Distributed Decode–Forward for Multicast

Sung Hoon Lim and Kwang Taik Kim
Digital Media & Communications Research Center
Samsung Electronics
Suwon, Gyeonggi-do 443-742, Korea

Young-Han Kim
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093

***Abstract*—A new coding scheme for multicasting a message over a general relay network is presented that extends both network coding for graphical networks by Ahlswede, Cai, Li, and Yeung, and partial decode–forward for relay channels by Cover and El Gamal. For the N-node Gaussian multicast network, the scheme achieves within $0.5N$ bits from the capacity, improving upon the best known capacity gap results. The key idea is to use multicoding at the source as in Marton coding for broadcast channels. Instead of recovering a specific part of the message as in the original partial decode–forward scheme, a relay in the proposed distributed decode–forward scheme recovers an auxiliary index that implicitly carries some information about the message and forwards it in block Markov coding. This scheme can be adapted to broadcasting multiple messages over a general relay network, extending and refining a recent result by Kannan, Raja, and Viswanath.**

I. Introduction

Relays are fundamental building blocks for communication networks. By propagating desired messages over the network—either directly or indirectly—they help increase throughput, improve reliability, and expand range, the effects of which are most pronounced in wireless networks with high path loss at low signal-to-noise ratio.

Since van der Meulen [1] studied the 3-node relay channel model in the context of mathematical communication theory, numerous relaying schemes have been proposed in the literature. Among these, decode–forward, compress–forward, and amplify–forward are particularly well studied and form a basis, along with direct transmission, for most other relaying schemes. In the decode–forward coding scheme, originally proposed by Cover and El Gamal [2], the relay recovers, re-encodes, and forwards the transmitted message. This digital-to-digital interface can be further combined with direct transmission via superposition coding, leading to the partial decode–forward coding scheme [2], [3].

Alternatively, the relay can communicate the received sequence itself instead of decoding it for all or part of the message. In the compress–forward coding scheme, proposed again by Cover and El Gamal [2], the relay serves as an analog-to-digital interface and compresses its received sequence via vector quantization and forwards the compression index. Going further in this direction, in the amplify–forward coding scheme, originally proposed by Schein and Gallager [4], and popularized by Laneman, Tse, and Wornell [5], the relay serves as an analog-to-analog interface and simply sends a scaled version of its received sequence. All three schemes are known to achieve the capacity within 1 bit for the single-antenna Gaussian relay channel [6], [7].

These schemes have been extended beyond relay channels with varying degrees of generality (in operation) and scalability (in performance). Amplify–forward can be readily applied to turn an arbitrary Gaussian multihop network into a single-hop network with intersymbol interference, but fails to bring in scalability as its achievable rate can have an unbounded gap from capacity. Noisy network coding [8], [9], a variant of compress–forward, provides a scalable performance for general *multimessage multicast* networks with a bounded gap from capacity; see also [6]. But it suffers from noise propagation (as relays do not decode for the message) and does not provide a scalable performance for *multiple-unicast* networks.

An extension of decode–forward has been developed [10], [11] for single-message multicast networks, whereby the relays form a route from the source to destinations, and an intermediate node recovers the codewords from all its predecessors en route and forwards the corresponding message to the next hop. The decoding requirement, however, is often too restrictive and the resulting achievable rate can have an unbounded gap from capacity. In principle, potential extensions of partial decode–forward (if any) would alleviate this issue, but with a notable exception of Aref's results on classes of deterministic networks [12, Sections 3.4 and 3.5], partial decode–forward has rarely been studied beyond the 3-node relay channel, as it is unclear which relay should be assigned to forward which part of the message. This problem seems even more intractable when multiple messages are to be communicated.

With these limitations of the existing relaying schemes in mind, our goal is to develop a scalable coding scheme for general *broadcast* networks (i.e., a single source with multiple independent messages for multiple destinations). The main difficulty for developing such a scheme can be perhaps recognized from the fact

978-1-4799-5186-4/14/$31.00

that the crucial component of Marton coding [13] for the (single-hop) broadcast channel is careful coordination between the codewords for different messages. For multihop networks, similar coordination becomes far more challenging since the source should control the codewords transmitted from multiple nodes. This coordination requirement excludes amplify–forward and noisy network coding for broadcast, as the behavior of the relays is not available at the source in these coding schemes, leaving decoding-based schemes as the only feasible alternatives.

Motivated by the need for a scalable extension of decode–forward to general broadcast networks, we develop the *distributed decode–forward* coding scheme in this paper. As in [13], [12], the coding scheme uses *multicoding* as a basis for coordination among distributed nodes. To some extent, the idea of "pruning" superletter codewords by Anand and Kumar [14] for multicast and extended by Kannan, Raja, and Viswanath [15] to broadcast is also reminiscent of multicoding. By encoding the message with compatible codewords via multicoding, the source can coordinate and control the transmission over the entire network, which results in a scalable performance. With streamlined operations and single-letter achievable rates for general *broadcast* networks, the proposed scheme can be viewed as an extension and refinement of the coding scheme of Kannan et al. [15].

To best explain the source-centric coordination aspect of the distributed decode–forward scheme, our main focus in this 5-page paper is on explaining how the multicoding approach can be applied to the single-message multicast network; see Sections II and III. For the special case of the relay channel, the achievable rate coincides with that of partial decode–forward. In this sense, the proposed coding scheme is an extension of partial decode–forward to networks. Unlike (partial) decode–forward whereby the source node controls the behavior of the relays by routing (parts of) the message itself with superposition coding over multiple hops, however, the source node in distributed decode–forward controls the relays by communicating auxiliary indices that carry information about the message rather implicitly and are forwarded in two hops. Although this multicoding-based coordination results in some noise propagation, one is rewarded by a scalable coding scheme for general networks with an arbitrary number of hops. For example, when applied to deterministic multicast networks, distributed decode–forward extends and improves upon the result by Avestimehr, Diggavi, and Tse [6]. In particular, it recovers network coding for graphical networks by Ahlswede, Cai, Li, and Yeung [16]. When applied to Gaussian multicast networks, the scheme achieves within $0.5N$ bits from the capacity, which improves upon the previously known tightest gap of $0.63N$ bits [8].

This two-hop multicoding approach can be adapted to broadcast, our stated goal of this study. In Section IV, we briefly describe how the multicast coding scheme can be modified to broadcast and present the resulting rate region for general broadcast networks.

Throughout the paper, we use the notation in [17]. In particular, a sequence of random variables with node index k and time index $i \in [1:n] := \{1,\dots,n\}$ is denoted as $X_k^n := (X_{k1},\dots,X_{kn})$. A tuple of random variables is denoted as $X(\mathcal{A}) := (X_k : k \in \mathcal{A})$.

II. Problem Setup and Main Results

Consider the N-node discrete memoryless multicast network (DM-MN) $p(y^N|x^N)$ [17, Section 18.1]. Suppose that source node 1 wishes to send a message M to a set of destination nodes $\mathcal{D} \subseteq [2:N]$. A $(2^{nR}, n)$ code for the DM-MN consists of

- a message set $[1:2^{nR}]$,
- a source encoder that assigns a code symbol $x_{1i}(m, y_1^{i-1})$ to each message $m \in [1:2^{nR}]$ and received sequence $y_1^{i-1} \in \mathcal{Y}_1^{i-1}$ for $i \in [1:n]$,
- a set of relay encoders, where encoder $k \in [2:N]$ assigns a $x_{ki}(y_k^{i-1})$ to each y_k^{i-1} for $i \in [1:n]$, and
- a set of decoders, where decoder $d \in \mathcal{D}$ assigns an estimate $\hat{m}_d$ or an error message e to each y_d^n.

We assume that the message M is uniformly distributed over the message set. The average probability of error is defined as $P_e^{(n)} = \mathsf{P}\{\hat{M}_d \neq M \text{ for some } d \in \mathcal{D}\}$. A rate R is said to be achievable if there exists a sequence of $(2^{nR}, n)$ codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$. The capacity of the DM-MN is the supremum of the set of achievable rates.

We are ready to state the main theorem of the paper.

Theorem 1. *The capacity of the DM-MN $p(y^N|x^N)$ with a set $\mathcal{D}$ of destination nodes is lower bounded as*

$$C \geq \max \min_{d\in\mathcal{D}} \min \Big[I(X(\mathcal{S}); U(\mathcal{S}^c), Y_d | X(\mathcal{S}^c)) - \sum_{k\in\mathcal{S}^c} I(U_k; U(\mathcal{S}_k^c), X^N | X_k, Y_k) \Big], \tag{1}$$

where $\mathcal{S}_k^c = \mathcal{S}^c \cap [2:k-1]$ for $k \in [2:N]$ and the second minimum is over all $\mathcal{S} \subseteq [1:N]$ such that $1 \in \mathcal{S}$ and $d \in \mathcal{S}^c$. The maximum in (1) is over all pmfs of the form $(\prod_{k=2}^N p(x_k)) p(x_1|x_2^N) p(u_2^N|x^N)$.

The proof of Theorem 1 along with the description and analysis of the associated distributed decode–forward coding scheme is deferred to Section III. Here we illustrate the utility of Theorem 1 via three canonical examples.

Example 1 (Relay channels)**.** Consider the discrete memoryless relay channel $p(y_2, y_3|x_1, x_2)$, in which the sender (node 1) communicates a message to the receiver (node 3) with the help of the relay (node 2). This corresponds to a DM-MN with $N = 3$, $\mathcal{D} = \{3\}$, and $Y_1 = X_3 = \emptyset$. For this case, Theorem 1 simplifies as

$$C \geq \max_{p(x_1,x_2,u_2)} \min\big\{I(X_1, X_2; Y_3), \\ I(U_2; Y_2|X_2) + I(X_1; Y_3|X_2, U_2)\big\},$$

which coincides with the partial decode–forward lower bound [2, Theorem 7]; see also the standalone version in [3] and [17, Theorem 16.3].

Example 2 (Deterministic networks)**.** Suppose $Y_k = g_k(X_1, \ldots, X_N)$, $k \in [1:N]$. Then, by setting $U_k = Y_k$, $k \in [2:N]$ in (1), Theorem 1 simplifies as

$$C \geq \max \min_{\mathcal{S}:1\in\mathcal{S},\, \mathcal{S}^c\cap\mathcal{D}\neq\emptyset} H(Y(\mathcal{S}^c)|X(\mathcal{S}^c)), \quad (2)$$

where the maximum is over all pmfs $(\prod_{k=2}^N p(x_k))p(x_1|x_2^N)$. This bound has the same form as the cutset bound (see, for example, [17, Section 18.3.1], whereby the maximum is taken over all joint pmfs $p(x^N)$. Thus, if the maximum of the cutset bound is attained by a pmf of the form $(\prod_{k=2}^N p(x_k))p(x_1|x_2^N)$, which includes product pmfs, then the lower bound in (2) is tight. In particular, it recovers the celebrated results by Ahlswede, Cai, Li, and Yeung [16] for graphical networks and by Avestimehr, Diggavi, and Tse [6] for linear deterministic networks, but from a completely different path. See Section IV for further discussions.

Example 3 (Gaussian networks)**.** Consider the additive white Gaussian noise network, in which the channel outputs are $Y_k = g_{k1}X_1 + \cdots + g_{kN}X_N + Z_k, \quad k \in [1:N]$. Here g_{kj} is the channel gain from node j to node k and $Z_1, \ldots, Z_N$ are independent Gaussian noise components with zero mean and unit variance. We assume average power constraint P on each X_k [17, Section 19.1]. On the one hand, the cutset bound on the capacity leads to

$$\begin{aligned} C &\leq I(X(\mathcal{S}); Y(\mathcal{S}^c)|X(\mathcal{S}^c)) \\ &= \frac{1}{2}\log\Big|I + G(\mathcal{S})K_{X(\mathcal{S})}G^T(\mathcal{S})\Big| \\ &= \frac{1}{2}\log\Big|I + K_{X(\mathcal{S})}G^T(\mathcal{S})G(\mathcal{S})\Big| \\ &\leq \frac{1}{2}\log\Big(\Big|I + PG^T(\mathcal{S})G(\mathcal{S})\Big|\Big|I + \frac{1}{P}K_{X(\mathcal{S})}\Big|\Big) \\ &\overset{(a)}{\leq} \frac{1}{2}\log\big|I + PG(\mathcal{S})G^T(\mathcal{S})\big| + \frac{|\mathcal{S}|}{2}, \end{aligned}$$

for all $\mathcal{S} \subseteq [1:N]$ such that $1 \in \mathcal{S}$ and $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$. Here $K_{X(\mathcal{S})}$ is the covariance matrix of $X(\mathcal{S})$ and (a) follows by the Hadamard inequality. On the other hand, in the lower bound (1) we set X_k, $k \in [1:N]$, i.i.d. $\mathrm{N}(0, P)$, and $U_k = g_{k1}X_1 + \cdots + g_{kN}X_N + \hat{Z}_k$, $k \in [2:N]$, where $\hat{Z}_k \sim \mathrm{N}(0, 1)$ are independent of each other and of (X^N, Y^N). Then,

$$\begin{aligned} &I(X(\mathcal{S}); U(\mathcal{S}^c), Y_d|X(\mathcal{S}^c)) \\ &\quad\geq I(X(\mathcal{S}); U(\mathcal{S}^c)|X(\mathcal{S}^c)) \\ &\quad= \frac{1}{2}\log\big|I + PG(\mathcal{S})G^T(\mathcal{S})\big| \end{aligned}$$

and

$$\begin{aligned} I(U_k; U(\mathcal{S}_k^c), X^N|X_k, Y_k) &= I(U_k; X^N|X_k, Y_k) \\ &= \frac{1}{2}\log\left(\frac{1 + 2S_k}{1 + S_k}\right) \leq \frac{1}{2}, \end{aligned}$$

where $S_k = \sum_{j\neq k} g_{kj}^2 P$. Hence, Theorem 1 simplifies as

$$C \geq \min_{\mathcal{S}:1\in\mathcal{S},\mathcal{S}^c\cap\mathcal{D}\neq\emptyset} \frac{1}{2}\log\big|I + PG(\mathcal{S})G^T(\mathcal{S})\big| - \frac{|\mathcal{S}^c|}{2}.$$

Comparing the upper and lower bounds we can conclude that distributed decode–forward achieves within $0.5N$ bits from the cutset bound and thus from the capacity. Note that this establishes the tightest known gap for single-message Gaussian multicast networks.

III. Proof of Theorem 1

We use a block Markov coding scheme in which a sequence of b i.i.d. messages M_j, $j \in [1:b]$, is sent over b blocks each consisting of n transmissions. For each block, we generate codewords U_k, $k \in [2:N]$, to be recovered at relay k. Using multicoding [17, Sections 7.8 and 8.3], we design these codewords to be dependent among themselves and on the transmitted codewords $X_1, \ldots, X_N$. The key difference from multicoding for single-hop networks is that here multicoding is performed using backward encoding over all blocks and the dependence among the codewords are satisfied simultaneously among them. Unlike partial decode–forward, there is no need for these codewords to have any layered superposition structure. In fact, the scheme does not keep track of which relay recovers exactly which part of the message from which node; relay k recovers *some* part of the message rather implicitly by recovering U_k. The recovered part of the message, captured by an auxiliary index, is then forwarded to the destination nodes in the next block. The details are as follows.

Codebook generation. Fix the pmf $(\prod_{k=2}^N p(x_k)) \cdot p(x_1|x_2^N)p(u_2^N|x^N)$ that attains the maximum in (1). For block $j \in [1:b]$, randomly and independently generate $2^{n\hat{R}_k}$ sequences $x_k^n(l_{k,j-1})$, $l_{k,j-1} \in [1:2^{n\hat{R}_k}]$, each according to $\prod_{i=1}^n p_{X_k}(x_{ki})$, $k \in [2:N]$. For each $l_{k,j-1} \in [1:2^{n\hat{R}_k}]$, randomly and independently generate $2^{n\hat{R}_k}$ sequences $u_k^n(l_{kj}|l_{k,j-1})$, $l_{kj} \in [1:2^{n\hat{R}_k}]$, each according to $\prod_{i=1}^n p_{U_k|X_k}(u_{ki}|x_{ki}(l_{k,j-1}))$. For each $\mathbf{l}_{j-1} = (l_{2,j-1}, \ldots, l_{N,j-1})$, randomly and independently generate $2^{n(R+\tilde{R})}$ sequences $x_1^n(m_j, t_j|\mathbf{l}_{j-1})$,

Block	1	2	$\cdots$	$b-1$	b
Multicoding	$(t_1, \mathbf{l}_0)$	$\leftarrow (t_2, \mathbf{l}_1)$	$\ldots$	$\leftarrow (t_{b-1}, \mathbf{l}_{b-2})$	$\leftarrow (t_b, \mathbf{l}_{b-1})$
X_1	$x_1^n(m_1, t_1 \| \mathbf{l}_0)$	$x_1^n(m_2, t_2 \| \mathbf{l}_1)$	$\ldots$	$x_1^n(m_{b-1}, t_{b-1} \| \mathbf{l}_{b-2})$	$x_1^n(m_b, t_b \| \mathbf{l}_{b-1})$
Y_k	$u_k^n(l_{k1} \| l_{k0})$	$u_k^n(l_{k2} \| l_{k1})$	$\ldots$	$u_k^n(l_{k,b-1} \| l_{k,b-2})$	$u_k^n(l_{kb} \| l_{k,b-1})$
X_k	$x_k^n(l_{k0})$	$x_k^n(l_{k1})$	$\ldots$	$x_k^n(l_{k,b-2})$	$x_k^n(l_{k,b-1})$
Y_d	$\hat{m}_1$	$\leftarrow (\hat{m}_2, \mathbf{l}_1)$	$\ldots$	$\leftarrow (\hat{m}_{b-1}, \mathbf{l}_{b-2})$	$\leftarrow (\hat{m}_b, \mathbf{l}_{b-1})$

TABLE I
ENCODING AND DECODING OF THE DISTRIBUTED DECODE–FORWARD CODING SCHEME.

$(m_j, t_j) \in [1:2^{nR}] \times [1:2^{n\tilde{R}}]$, each according to $\prod_{i=1}^n p_{X_1|X_2,\ldots,X_N}(x_{1i}|x_{2i}(l_{2,j-1}),\ldots,x_{Ni}(l_{N,j-1}))$.

Encoding. For $j = b, b-1, \ldots, 1$ and for each m_j, find an index tuple $(t_j, \mathbf{l}_{j-1})$ such that

$$(x_1^n(m_j, t_j|\mathbf{l}_{j-1}), x_2^n(l_{2,j-1}), \ldots, x_N^n(l_{N,j-1}),\\ u_2^n(l_{2j}|l_{2,j-1}), \ldots, u_N^n(l_{Nj}|l_{N,j-1})) \in \mathcal{T}_{\epsilon'}^{(n)}$$

successively with the initial condition $l_{2b} = \cdots = l_{Nb} = 1$. If there is more than one such index tuple, select one of them arbitrarily. If there is none, select one from $[1:2^{n\tilde{R}}] \times [1:2^{n\hat{R}_2}] \times \cdots \times [1:2^{n\hat{R}_N}]$ arbitrarily. By direct application of the properties of multivariate typicality [17, Section 2.5], induction on backward encoding, and steps similar to those of the multivariate covering lemma [17, Lemma 8.2], it can be shown that encoding is successfully with high probability if

$$\tilde{R} + \sum_{k=2}^N \hat{R}_k > \sum_{k=2}^N I(U_k; U^{k-1}, X^N | X_k) + \delta(\epsilon'), \quad (3)$$

$$\hat{R}(\mathcal{T}) > \sum_{k\in\mathcal{T}} I(U_k; U(\mathcal{T}_k), X(\mathcal{T})|X_k) + \delta(\epsilon') \quad (4)$$

for all $\mathcal{T} \subseteq [2:N]$, where $\mathcal{T}_k = \mathcal{T} \cap [2:k-1]$.

Before the actual transmission of the messages, we use additional $(N-1)^2$ blocks to transmit each l_{k0} to node $k \in [2:N]$ using multi-hop coding, as in the initialization phase for short-message noisy network coding in [18]. The additional transmission needed for this phase is in the order of $O(nN^2)$, independent of b. Thus, the realized transmission rate converges to R as $b \to \infty$. In the following, we assume that all l_{k0} is known prior to transmission.

To send message m_j in block j, the source node transmits $x_1^n(m_j, t_j|\mathbf{l}_{j-1})$, where $(t_j, \mathbf{l}_{j-1})$ is the chosen index tuple.

Relay encoding. Let $\epsilon > \epsilon'$. At the end of block j, node k finds a *unique* index $\tilde{l}_{kj} \in [1:2^{n\hat{R}_k}]$ such that

$$(u_k^n(\tilde{l}_{kj}|\tilde{l}_{k,j-1}), x_k^n(\tilde{l}_{k,j-1}), y_k^n(j)) \in \mathcal{T}_\epsilon^{(n)}.$$

By the packing lemma [17, Lemma 3.1], this is successful with high probability if

$$\hat{R}_k < I(U_k; Y_k|X_k) - \delta(\epsilon). \quad (5)$$

In block $j+1$, the relay then transmits $x_k^n(\tilde{l}_{kj})$.

Decoding. We use backward decoding. For $j = b, \ldots, 1$, decoder $d \in \mathcal{D}$ finds a unique index tuple $(\hat{m}_j, \hat{\mathbf{l}}_{j-1}) = (\hat{m}_j, \hat{l}_{2,j-1}, \ldots, \hat{l}_{N,j-1})$ with $\hat{l}_{d,j-1} = l_{d,j-1}$ such that

$$(x_1^n(\hat{m}_j, t_j|\hat{\mathbf{l}}_{j-1}), x_2^n(\hat{l}_{2,j-1}), \ldots, x_N^n(\hat{l}_{N,j-1}),\\ u_2^n(\hat{l}_{2j}|\hat{l}_{2,j-1}), \ldots, u_N^n(\hat{l}_{Nj}|\hat{l}_{N,j-1}), y_d^n(j)) \in \mathcal{T}_\epsilon^{(n)}$$

for some $t_j \in [1:2^{n\tilde{R}}]$, successively with the initial condition $\hat{l}_{2b} = \cdots = \hat{l}_{Nb} = 1$. By the independence of the codebooks, the Markov lemma [17, Lemma 12.1], the joint typicality lemma, and induction on backward decoding, this step is successful with high probability if

$$R + \tilde{R} + \hat{R}(\mathcal{T}) < I(X_1, X(\mathcal{T}); U(\mathcal{T}^c), Y_d | X(\mathcal{T}^c))\\ + \sum_{k\in\mathcal{T}} I(U_k; U(\mathcal{T}_k), U(\mathcal{T}^c), X^N, Y_d|X_k) - \delta(\epsilon) \quad (6)$$

for some $\mathcal{T} \subseteq [2:N]$ such that $d \in \mathcal{T}^c$.

By identifying $\mathcal{T} = \{1\} \cup \mathcal{S}$ and eliminating the auxiliary rates $\hat{R}_2, \ldots, \hat{R}_N$, and $\tilde{R}$ from (3)–(6), we obtain the inequalities in (1) and

$$\min\Big[I(X(\mathcal{S}); U(\mathcal{S}^c)|X(\mathcal{S}^c))\\ - \sum_{k\in\mathcal{S}^c} I(U_k; U(\mathcal{S}_k^c), X^N|X_k, Y_k)\Big] \geq 0, \quad (7)$$

where the minimum is over all $\mathcal{S} \subseteq [1:N]$ such that $1 \in \mathcal{S}$ and $\mathcal{S}^c \cap \mathcal{D} \neq \emptyset$. Finally, noting that the maximizing input pmf must satisfy (7) completes the proof of Theorem 1.

IV. DISCUSSION

Our coding scheme can be adapted to broadcast by merging the role of the auxiliary variables for Marton coding (broadcast) with that of the auxiliary variables for relaying. This results in the following inner bound on the capacity region of a general broadcast network.

Theorem 2. *For the discrete memoryless network $p(y^N|x^N)$ with $N-1$ messages broadcast to receivers 2 through N, a rate tuple $(R_2, \ldots, R_N)$ is achievable if*

$$R(\mathcal{S}^c) < I(X_1, X(\mathcal{S}); U(\mathcal{S}^c)\,|\,X(\mathcal{S}^c)) - \sum_{k \in \mathcal{S}^c} I(U_k; U(\mathcal{S}_k^c), X^N\,|\,X_k, Y_k)$$

for some pmfs $(\prod_{k=2}^n p(x_k))p(x_1, u_2^N|x_2^N)$, where $\mathcal{S}_k^c = \mathcal{S}^c \cap [2:k-1]$.

Once again the source node controls the relays by auxiliary indices L_k that carry some information about *all* messages. For deterministic networks, Theorem 2 simplifies to the same rate expression as the cutset outer bound [15, Eq. (9)] except that input pmfs are of the form $p(x_1|x_2^N)(\prod_{k=2}^n p(x_k))$. For Gaussian networks, Theorem 2 simplifies to the cutset outer bound within $0.5N$ bits per dimension. Details of this coding scheme and the proof of Theorem 2 will be presented elsewhere [19].

It is gratifying to observe that decode–forward and compress–forward—two fundamental coding schemes for relay channels—are now generalized to arbitrary networks by distributed decode–forward (DDF) and noisy network coding (NNC). For *single-message multicast*, both schemes achieve similar performance; both achieve the Gaussian network capacity within a finite gap and the max-flow min-cut capacity for graphical networks.[1]

Operationally, the two coding schemes have several distinct (and somewhat dual) features. In destination-centric NNC, the source and the relays are relatively simple, but the major burden is on the destinations that need to recover the messages and the compression indices from the entire network over multiple blocks. This scheme fits well with (and currently is the only reasonable solution to) general *multiple access* and *multimessage multicast* relay networks. In source-centric DDF, the relays and the destinations are relatively simple, but the source needs to precode dependent codewords for the entire network over multiple blocks. The scheme fits well with (and currently is the only reasonable solution to) general *broadcast* relay networks.

This operational reciprocity in the roles of source and destination for multiple access and broadcast has been well noted by Kannan, Raja, and Viswanath [15], which was the key intuition for their coding scheme that parallels the quantize–map–forward scheme by Avestimehr, Diggavi, and Tse [6]. Compared to these nested multiletter schemes [6], [15], however, NNC and DDF provide a more conclusive example on the duality between multiple access and broadcast—NNC achieves the capacity region of the (single-hop) multiple access channel, while DDF achieves Marton's inner bound for the broadcast channel. It remains to be seen how this duality can be exploited in building scalable schemes for *multiple unicast* beyond single-hop interference channels.

[1]For general deterministic multicast networks, DDF performs better than NNC in general, due to input pmfs of the form $p(x_1|x_2^N)\prod_{k=1}^N p(x_k)$ instead of product pmfs used in NNC.

References

[1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, no. 1, pp. 120–154, 1971.

[2] 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.

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

[4] B. Schein and R. G. Gallager, "The Gaussian parallel relay channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Sorrento, Italy, Jun. 2000, p. 22.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[6] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.

[7] W. Chang, S.-Y. Chung, and Y. H. Lee, "Gaussian relay channel capacity to within a fixed number of bits," 2010, preprint available at http://arxiv.org/abs/1011.5065/.

[8] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[9] M. H. Yassaee and M. R. Aref, "Slepian-Wolf coding over cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3462–3482, 2011.

[10] 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.

[11] 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.

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

[13] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, 1979.

[14] M. Anand and P. R. Kumar, "A digital interface for Gaussian relay and interference networks: Lifting codes from the discrete superposition model," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2548–2564, 2011.

[15] S. Kannan, A. Raja, and P. Viswanath, "Approximately optimal wireless broadcasting," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7154–7167, 2012.

[16] 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.

[17] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[18] J. Hou and G. Kramer, "Short message noisy network coding with a decode–forward option," 2013. [Online]. Available: http://arxiv.org/abs/1304.1692/

[19] S. H. Lim, K. T. Kim, and Y.-H. Kim, "Distributed decode–forward for broadcast," 2014, in preparation.