] \\ 1 \in \mathcal{S}, d \in \mathcal{S}^c}} (I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c)) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d)), \tag{6}$$

where the maximum is over $\prod_{k=1}^N p(x_k)p(\hat{y}_k|y_k, x_k)$. Note that except for the relay channel case, this lower bound is tighter than the previous bound (4), as demonstrated via a simple example in Section IV-A. Our lower bound, however, does not uniformly outperform the hybrid compress–forward/decode–forward scheme in [15, Theorem 3], but can be combined with decode–forward in various ways to yield similar lower bounds to Theorem 7 in [9] and to Theorem 3 in [15].

Our coding scheme naturally extends and unifies the results on network coding and deterministic networks to noisy networks. For example, to recover the lower bound (1) for deterministic networks [6], we can take $\hat{Y}^N = Y^N$ in (6). As we show, it also recovers results on approximation for Gaussian networks [6] and coding for wireless erasure networks [4], [7]. While the coding techniques 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 that holds for arbitrary discrete memoryless and Gaussian networks.

The key idea behind our compress–forward coding scheme is to use block Markov message repetition coding and simultaneous decoding. Instead of sending different messages over multiple blocks and decoding one message at a time as in previous coding schemes, the source transmits the same message over multiple blocks using independently generated codebooks and the destination decodes the message using all received blocks. To be fair, the idea of transmitting a single message over multiple blocks was previously used in [1] to extend their coding scheme from acyclic to cyclic networks via an unfolding technique, which was later used in [6] to extend the coding scheme from layered to nonlayered networks. In comparison, our coding scheme applies to general networks without the need for unfolding.

The rest of the paper is organized as follows. In the next section, we formally set up the problem of multi-source multicast over a general network and state the main result (noisy network coding inner bound). The proof of the bound using the new compress–forward scheme is given in Section III. In Section IV, we specialize Theorem 1 to the class of semideterministic networks (which includes both deterministic [6] and erasure [4], [7] networks) and Gaussian networks [6]. Throughout the paper, we follow the notation in [19].

II. Problem Statement and Main Result

An N-node discrete memoryless multi-source multicast network (DM-MMN) $(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N, p(y^N|x^N), \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_N)$ consists of N sender–receiver alphabet pairs $(\mathcal{X}_k, \mathcal{Y}_k)$, $k \in [1:N] := \{1, \ldots, N\}$, and a collection of conditional pmfs $p(y_1, \ldots, y_N | x_1, \ldots, x_N)$. Each node $k \in [1:N]$ wishes to send a message M_k to a set of destination nodes,

$\mathcal{D} \subseteq [1:N]$. Formally, a $(2^{nR_1}, \dots, 2^{nR_N}, n)$ code for a DM-MMN consists of N message sets $[1:2^{nR_1}], \dots, [1:2^{nR_N}]$, 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 $d \in \mathcal{D}$ that assigns message estimates $(m_{1d}, \dots, m_{Nd})$ to each (y_d^n, m_d).

We assume that messages M_k, $k \in [1:N]$, are each uniformly distributed over $[1:2^{nR_k}]$ and are independent of each other. The probability of error is defined by $P_e^{(n)} = \mathsf{P}\{(\hat{M}_{1d}, \dots, \hat{M}_{Nd}) \neq (M_1, \dots, M_N) \text{ for some } d \in \mathcal{D}\}$. A rate tuple $(R_1, \dots, R_N)$ is said to be achievable if there exists a sequence of $(2^{nR_1}, \dots, 2^{nR_N}, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity region of the DM-MMN is the closure of the set of achievable rate tuples.

Theorem 1 (Noisy network coding inner bound): A rate tuple $(R_1, \dots, R_N)$ is achievable for the DM-MMN if there exists some joint pmf $p(q)\prod_{k=1}^N p(x_k|q)p(\hat{y}_k|y_k, x_k, q)$ such that

$$R(\mathcal{S}) < \min_{d \in \mathcal{S}^c \cap \mathcal{D}} (I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_d | X(\mathcal{S}^c), Q) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_d, Q)) \quad (7)$$

for all cutsets $\mathcal{S}$, where $R(\mathcal{S}) = \sum_{k \in \mathcal{S}} R_k$.

Recall the cutset outer bound on the capacity region [13]: If the rate tuple $(R_1, \dots, R_N)$ is achievable, then there exists some joint pmf $p(x_1, \dots, x_N)$ such that

$$R(\mathcal{S}) \leq I(X(\mathcal{S}); Y(\mathcal{S}^c) | X(\mathcal{S}^c)) \quad (8)$$

for all cutsets $\mathcal{S}$. In comparison, the inner bound (7) in Theorem 1 has the first term with Y replaced by the "compressed" version $\hat{Y}$, the additional negative term that quantifies the rate requirement to convey the compressed version, and the maximum over independent X^N.

Note that Theorem 1 can be easily specialized to different source–destination configurations. For example, by taking $R_k = 0$ for some nodes k, we can subsume the situation in which only a subset of source nodes send messages. Also when each message M_k is communicated to different sets $\mathcal{D}_k$ of destination nodes, Theorem 1 continues to hold with multicast completion $\mathcal{D} = \cup_{k=1}^N \mathcal{D}_k$ of destinations nodes.

III. Proof of Theorem 1 for the Relay Channel

To illustrate the main idea of the coding scheme and highlight the major differences from the traditional compress–forward coding scheme [9], [15], we prove the achievability for the relay channel. For the proof for the general DM-MMN, we refer the reader to [20].

Let $\mathbf{x}_{kj}$ denote $(x_{k,(j-1)n+1}, \dots, x_{k,jn})$, $j \in [1:b]$; thus $x_k^{bn} = (x_{k1}, \dots, x_{k,nb}) = (\mathbf{x}_{k1}, \dots, \mathbf{x}_{kb}) = \mathbf{x}_k^b$. 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 $\mathbf{y}_{2j}$ with $\mathbf{x}_{2j}$ as state information, 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 $m \in [1:2^{nbR}]$ using $(\mathbf{y}_{31}, \dots, \mathbf{y}_{3b})$ by joint typical decoding for each of b blocks simultaneously. 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^{nR_2} sequences $\mathbf{x}_{2j}(l_{j-1})$, $l_{j-1} \in [1:2^{nR_2}]$, each according to $\prod_{i=1}^n p_{X_2}(x_{2,(j-1)n+i})$. For each $\mathbf{x}_{2j}(l_{j-1})$, randomly and conditionally independently generate 2^{nR_2} sequences $\hat{\mathbf{y}}_{2j}(l_j|l_{j-1})$, $l_j \in [1:2^{nR_2}]$, 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 $\mathcal{C}_j = \{\mathbf{x}_{1j}(m), \mathbf{x}_{2j}(l_{j-1}), \hat{\mathbf{y}}_{2j}(l_j|l_{j-1}) : m \in [1:2^{nbR}], l_{j-1}, l_j \in [1:2^{nR_2}]\}$ for $j \in [1:b]$. Encoding and decoding are explained with help of Table I on the next page.

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. The codeword pair $(\mathbf{x}_{1j}(m), \mathbf{x}_{2j}(l_{j-1}))$ is transmitted in block $j \in [1:b]$.

Decoding: Let $\epsilon > \epsilon'$. At the end of block b, the decoder finds a unique index m such that $(\mathbf{x}_{1j}(m), \hat{\mathbf{y}}_{2j}(l_j|l_{j-1}), \mathbf{x}_{2j}(l_{j-1}), \mathbf{y}_{3j}) \in \mathcal{T}_\epsilon^{(n)}$ for all $j \in [1:b]$ for some $l_1, l_2, \dots, l_b$.

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

$$\mathcal{E}_0 := \cup_{j=1}^b \{(\hat{\mathbf{Y}}_{2j}(l_j|L_{j-1}), \mathbf{X}_{2j}(L_{j-1}), \mathbf{Y}_{2j}) \notin \mathcal{T}_{\epsilon'}^{(n)} \ \forall l_j\},$$
$$\mathcal{E}_m := \{(\mathbf{X}_{1j}(m), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)}, \ j \in [1:b] \text{ for some } l_1, l_2, \dots, l_b\}.$$

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

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

By the covering lemma [19], $\mathsf{P}(\mathcal{E}_0) \to 0$ as $n \to \infty$, if $R_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(\epsilon')$, and by the conditional typicality lemma [19], $\mathsf{P}(\mathcal{E}_0^c \cap \mathcal{E}_1^c) \to 0$ as $n \to \infty$.

To bound $\mathsf{P}(\cup_{m \neq 1} \mathcal{E}_m)$, assume without loss of generality that $(L_1, \dots, L_b) = (1, \dots, 1)$; recall the symmetry of the codebook construction. Define $\mathcal{A}_j(m, l_{j-1}, l_j) := \{(\mathbf{X}_{1j}(m), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)}\}$. Then

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_m) &= \mathsf{P}(\cup_{l^b} \cap_{j=1}^b \mathcal{A}_j(m, l_{j-1}, l_j)) \\
&\leq \sum_{l^b} \mathsf{P}(\cap_{j=1}^b \mathcal{A}_j(m, l_{j-1}, l_j)) \\
&= \sum_{l^b} \prod_{j=1}^b \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) \qquad (9) \\
&\leq \sum_{l^b} \prod_{j=2}^b \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)),
\end{aligned}$$

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

where (9) follows since the codebook is generated independently for each block j and the channel is memoryless. Note that if $m \neq 1$ and $l_{j-1} = 1$, then $\mathbf{X}_{1j}(m) \sim \prod_{i=1}^{n} p_{X_1}(x_{1,(j-1)n+i})$ is independent of $(\hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j})$ (given $L_{j-1} = L_j = 1$). Hence, by the joint typicality lemma [19],

$$\begin{aligned}\mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) &= \mathsf{P}\{(\mathbf{X}_{1j}(m), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}), \mathbf{X}_{2j}(l_{j-1}), \mathbf{Y}_{3j}) \in \mathcal{T}_\epsilon^{(n)}\} \\ &\leq 2^{-n(I(X_1;\hat{Y}_2,Y_3|X_2)-\delta(\epsilon))} =: 2^{-n(I_1-\delta(\epsilon))}. \end{aligned} \tag{10}$$

Similarly, if $m \neq 1$ and $l_{j-1} \neq 1$, then $(\mathbf{X}_{1j}(m), \mathbf{X}_{2j}(l_{j-1}), \hat{\mathbf{Y}}_{2j}(l_j|l_{j-1}))$ is independent of $\mathbf{Y}_{3j}$ (given $L_{j-1} = L_j = 1$). Hence, by the joint typicality lemma [19],

$$\begin{aligned}\mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) &\leq 2^{-n(I(X_1,X_2;Y_3)+I(\hat{Y}_2;X_1,Y_3|X_2)-\delta(\epsilon))} \\ &=: 2^{-n(I_2-\delta(\epsilon))}. \end{aligned} \tag{11}$$

If l^{b-1} has k 1s, then by (10) and (11)

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

Therefore

$$\begin{aligned}&\sum_{l^b}\prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) \\ &= \sum_{l_b}\sum_{l^{b-1}}\prod_{j=2}^{b} \mathsf{P}(\mathcal{A}_j(m, l_{j-1}, l_j)) \\ &\leq \sum_{l_b}\sum_{k=0}^{b-1}\binom{b-1}{k} 2^{n(b-1-k)R_2} 2^{-n(kI_1+(b-1-k)I_2-(b-1)\delta(\epsilon))} \\ &= \sum_{l_b}\sum_{k=0}^{b-1}\binom{b-1}{k} 2^{-n(kI_1+(b-1-k)(I_2-R_2)-(b-1)\delta(\epsilon))} \\ &\leq \sum_{l_b}\sum_{k=0}^{b-1}\binom{b-1}{k} 2^{-n((b-1)(\min\{I_1,\, I_2-R_2\}-\delta(\epsilon)))} \\ &\leq 2^{nR_2} 2^{b-1} \cdot 2^{-n((b-1)(\min\{I_1,\, I_2-R_2\}-\delta(\epsilon)))},\end{aligned}$$

which tends to zero as $n \to \infty$, if $R < \frac{b-1}{b}(\min\{I_1, I_2 - R_2\} - \delta(\epsilon)) - \frac{1}{b}R_2$. Finally, by eliminating $R_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(\epsilon')$ and letting $b \to \infty$, we have shown the achievability of any rate $R < \min\{I(X_1; \hat{Y}_2, Y_3|X_2), I(X_1, X_2; Y_3) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_3)\} - \delta'(\epsilon)$. This concludes the proof of Theorem 1 for the special case of the relay channel.

IV. Examples

A. A Simple Line Network

As mentioned before, the noisy network coding lower bound on the single-source capacity (6) is tighter than the previous compress–forward lower bound (4). The following example highlights the performance improvement.

Consider a noiseless DM-MMN with $N = 4, \mathcal{D} = \{4\}$, where $Y_2 = X_1$, $Y_3 = X_2$, $Y_4 = X_3$ are all binary; that is, one bit can be noiselessly routed from node 1 to 2 to 3 to 4. Consider the single-source setup ($R_2 = R_3 = R_4 = 0$). Trivially the capacity $C = 1$. By taking $\hat{Y}^N = Y^N$ and the uniform input pmf, the noisy network coding lower bound achieves the capacity. On the other hand, it can be readily checked by simple algebra that an achievable rate R in the bound (4) satisfies

$$R \leq I(X_1; \hat{Y}_2, \hat{Y}_3, Y_4|X_2, X_3) \leq I(X_1; \hat{Y}_3|X_2, X_3) = 0.$$

Therefore, no positive rate is achievable. One performance limitation of the traditional compress–forward coding scheme is due to the fact that the receiver is required to decode the compression index of Y_2 correctly, even though the goal is to decode the intended message only. Note that for this particular example, the hybrid coding scheme in [15, Theorem 3] does not achieve any positive rate.

B. Semideterministic Networks

A DM-MMN is said to be *semideterministic* if the channel output symbol $Y_k = f_{kd}(X_1, \ldots, X_N, Y_d)$ at node k is a deterministic function of input symbols $(X_1, \ldots, X_N)$ and the output symbol Y_d at a destination for all $k \in [1 : N]$ and $d \in \mathcal{D}$. For this class of networks, the noisy network coding inner bound can be simplified by taking $\hat{Y}^N = Y^N$ as follows:

$$R(\mathcal{S}) < I(X(\mathcal{S}); Y(\mathcal{S}^c)|X(\mathcal{S}^c), Q) \tag{12}$$

for all cutsets $\mathcal{S}$. This rate region is of the same form as the cutset outer bound except that the input distribution is chosen from the set of product input distributions. Hence, for the semideterministic DM-MMN for which the cutset bound is attained by product input distributions, both bounds are tight.

We consider examples to illustrate applications of (12).

Example 1 (Noiseless networks): Consider a noiseless DM-MMN modeled by a weighted directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{C})$, where $\mathcal{N} = [1 : N]$ is the set of nodes, $\mathcal{E} \subseteq [1 : N] \times [1 : N]$ is the set of edges, and $\mathcal{C} = \{C_{jk} \in \mathbb{R}^+ : (j, k) \in \mathcal{E}\}$ is the set of weights. Each edge represents a noiseless communication link with capacity C_{jk}. Then, a noiseless DM-MMN with destination nodes $\mathcal{D}$ is a special case of the semideterministic DM-MMN with $Y_k = \{X_{jk} : (j, k) \in \mathcal{E}\}$, $k \in [1 : N]$. Each link

$(j,k) \in \mathcal{E}$ carries an input symbol $x_{jk} \in \mathcal{X}_{jk}$ with link capacity $C_{jk} = \log|\mathcal{X}_{jk}|$. By taking the uniform pmf on all input symbols in (12), the inner bound coincides with the cutset bound

$$R(\mathcal{S}) \le C(\mathcal{S}) \tag{13}$$

for all cutsets $\mathcal{S}$, where $C(\mathcal{S}) = \sum_{(j,k)\in\mathcal{E}:j\in\mathcal{S},k\in\mathcal{S}^c} C_{jk}$ is the cut capacity from $\mathcal{S}$ to $\mathcal{S}^c$. This recovers previously known results by Ahlswede *et al.* [1] and Dana *et al.* [4].

Example 2 (Deterministic networks): A DM-MMN is said to be deterministic if $Y_k = f_k(X_1, \ldots, X_N)$, $k \in [1:N]$. Then (12) can be further simplified to

$$R(\mathcal{S}) < H(Y(\mathcal{S}^c)|X(\mathcal{S}^c), Q) \tag{14}$$

for all cutsets $\mathcal{S}$, which generalizes (1) by Avestimehr *et al.* [6]. For linear finite field networks $Y_k = \sum_{j=1}^{N} G_{jk}X_j$, $k \in [1:N]$, where the channel gains G_{jk} and inputs X_j take values from $GF(q)$, (14) characterizes the capacity region.

We now consider a *deterministic* DM-MMN $p(y^N|x^N, s)$ *with state*, where $Y_k = f_k(X^N, S)$ for all $k \in [1:N]$ is a deterministic function of (X^N, S) and S is a discrete memoryless (i.i.d.) random state that models fading, erasure, interference, or more generally channel uncertainties. Suppose the destination nodes have access to the state sequence s^n strictly causally, i.e., the encoders are given by $x_{ki}(m_k, s^{i-1}, y_k^{i-1})$, $i \in [1:n]$ for $k \in \mathcal{D}$, and $x_{ki}(m_k, y_k^{i-1})$, $i \in [1:n]$ for $k \notin \mathcal{D}$, and the decoders are $(m_{1d}, \ldots, m_{Nd})(m_d, y_d^n, s^n)$ for $d \in \mathcal{D}$. Then, (12) reduces to

$$R(\mathcal{S}) < H(Y(\mathcal{S}^c)|X(\mathcal{S}^c), S, Q), \tag{15}$$

which is of the same form as the cutset bound except being evaluated for product pmfs instead of joint pmfs.

Example 3 (Erasure networks): Consider the erasure network in which the channel output at node $k \in [1:N]$ is $Y_k = \{Y_{jk} : j \in [1:N]\}$ where $Y_{jk} = \varepsilon$ if the input signal is erased, and $Y_{jk} = X_j$, otherwise. Since Y_k is a function of the channel inputs X^N and the state, the erasure network is a special case of the deterministic DM-MMN with state. Suppose each destination node has complete access to the erasure pattern of the network. By taking the uniform pmf on all input symbols and $Q = \emptyset$, (15) becomes

$$R(\mathcal{S}) < \sum_{j\in\mathcal{S}} \log|\mathcal{X}_j|(1 - \mathsf{P}\{Y_{jk} = \varepsilon \text{ for all } k \in \mathcal{S}^c\}),$$

which coincides with the cutset bound, recovering the result by Dana *et al.* [4].

A similar example can be provided for linear finite-field networks with state, which generalizes the previous result by Smith and Vishwanath [7]; see [20] for details.

C. Gaussian Networks

Consider the additive white Gaussian noise MMN in which the channel outputs are given by $Y^N = GX^N + Z^N$, where $G \in \mathbb{R}^{N\times N}$ is the channel gain matrix and Z^N is a vector of independent white Gaussian noise processes with zero mean and unit variance. We assume average power constraint P on each sender, i.e., $\sum_{i=1}^{n} \mathsf{E}\left(x_{ki}^2(m_k, Y_k^{i-1})\right) \le nP$ for all $k \in [1:N]$ and $m_k \in [1:2^{nR_k}]$. For each cutset $\mathcal{S} \subseteq [1:N]$, define a channel gain matrix $G(\mathcal{S})$ such that $Y(\mathcal{S}^c) = G(\mathcal{S})X(\mathcal{S}) + G'(\mathcal{S})X(\mathcal{S}^c) + Z(\mathcal{S}^c)$. Now it can be shown [20] that the noisy network coding inner bound yields

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

Compared to the cutset bound

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

the inner bound is within $(N/2)\log(4N)$ bits of the capacity region (for each rate R_k), regardless of the channel gain and power constraint. This generalizes the result of Avestimehr *et al.* [6] to multiple sources with a slightly better factor.

References

[1] 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.

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

[3] 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.

[4] 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.

[5] 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.

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

[7] B. Smith and S. Vishwanath, "Unicast transmission over multiple access erasure networks: Capacity and duality," in *Proc. Information Theory Workshop*, Tahoe City, California, May 2007, pp. 331–336.

[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] A. El Gamal and M. R. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, p. 536, 1982.

[11] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1815–1817, 2005.

[12] M. R. Aref, "Information flow in relay networks," Ph.D. Thesis, Stanford University, Stanford, CA, Oct. 1980.

[13] A. El Gamal, "On information flow in relay networks," in *Proc. IEEE National Telecom Conference*, Nov. 1981, vol. 2, pp. D4.1.1–D4.1.4.

[14] 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.

[15] 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.

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

[17] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channel," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007, pp. 596–600.

[18] 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.

[19] A. El Gamal and Y.-H. Kim, *Lecture Notes on Network Information Theory*, Stanford University and UCSD, 2009.

[20] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," 2009.