Fifty-second Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 3, 2014

Partial Decode–Forward Relaying for the Gaussian Two-Hop Relay Network

Jing Li
School of Telecommunications Engineering
Xidian University
Email: jli_8713@hotmail.com

Young-Han Kim
Department of Electrical and Computer Engineering
University of California, San Diego
Email: yhk@ucsd.edu

***Abstract*—The multicast capacity of the Gaussian two-hop relay network with one source, N relays, and L destinations is studied. It is shown that a careful modification of the partial decode–forward coding scheme, in which the relays cooperate through degraded sets of message parts, achieves the cutset upper bound within $\log N$ bits regardless of the channel gains and power constraints. This scheme improves upon a previous scheme by Chern and Özgür, which is also based on partial decode–forward yet has an unbounded gap from the cutset bound for $L \geq 2$ destinations. As an added benefit, the achievable rate of the proposed scheme involves evaluating rate for $L(N+1)$ cuts out of the total $L2^N$ possible cuts.**

I. Introduction

Consider the Gaussian two-hop relay network with one source, N relays, and L destinations as depicted in Fig. 1, which is a cascade of a broadcast channel (BC) from the source to the relays and multiple multiple access channels (MAC) from the relays to the destinations. The source (node 0) wishes to reliably communicate a common message to the L destination nodes (nodes $N+1, \ldots, N+L$) with help of the relays (nodes $1, \ldots, N$). The special case of $L = 1$, often referred to as the diamond network, was introduced by Schein and Gallager [1], [2]. The capacity is not known in general except for the trivial case of $L = N = 1$.

The best known capacity upper bound is the cutset bound [3], which is the minimum cut capacity over all possible cuts that separate the source and the destinations. There are several capacity lower bounds based on different coding schemes. The compress–forward scheme for the relay channel by Cover and El Gamal [4] have been extended to relay networks in several forms, such as quantize–map–forward (QMF) by Avestimehr, Diggavi, and Tse [5], and noisy network coding (NNC) [6], [7]. The standard analysis [6] shows that when specialized to our channel model in Fig. 1, these coding schemes achieve the cutset bound within $O(N)$ bits for any channel parameters. More recently, Chern and Özgür [8] provided a more refined study on the performance of NNC and showed that it is within $O(\log N)$ bits from the cutset bound regardless of the number of destinations.

In the same paper [8], Chern and Özgür extended the partial decode–forward (PDF) scheme for the relay channel by Cover and El Gamal [4] to the Gaussian diamond network ($L = 1$). In the scheme by Chern and Özgür, the source broadcasts independent parts of the message to the relays, who then decode and forward their corresponding parts to the destination over the MAC. Thus, the Chern–Özgür scheme can achieve the rate characterized by the intersection of the BC capacity region and the MAC capacity region, which can be shown to be within $O(\log N)$ bits from the cutset bound. However, the gap from the cutset bound becomes unbounded when there are more than one destination node.

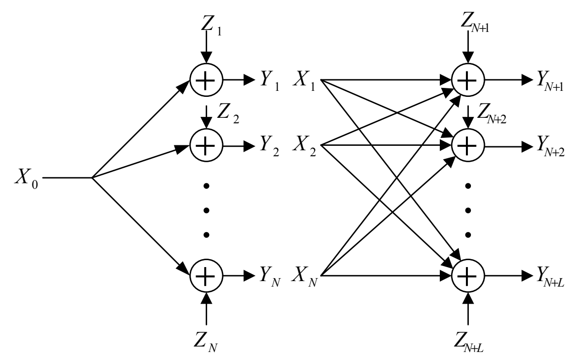

Fig. 1. The Gaussian two-hop relay network.

In this paper, we develop another variant of partial decode–forward that achieves the cutset bound within $O(\log N)$ bits for any number of destination nodes. In the proposed scheme, the relays decode for multiple message parts based on their respective decoding capabilities (as in the degraded BC with degraded message sets [9]) and forward these parts cooperatively (as in the MAC with degraded message sets [10], [11]). Thus, the proposed scheme achieves the rate characterized by the intersection of the capacity region of the BC with degraded message sets and the capacity regions of the group of multiple access channels with degraded message sets.

Although this improvement may be viewed at first as an unnatural complication (except for the obvious benefit for multicasting), it actually yields a simpler characterization of the achievable rate and hence a simpler approximation of the capacity. Note that direct computation of achievable rates for NNC and the Chern–Özgür scheme requires evaluating the cut rate over $L2^N$ different cuts and then taking the minimum. It was shown in [12] that the cut rates are submodular functions; therefore the minimization can be solved in polynomial time (see also [13]). In this paper, we show that the achievable rate of the proposed scheme involves evaluating only $L(N+1)$ cut rates and therefore computation of the achievable rate as well as approximate computation of the capacity (and the cutset bound) takes linear time complexity.

The rest of the paper is organized as follows. In the next section, we formally define the problem of the Gaussian two-hop relay network. In Section III, we review the cutset upper bound on the capacity, which will be benchmarked throughput. In Section IV, we present our coding scheme for the special case of the diamond network ($L=1$) and characterize the achievable rate. Finally, we extend this result to the general L-destination case in Section V.

Throughout the paper, we mostly follow the notation in [14]. In particular, we denote $[1:N] := \{1, 2, \cdots, N\}$. A tuple of random variables is denoted as $X(\mathcal{S}) := (X_k : k \in \mathcal{S})$. The Gaussian capacity function is defined as $\mathsf{C}(x) := (1/2)\log(1+x)$.

II. Formal Statement of the Problem

Recall the Gaussian two-hop relay network depicted in Fig. 1. Here node $j \in [0:N] := \{0\} \cup [1:N]$ transmits the signal X_j, respectively. The received signals at the relays corresponding to the signal X_0 transmitted from the source node are

$$Y_j = g_j X_0 + Z_j, \quad j \in [1:N],$$

where $g_1, \ldots, g_N$ are the channel gains from node 0 to node 1 through N, respectively, and $Z_1, \ldots, Z_N$ are independent $\mathrm{N}(0,1)$ noise components. We assume without loss of generality that

$$|g_1| \geq |g_2| \geq \cdots \geq |g_N|. \tag{1}$$

Similarly, the received signals at the destinations corresponding to the signals $X_1, \ldots, X_N$ transmitted from the relays are

$$Y_{N+d} = \sum_{j=1}^{N} h_{dj} X_j + Z_{N+d}, \quad d \in [1:L],$$

where h_{dj}, $j \in [1:N]$, $d \in [1:L]$, are the channel gains and $Z_{N+1}, \ldots, Z_{N+L}$ are independent $\mathrm{N}(0,1)$ noise components. The first-hop of the network (source-to-relays) can be viewed as a Gaussian broadcast channel, while the second-hop of the network (relays-to-destinations) can be viewed as multiple Gaussian multiple access channels. All nodes are subject to (expected) average power constraint P, and we denote by $S_j = g_j^2 P$ and $S_{dj} = h_{dj}^2 P$ the received signal-to-noise ratios (SNRs) at the relays and the receivers, respectively.

A $(2^{nR}, n)$ code for a Gaussian two-hop relay network consists of

- a message set $[1:2^{nR}]$,
- an source encoder that assigns a codeword $x_0^n(m)$ to each message $m \in [1:2^{nR}]$,
- a set of relay encoders, where encoder $j \in [1:N]$ assigns a symbol $x_{it}(y_i^{t-1})$ to each past received sequence y_i^{t-1} for each time $t \in [1:n]$, and
- a set of decoders, where decoder $d \in [N+1:N+L]$ assigns an estimate $\hat{m}_d$ or an error message e to each received sequence y_j^n.

We assume that the message M is uniformly distributed over the message set. The average probability of error is defined as $P_e^{(n)} = \mathsf{P}\{\hat{M}_d \neq M \text{ for some } d \in [N+1:N+L]\}$. A rate R is said to be achievable for the Gaussian two-hop relay network if there exists a sequence of $(2^{nR}, n)$ codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$. The capacity C is the supremum of all achievable rates.

III. The Cutset Bound on the Capacity

Since the network consists of two noninteracting layers, the cutset bound [3] on the capacity of a general unicast network can be simplified as

$$\begin{aligned} C &\leq R_{\mathrm{CS}} \\ &:= \sup_F \min_{d,\mathcal{S}} I(X_0; Y(\mathcal{S}^c)) + I(X(\mathcal{S}); Y_{N+d} | X(\mathcal{S}^c)), \end{aligned} \tag{2}$$

where the supremum is over all joint distributions $F(x_0)F(x^N)$ satisfying $\mathsf{E}(X_j^2) \leq P$, $j \in [0:N]$, the minimum is over all $d \in [1:L]$ and $\mathcal{S} \subseteq [1:N]$, and $\mathcal{S}^c$ denotes $[1:N] \setminus \mathcal{S}$. By the maximum differential entropy lemma (see, for example, [14, Section 2.2]), the supremum in (2) is attained by Gaussian X_0 and jointly Gaussian $(X_1, \ldots, X_N)$. By switching the order of the supremum (over Gaussian distributions) and the minimum, the cutset bound is upper bounded as

$$\begin{aligned} R_{\mathrm{CS}} &\leq \min_{d,\mathcal{S}} \sup_F I(X_0; Y(\mathcal{S}^c)) + I(X(\mathcal{S}); Y_{N+d} | X(\mathcal{S}^c)) \\ &= \min_{d,\mathcal{S}} \mathsf{C}\Big(\sum_{j\in\mathcal{S}^c} S_j\Big) + \mathsf{C}\Big(\Big(\sum_{j\in\mathcal{S}} \sqrt{S_{d,j}}\Big)^2\Big). \end{aligned} \tag{3}$$

Note that direct computation of (3) (of (3) for a fixed distribution) involves evaluation of the minimum over the combination of 2^N choices of $\mathcal{S}$ and L choices of d, that is, the total $L2^N$ cuts that separate the source and destinations.

IV. Partial Decode–Forward for the Gaussian Diamond Network

For simplicity, we first consider the case $L=1$. In the partial decode–forward scheme by Chern and Özgür [8], the source node divides the message M into N independent parts $M_1, \ldots, M_N$ (rate splitting), relay j recovers M_j and forwards it (decode–forward), and the destination node forms the estimates of $M_1, \ldots, M_N$, and thus of M itself; see Fig. 2. This scheme is implemented over two hops in a block Markov fashion, and the achievable rate can be characterized as

$$R_{\mathrm{PDF}} = \max\Big\{\sum_{j=1}^{N} R_j : (R_1, \ldots, R_N) \in \mathscr{R}_{\mathrm{BC}} \cap \mathscr{R}_{\mathrm{MAC}}\Big\},$$

where $\mathscr{R}_{\mathrm{BC}}$ is the capacity region of the standard N-receiver Gaussian broadcast channel with SNRs $S_1, \ldots, S_N$, that is, the set of rate tuples $(R_1, \ldots, R_N)$ such that

$$R_j \leq \mathsf{C}\Big(\frac{\alpha_j S_j}{\sum_{k=1}^{j-1} \alpha_k S_k + 1}\Big), \quad j \in [1:N], \tag{4}$$

for some $(\alpha_1, \ldots, \alpha_N)$ satisfying $\alpha_j \geq 0$, $j \in [1:N]$, and $\sum_{j=1}^{N} \alpha_j = 1$, and $\mathscr{R}_{\mathrm{MAC}}$ is the capacity region of

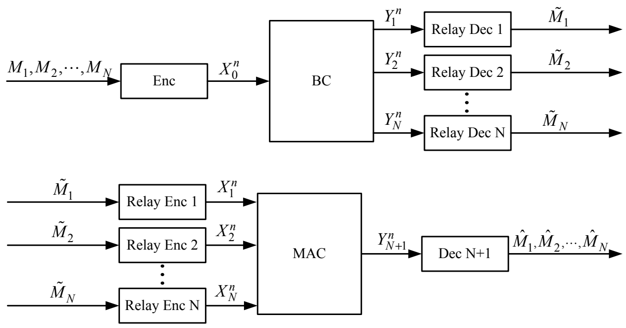

Fig. 2. The partial decode–forward coding scheme by Chern and Özgür.

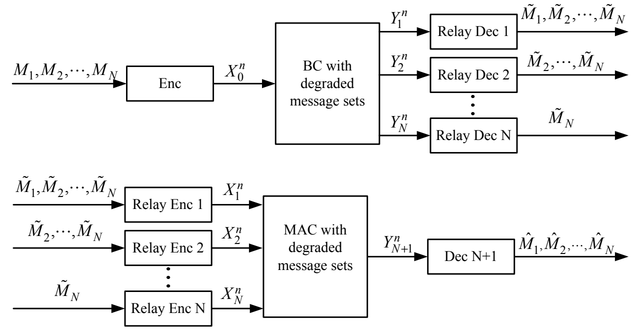

Fig. 3. The modified partial decode–forward scheme.

the standard N-sender Gaussian multiple access channel with SNRs $S_{11}, \dots, S_{1N}$, i.e., the set of rate tuples $(R_1, \dots, R_N)$ such that

$$R(\mathcal{S}) := \sum_{j \in \mathcal{S}} R_j \le \mathsf{C}\Bigl(\sum_{j \in \mathcal{S}} S_{1,j}\Bigr), \quad \mathcal{S} \subseteq [1:N].$$

The region $\mathscr{R}_{\mathrm{MAC}}$ is a polymatroid [15], but the region $\mathscr{R}_{\mathrm{BC}}$ is not in general. Consequently, the maximum sum-rate of the intersection of the two regions is rather cumbersome to calculate. Chern and Özgür relaxed $\mathscr{R}_{\mathrm{BC}}$ into a polymatroidal inner bound, and used Edmonds's polymatroid intersection theorem to find the maximum sum-rate of the intersection. The achievable rate thus computed can be shown to satisfy

$$\Delta_{\mathrm{PDF}} = R_{\mathrm{CS}} - R_{\mathrm{PDF}} \le O(\log N),$$

regardless of S_j and S_{1k}, $j, k \in [1:N]$.

We propose a modified partial decode–forward (MPDF) scheme as depicted in Fig. 3. Here, the relays recover degraded sets of the message parts in the natural order—recall the assumption on the channel gains in (1)—say, relay 1 recovers $M_1, \dots, M_N$; relay 2 recovers $M_2, \dots, M_N$; and so on. The relays then cooperatively communicate these message parts as in the multiple access channel with degraded message sets [10], [11]. The achievable rate can be characterized as

$$R_{\mathrm{MPDF}} = \max\Bigl\{\sum_{j=1}^{N} R_j \colon (R_1, \dots, R_N) \in \mathscr{R}_{\text{BC-DMS}} \cap \mathscr{R}_{\text{MAC-DMS}}\Bigr\},$$

where $\mathscr{R}_{\text{BC-DMS}}$ is the capacity region of the standard N-receiver Gaussian broadcast channel (BC) with degraded message sets and $\mathscr{R}_{\text{MAC-DMS}}$ is the capacity region of the N-sender Gaussian multiple access channel (MAC) with degraded message sets. Since the broadcast channel is degraded in the order of $1 \to 2 \to \cdots \to N$, $\mathscr{R}_{\text{BC-DMS}} = \mathscr{R}_{\mathrm{BC}}$ as in (4). The capacity region of the multiple access channel with degraded message sets [10], [11] consists of all rate tuples $(R_1, \dots, R_N)$ such that

$$\sum_{j=1}^{k} R_j \le I(X^k; Y_{N+1} | X_{k+1}^N), \quad k \in [1:N], \tag{5}$$

for some $F(X_1, \dots, X_N)$ such that $\mathsf{E}(X_j^2) \le P$, $j \in [1:N]$. Again by the maximum differential entropy lemma, there is no loss of generality in setting $(X_1, \dots, X_N)$ to be jointly Gaussian in (5).

As before, exact calculation of the maximum sum-rate of the intersection of $\mathscr{R}_{\mathrm{BC}}$ and $\mathscr{R}_{\text{MAC-DMS}}$ is difficult. We instead consider polymatroidal relaxations of the two regions. For $\mathscr{R}_{\mathrm{BC}}$, we follow the same argument as in [8]. We first use the BC–MAC duality [16] and represent the region as the union of the capacity regions of Gaussian MACs with SNRs $\alpha_1 S_1, \dots, \alpha_N S_N$ over all $\alpha_j \ge 0$, $j \in [1:N]$, such that $\sum_{j=1}^{N} \alpha_j = 1$. By setting $\alpha_j \equiv 1/N$, we obtain a polymatroidal inner bound $\mathscr{R}'_{\mathrm{BC}}$ on $\mathscr{R}_{\mathrm{BC}}$, which consists of all rate tuples $(R_1, \dots, R_N)$ such that

$$R(\mathcal{S}) \le \mathsf{C}\Bigl(\frac{1}{N} \sum_{j \in \mathcal{S}} S_j\Bigr), \quad \mathcal{S} \subseteq [1:N]. \tag{6}$$

As for $\mathscr{R}_{\text{MAC-DMS}}$, we restrict $F(X_1, \dots, X_N)$ in (5) to be independent and identically distributed $X_j \sim \mathrm{N}(0, P)$, $j \in [1:N]$. The resulting inner bound $\mathscr{R}'_{\text{MAC-DMS}}$ on $\mathscr{R}_{\text{MAC-DMS}}$, which consists of all rate tuples $(R_1, \dots, R_N)$ such that

$$\sum_{j=1}^{k} R_j \le \mathsf{C}\Bigl(\sum_{j=1}^{k} S_{1j}\Bigr), \quad k \in [1:N]. \tag{7}$$

It can be readily checked that this inner bound is a polymatroid, since the system of inequalities in (7) can be rewritten as

$$R(\mathcal{S}) \le \mathsf{C}\Bigl(\sum_{j=1}^{\max(\mathcal{S})} S_{1j}\Bigr), \quad \mathcal{S} \subseteq [1:N], \tag{8}$$

by adding inactive inequalities. Now by Edmonds's polymatroid intersection theorem [17], we have

$$\begin{aligned} R_{\mathrm{MPDF}} &\ge \max\Bigl\{\sum_{j=1}^{N} R_j \colon (R_1, \dots, R_N) \in \mathscr{R}'_{\mathrm{BC}} \cap \mathscr{R}'_{\text{MAC-DMS}}\Bigr\} \\ &= \min_{\mathcal{S} \subseteq [1:N]} \mathsf{C}\Bigl(\frac{1}{N} \sum_{j \in \mathcal{S}^c} S_j\Bigr) + \mathsf{C}\Bigl(\sum_{j=1}^{\max(\mathcal{S})} S_{1j}\Bigr) \end{aligned} \tag{9}$$

$$\stackrel{(a)}{=} \min_{k\in[0:N]} \mathsf{C}\biggl(\frac{1}{N}\sum_{j=k+1}^{N} S_j\biggr) + \mathsf{C}\biggl(\sum_{j=1}^{k} S_{1j}\biggr), \tag{10}$$

where the last equality follows since the minimum in (9) is attained by $\mathcal{S} = \emptyset$ or $\mathcal{S} = [1:k]$ for some k. In comparison, by setting $\mathcal{S}$ to be of the form $[1:k]$ in (3), the cutset upper bound can be further relaxed as

$$\begin{aligned} R_{\mathrm{CS}} &\le \min_{k\in[0:N]} \mathsf{C}\biggl(\sum_{j=k+1}^{N} S_j\biggr) + \mathsf{C}\biggl(\Bigl(\sum_{j=1}^{k} \sqrt{S_{1j}}\Bigr)^2\biggr) \\ &\le \min_{k\in[0:N]} \mathsf{C}\biggl(\sum_{j=k+1}^{N} S_j\biggr) + \mathsf{C}\biggl(N\sum_{j=1}^{k} S_{1j}\biggr). \end{aligned} \tag{11}$$

By comparing (10) and (11), we finally establish the following.

Theorem 1. $\Delta_{\mathrm{MPDF}} := R_{\mathrm{CS}} - R_{\mathrm{MPDF}} \le \log N$.

Note that computation of (10) or (11) involves minimization over $N+1$ values. Consequently, the cutset bound as well as the capacity can be computed approximately within $\log N$ in $O(N)$ complexity by considering only $N+1$ cuts rather than all 2^N possible cuts as in (2) and (3).

V. Extension to Multiple Destinations

The advantage of the modified partial decode–forward coding scheme is fully realized when there are multiple destinations ($L \ge 2$), in which case the scheme by Chern and Özgür has an unbounded gap from the capacity [8, Section VI]. In the modified partial decode–forward scheme, the communication is viewed as a cascade of a BC (with degraded message sets) and multiple MACs with degraded message sets. The achievable rate can be characterized as

$$\begin{aligned} &R_{\mathrm{MPDF}} \\ &= \max\biggl\{\sum_{j=1}^{N} R_j : (R_1,\ldots,R_N) \in \mathscr{R}_{\mathrm{BC}} \cap \mathscr{R}_{\text{MMAC-DMS}}\biggr\}, \end{aligned}$$

where $\mathscr{R}_{\text{MMAC-DMS}}$ is the capacity region of the N-sender L-receiver Gaussian multiple MAC with degraded message sets, which is identical to the capacity region of the N-sender L-state Gaussian compound MAC with degraded message sets, that is, the set of rate tuples $(R_1,\ldots,R_N)$ such that

$$\sum_{j=1}^{k} R_j \le \min_{d\in[1:L]} I(X^k; Y_{N+d} | X_{k+1}^N), \quad k \in [1:N], \tag{12}$$

for some jointly Gaussian X^N with $\mathsf{E}(X_j^2) \le P$, $j \in [1:N]$.

We can now proceed in the exactly same manner as in the previous section, except that in place of (8) we use

$$R(\mathcal{S}) \le \min_{d\in[1:L]} \mathsf{C}\biggl(\sum_{j=1}^{\max(\mathcal{S})} S_{dj}\biggr), \quad \mathcal{S} \subseteq [1:N], \tag{13}$$

Consequently, we can lower bound the achievable rate of the scheme as

$$R_{\mathrm{MPDF}} \ge \min_{d\in[1:L]} \min_{k\in[0:N]} \mathsf{C}\biggl(\frac{1}{N}\sum_{j=k+1}^{N} S_j\biggr) + \mathsf{C}\biggl(\sum_{j=1}^{k} S_{dj}\biggr), \tag{14}$$

Starting with (3) and following the same argument as before, we can relax the cutset upper bound as

$$R_{\mathrm{CS}} \le \min_{d\in[1:L]} \min_{k\in[0:N]} \mathsf{C}\biggl(\sum_{j=k+1}^{N} S_j\biggr) + \mathsf{C}\biggl(N\sum_{j=1}^{k} S_{dj}\biggr). \tag{15}$$

This establishes the following.

Theorem 2. $\Delta_{\mathrm{MPDF}} \le \log N$.

A few remarks are in order.

1) The gap from the cutset bound does not depend on the number L of destinations.
2) The gap of $\log N$ is somewhat conservative. In particular, the modified PDF scheme in full generality allows for coherent transmission among the relays, which achieves the rate that is comparable to the second term in (2).
3) Computation of (14) or (15) involves minimization of rates over $L(N+1)$ cuts instead of all $L2^N$ cuts. Thus, the capacity and the cutset bound can be approximated within $\log N$ in $O(LN)$ complexity (cf. [12], [13]).

References

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

[2] B. Schein, "Distributed coordonation in network information theory," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, USA, Oct. 2001.

[3] A. El Gamal, "On information flow in relay networks," in *Proc. IEEE National Telecomm. Conf.*, New Orleans, LA, Nov. 1981, vol. 2, pp. D4.1.1–D4.1.4.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] B. Chern and A. Özgür, "Achieving the capacity of the n-relay Gaussian diamond network within $\log n$ bits," in *Proc. IEEE Inf. Theory Workshop*, Lausanne, Switzerland, Sep 2012, pp. 377–380.

[9] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 279–280, 1974.

[10] V. V. Prelov, "Transmission over a multiple-access channel with a special source hierarchy," *Probl. Peredachi Inf.*, vol. 20, no. 4, pp. 3–10, 1984.

[11] T. S. Han, "The capacity region of general multiple-access channel with certain correlated sources," *Inf. Control*, vol. 40, no. 1, pp. 37–60, 1979.

[12] F. Parvaresh and R. H. Etkin, "On computing the capacity of relay networks in polynomial time," in *Proc. IEEE Int. Symp. Inf. Theory*, Palo Alto, Carlifornia, July/Aug. 2011, pp. 1342–1346.

[13] S. Brahma, A. Sengupta, and C. Fragouli, "Efficient subnetwork selection in relay networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, june/july 2014, pp. 1927–1931.

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

[15] A. Schrijver, *Combinatorial Optimization.* 3 vols. Berlin: Springer-Verlag, 2003.

[16] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.

[17] J. Edmonds, "Submodular functions, matroids, and certain polyhedra," in *Combinatorial Structures and their Applications*, 1970, pp. 69–87.