Cooperative Strategies and Capacity Theorems for Relay Networks

Gerhard Kramer, Michael Gastpar, and Piyush Gupta

***Abstract*— Coding strategies that exploit terminal cooperation are developed for relay networks. Two basic schemes are studied: the relays multi-hop the source message to the destination, or they transmit compressed channel outputs to the destination. Strategies that mix these schemes are also considered. The multi-hopping is done in a sophisticated way: the transmitters cooperate and each receiver uses several or all of its past channel output blocks to decode. For compression, the relays take advantage of the statistical dependence between their channel outputs and the destination's channel output. The strategies are applied to several wireless channels, and it is shown that one can approach capacity if the terminals form two closely spaced clusters. One can further achieve the ergodic capacity with phase fading if the relays are in a region near the source terminal, and if phase information is available only locally. The ergodic capacity results generalize to multi-antenna transmission with Rayleigh fading, single-bounce fading, certain quasistatic fading problems, cases where partial channel knowledge is available at the transmitters, and cases where local user cooperation is permitted. The results further extend to multi-source networks such as multi-access and broadcast relay channels.**

***Index Terms*— antenna arrays, capacity, coding, multi-user channels, relay channels**

I. INTRODUCTION

Relay channels model problems where one or more relays help a pair of terminals communicate. This might occur, for example, in a multi-hop or sensor network where terminals have limited power to transmit data. We briefly summarize the history of information theory for such channels, as well as some recent developments concerning coding strategies.

A model for relay channels was introduced and studied by van der Meulen in [1], [2] (see also [3, Sec. IX]). Two fundamental coding strategies for a single relay were developed by Cover and El Gamal [4, Thm. 1 and Thm. 6]. A combination of these strategies [4, Thm. 7] achieves capacity for several classes of channels, as discussed in [4]–[7]. Capacity-achieving codes appeared in [8] for deterministic relay channels, and in [9], [10] for "permuting" relay channels with states or memory.

We will consider only *random* coding, and concentrate on generalizing the two basic strategies in [4]. The first strategy achieves the rates in [4, Thm. 1], and it uses block Markov superposition encoding, random partitioning (binning) and *successive* decoding. The encoding is done using codebooks of different sizes, and we call this *irregular* block Markov encoding.

G. Kramer and P. Gupta are with Bell Laboratories, Lucent Technologies, 600 Mountain Ave, Murray Hill NJ 07974 (email: gkr@research.bell-labs.com, pgupta@research.bell-labs.com). M. Gastpar is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720-1770 (email: gastpar@eecs.berkeley.edu).

Two alternatives to irregular encoding/successive decoding were developed in the context of the *multi-access channel with generalized feedback* (MAC-GF) studied by King [11]. This channel has three terminals like a single-relay channel, but now two of the terminals transmit messages to the third terminal, e.g., terminals 1 and 2 transmit the respective messages M_1 and M_2 to terminal 3. Terminals 1 and 2 further receive a *common* channel output Y^* that can be different than terminal 3's output Y. King developed an achievable rate region for this channel that generalizes results of Slepian and Wolf [12], Gaarder and Wolf [13], and Cover and Leung [14].

Carleial extended King's model by giving the transmitters *different* channel outputs Y_1 and Y_2 [15], and it is this model that we call a MAC-GF. Carleial further designed a strategy and derived an achievable rate region with 17 bounds [15, eq. (7a)-(7q)]. Although this region can be difficult to evaluate, there are several interesting features of the approach. First, the model includes the relay channel as a special case by making M_2 have zero rate and by setting $Y_1 = 0$ (note that King's version of the relay channel requires $Y_1 = Y_2$ [11, p. 36]). Second, Carleial achieves the same rates as in [4, Thm. 1] by appropriately choosing the random variables in [15, eq. (7)], i.e., choose $V_1 = U_2 = V_2 = W_2 = 0$ and $X_2 = W_1$. This is remarkable because Carleial's strategy is different than Cover and El Gamal's: the transmitter and relay codebooks have the *same* size, and the receiver employs a sliding *window decoding* technique that uses two consecutive blocks of channel outputs [15, p. 842]. A descriptive name for this strategy might be *regular* encoding/*window* decoding.

Yet a third relaying strategy is based on work for the MAC-GF by Willems [16, Ch. 7]. Willems designed an encoding technique that seems more powerful than Carleial's in general, but for the relay channel his encoders are basically the same as in [15]. Moreover, instead of using window decoding, Willems introduced a *backward decoding* technique. The resulting regular encoding/backward decoding method achieves the same rates as irregular encoding/successive decoding and regular encoding/window decoding. Backward decoding does, however, incur a substantial decoding delay.

Subsequent work focused on generalizing these strategies to multiple relays. Irregular encoding/successive decoding was extended to degraded relay networks by Aref [5, Ch. 4]. Aref further developed binning strategies for deterministic broadcast relay networks and deterministic relay networks without interference. For each of these networks, the corresponding strategy was shown to achieve capacity by applying a (then new) cut-set bound [5, p. 23]. This bound generalizes to networks with many messages [17, p. 445] and has become a standard tool for bounding capacity regions.

More recently, the paper [18] sparked a renewed interest in network information theory for wireless channels. Gupta and Kumar also applied the irregular encoding/successive decoding technique to multi-relay networks in [19]. They further extended this method to multi-source networks by associating one or more feedforward flowgraphs with every message (each of these flowgraphs can be interpreted as a "generalized path" in a graph representing the network [19, p. 1883]). We interpret their relaying approach, and that of [5, Ch. 4], as a *multi-hopping* strategy. By this we mean that the source message is decoded successively by the relays, and finally by the destination. We remark that, in contrast to many other multi-hopping schemes, the transmitters *cooperate* and each receiver uses *several or all* of its past channel output blocks to decode, and not only its most recent one.

Regular encoding/window decoding was developed for multiple relays by Xie and Kumar [20], [21], and one can similarly generalize regular encoding/backward decoding [22]. It is interesting to note that the rates of the two regular encoding strategies are the same, and this rate is better than that of [5], [19] for two or more relays. Regular encoding/window decoding is therefore currently the preferred multi-hopping strategy since it achieves the best rates in the simplest way.

Consider next the second basic strategy of Cover and El Gamal that achieves the rates given by [4, Thm. 6]. Instead of multi-hopping, the relays transmit compressed versions of their channel outputs to the destination. The relays further use the statistical dependence between these outputs and the destination's channel output. More precisely, the relays use Wyner-Ziv source coding to exploit side information at the destination [23]. This approach was generalized to the MAC-GF in [11, Ch. 3], and to multiple relays in [24] by adding partial decoding at the transmitters or relays. One can, of course, also mix the coding methods described above (irregular/regular encoding, successive/window/backward/partial decoding).

This paper extends several of the above strategies to relay networks with many terminals, antennas and sources. We further determine new capacity theorems for additive white Gaussian noise (AWGN) relay channels. The paper is divided into two main parts. The first part deals with general relay channels and includes Sections II to V. In Section II, we define the network model and review a capacity upper bound. Section III develops the multi-hopping strategies, also known as *decode-and-forward* strategies, and generalizes them to multi-access relay channels (MARCs) and broadcast relay channels (BRCs). Section IV extends the *compress-and-forward* strategy of [4, Thm. 6] to multiple relays. Section V describes a mixed strategy where each relay uses either decode-and-forward or compress-and-forward, and refines this strategy to include partial decoding.

The second part of the paper is Section VI that specializes the information theory to wireless networks with geometries (distances) and fading. We begin by showing that the mixed strategy of Section V achieves capacity when the terminals form two closely spaced clusters. We next consider channels with phase or Rayleigh fading, and where phase information is available only locally. We show that the decode-and-forward strategy achieves the ergodic capacity when all relays are in a region near the source terminal. The capacity results generalize to certain quasistatic models, and to MARCs and BRCs. Section VII concludes the paper.

We remark that, due to a surge of interest in relay channels, we cannot do justice to all the recent advances in the area here. For example, we do not discuss *cooperative diversity* that is treated in [25]–[31]. Many other results can be found in [32]–[52] and references therein. In particular, Schein developed several decode-and-forward, compress-and-forward, and amplify-and-forward strategies for a two-relay network in [32], [33]. His model is, however, somewhat restrictive in that there is no direct link between the source and destination. This has the advantage of simplifying the theory because transmission strategies do not need to deal with interference at the relays.

II. PRELIMINARIES

A. Abstract Model

We consider the network model of [5, p. 9]. The T-terminal relay network has a source terminal (terminal 1), $T-2$ relays (terminals t with $t \in \mathcal{T} = \{2,3,\ldots,T-1\}$), and a destination terminal (terminal T). The network random variables are: the message W, the channel inputs X_{ti}, $t = 1,2,\ldots,T-1$, $i = 1,2,\ldots n$, the channel outputs Y_{ti}, $t = 2,3,\ldots,T$, $i = 1,2,\ldots n$, and the message estimate $\hat{W}$. The X_{1i} are a function of W, and the X_{ti} are functions of terminal t's past outputs $Y_t^{i-1} = (Y_{t1}, Y_{t2}, \ldots, Y_{t(i-1)})$. The networks we consider are *memoryless* and *time invariant* in the sense that

$$\begin{aligned} &p(y_{2i},\ldots,y_{Ti}|w,x_1^i,\ldots,x_{T-1}^i,y_2^{i-1},\ldots,y_T^{i-1}) \\ &= p_{Y_2\ldots Y_T|X_1\ldots X_{T-1}}(y_{2i},\ldots,y_{Ti}|x_{1i},\ldots,x_{(T-1)i}), \end{aligned} \quad (1)$$

for all i, where $p(a|b)$ is the conditional probability that $A=a$ given $B=b$, and where the X_t and Y_t, $t=1,\ldots,T$, are random variables representing the respective channel inputs and outputs. As in (1), we adopt the convention of dropping subscripts on probability distributions when the arguments are lowercase versions of the random variables. The condition (1) lets one focus on the channel distribution

$$p(y_2,\ldots,y_T|x_1,\ldots,x_{T-1}) \quad (2)$$

for further analysis.

The destination computes its message estimate $\hat{W}$ as a function of Y_T^n. Suppose that W has B_W bits. The *capacity* C is the supremum of rates $R = B_W/n$ at which the destination's message estimate $\hat{W}$ can be made to satisfy $\Pr(\hat{W} \neq W) < \epsilon$ for any positive ϵ.

B. Capacity Upper Bound

Let $X_{\mathcal{S}} = \{X_t : t \in \mathcal{S}\}$. A capacity upper bound is given by the cut-set bound in [5, p. 23] (see also [17, p. 445]).

Proposition 1: The T-terminal relay network capacity satisfies

$$C \le \max_{p(x_1,x_2,\ldots,x_{T-1})} \min_{\mathcal{S}\subseteq\mathcal{T}} I(X_1 X_{\mathcal{S}}; Y_{\mathcal{S}^C} Y_T \mid X_{\mathcal{S}^C}) \quad (3)$$

where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{T}$. For example, for $T = 3$ we have

$$C \leq \max_{p(x_1,x_2)} \min\left\{I(X_1;Y_2Y_3|X_2),\ I(X_1X_2;Y_3)\right\}. \quad (4)$$

Remark 1: The set of $p_{X_1 \ldots X_{T-1}}(\cdot)$ is convex, and the mutual informations in (3) are concave in $p_{X_1 \ldots X_{T-1}}(\cdot)$ [17, p. 31]. Furthermore, the point-wise minimum of a collection of concave functions is concave [53, p. 35]. One can thus perform the maximization in (3) efficiently with convex optimization algorithms (see [47]).

III. DECODE-AND-FORWARD

The multi-hopping strategies have as a common feature that the source controls what the relays transmit. For wireless networks, one consequently achieves gains related to multi-antenna *transmission*. The strategy has been named "decode-and-forward" in [28], or simply decode-forward, and we label the corresponding rates R_{DF}.

A. Single Relay Rates

We interpret the strategy of Cover and El Gamal [4, Thm. 1] as a decode-forward strategy. Again, however, we emphasize that in addition to the usual multi-hopping, the transmitters cooperate and each receiver uses several or all of its past channel output blocks to decode. The strategy achieves any rate up to

$$R_{DF} = \max_{p(x_1,x_2)} \min\left\{I(X_1;Y_2|X_2),\ I(X_1X_2;Y_3)\right\}. \quad (5)$$

The difference between (4) and (5) is that Y_3 is included in the first information on the right hand side of (4).

Remark 2: We can apply Remark 1 to (5), i.e., convex optimization algorithms can efficiently perform the maximization over $p(x_1, x_2)$.

Remark 3: Suppose we have a wireless network. The second mutual information in (5) can be interpreted as the information between two transmit antennas X_1 and X_2, and one receive antenna Y_3 [24], [33, p. 15]. Decode-forward also achieves a cooperative gain reflected by the maximization over all *joint* distributions $p(x_1, x_2)$.

Remark 4: The rate (5) requires the relay to decode the source message, and this can be a rather severe constraint. For example, consider the network of discrete memoryless channels (DMCs) shown in Fig. 1. The channel inputs are $X_1 = [X_{11}, X_{12}]$ and X_2, and the outputs are Y_2 and $Y_3 = [Y_{31}, Y_{32}]$. Suppose that X_{11}, X_{12} and X_2 are binary, and that $Y_2 = X_{11}$, $Y_{31} = X_2$ and $Y_{32} = X_{12}$. The capacity is clearly 2 bits per use, but (5) gives only 1 bit per use.

Remark 5: One can generalize (5) by allowing the relay to *partially decode* the message. This is done in [4, Thm. 7] and [6] by introducing a random variable, say U, that represents the information decoded by the relay. The strategy of [6] (which is a special case of [4, Thm. 7]) achieves rates up to

$$R_{PDF} = \min\left\{I(U;Y_2|X_2) + I(X_1;Y_3|X_2U), I(X_1X_2;Y_3)\right\} \quad (6)$$

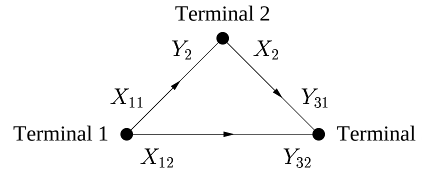

Fig. 1. A network of DMCs.

Block 1	Block 2	Block 3	Block 4
$\underline{x}_1(1, w_1)$	$\underline{x}_1(w_1, w_2)$	$\underline{x}_1(w_2, w_3)$	$\underline{x}_1(w_3, 1)$
$\underline{x}_2(1)$	$\underline{x}_2(w_1)$	$\underline{x}_2(w_2)$	$\underline{x}_2(w_3)$

Fig. 2. A two-hop strategy for the single-relay network.

where $p(u, x_1, x_2)$ is arbitrary up to the alphabet constraints on X_1 and X_2. For instance, choosing $U = X_1$ gives (5). Moreover, the rate (6) is the capacity of the network in Fig. 1 and Remark 4 by choosing U, X_{12}, and X_2 as independent coin-flipping random variables, and $X_{11} = U$.

B. Three Strategies for a Single Relay

The rate (5) has in the past been achieved with three different methods, as discussed in the introduction. We refer to [4, Thm. 1] for a description of the irregular encoding/successive decoding strategy. We instead review the regular encoding approach of [15], [16] that is depicted in Fig. 2.

The message w is divided into B blocks $w_1, w_2, \ldots, w_B$ of 2^{nR} bits each. The transmission is performed in $B + 1$ blocks by using codewords $\underline{x}_1(i, j)$ and $\underline{x}_2(i)$ of length n, where i and j range from 1 to 2^{nR}. The $\underline{x}_1(i, j)$, $j = 1, 2, \ldots, 2^{nR}$, can be "correlated" with $\underline{x}_2(i)$. For example, for real alphabet channels one might choose

$$\underline{x}_1(i, j) = \alpha\, \underline{x}_2(i) + \beta\, \underline{x}_1'(j) \quad (7)$$

where α and β are scaling constants, and where the $\underline{x}_1'(j)$, $j = 1, 2, \ldots, 2^{nR}$, form a separate code book.

Continuing with the strategy, in the first block terminal 1 transmits $\underline{x}_1(1, w_1)$ and terminal 2 transmits $\underline{x}_2(1)$. The receivers use either maximum likelihood or typical sequence decoders. Random coding arguments guarantee that terminal 2 can decode reliably as long as n is large and

$$0 \leq R < I(X_1;Y_2|X_2) \quad (8)$$

where we assume that $I(X_1;Y_2|X_2)$ is positive. We similarly assume the other mutual informations below are positive. So suppose terminal 2 correctly obtains w_1. In the second block, terminal 1 transmits $\underline{x}_1(w_1, w_2)$ and terminal 2 transmits $\underline{x}_2(w_1)$. Terminal 2 can decode w_2 reliably as long as n is large and (8) is true. One continues in this way until block $B + 1$. In this last block, terminal 1 transmits $\underline{x}_1(w_B, 1)$ and terminal 2 transmits $\underline{x}_2(w_B)$.

Consider now the destination (terminal 3), and let $\underline{y}_{3b}$ be its bth block of channel outputs. Suppose these blocks are collected until the last block of transmission is completed. Terminal 3 can then perform Willems' *backward decoding* by first decoding w_B from $\underline{y}_{3(B+1)}$. Note that $\underline{y}_{3(B+1)}$ depends on $\underline{x}_1(w_B, 1)$ and $\underline{x}_2(w_B)$, which in turn depend only on w_B. One can show (see [16, Ch. 7]) that terminal 3 can decode reliably as long as n is large and

$$0 \leq R < I(X_1 X_2; Y_3). \tag{9}$$

Terminal 3 next decodes w_{B-1} from $\underline{y}_{3B}$ which depends on $\underline{x}_1(w_{B-1}, w_B)$ and $\underline{x}_2(w_{B-1})$. Since terminal 3 knows w_B, it can again decode reliably as long as (9) is true. One continues in this fashion until all message blocks have been decoded. The overall rate is $R \cdot B/(B+1)$ bits per use, and by making B large we can get the rate as close to R as desired.

Finally, we describe Carleial's sliding *window decoding* technique [15]. Consider again Fig. 2, but suppose terminal 3 decodes w_1 *after block 2* by using a window of the two past received blocks $\underline{y}_{31}$ and $\underline{y}_{32}$. One can again show (see [15]) that terminal 3 can decode reliably as long as n is large and

$$0 \leq R < I(X_1 X_2; Y_3). \tag{10}$$

The mutual information (10) has a contribution of $I(X_2; Y_3)$ from $\underline{y}_{32}$, and $I(X_1; Y_3|X_2)$ from $\underline{y}_{31}$. After receiving $\underline{y}_{3b}$, $b \geq 3$, Terminal 3 similarly decodes w_{b-1} by using $\underline{y}_{3(b-1)}$ and $\underline{y}_{3b}$, all the while assuming its past message estimate $\hat{w}_{b-2}^{(3)}$ is w_{b-2}. The overall rate is again $R \cdot B/(B+1)$ bits per use, and by making B large we can get the rate as close to R as desired.

Remark 6: Window decoding enjoys the advantages of both the Cover/El Gamal strategy (two block decoding delay) and the Willems strategy (regular block Markov encoding). Furthermore, regular encoding and window decoding are easy to extend to multiple relays.

C. Multiple Relays

A natural first approach to multi-hop with several relays is to generalize the irregular encoding/successive decoding strategy. This was done in [5], [19]. However, we here consider only the improved strategy of [20], [21].

Consider two relays. We divide the message w into B blocks of 2^{nR} bits each. The transmission is performed in $B+2$ blocks by using $\underline{x}_1(i,j,k)$, $\underline{x}_2(i,j)$, and $\underline{x}_3(i)$, where i, j, k range from 1 to 2^{nR}. For example, the encoding for $B = 6$ is depicted in Fig. 3. Terminal 2 can reliably decode w_b after transmission block b if n is large, its past message estimates $\hat{w}_{b-2}^{(2)}, \hat{w}_{b-1}^{(2)}$ were correct, and

$$0 \leq R < I(X_1; Y_2|X_2 X_3). \tag{11}$$

Terminal 3 decodes w_{b-1} by using $\underline{y}_{3(b-1)}$ and $\underline{y}_{3b}$. This can be done reliably if n is large, its past message estimates $\hat{w}_{b-3}^{(3)}$, $\hat{w}_{b-2}^{(3)}$ were correct, and

$$0 \leq R < I(X_1 X_2; Y_3|X_3). \tag{12}$$

The mutual information (12) has a contribution of $I(X_2; Y_3|X_3)$ from $\underline{y}_{3b}$, and a contribution of $I(X_1; Y_3|X_2 X_3)$ from $\underline{y}_{3(b-1)}$. Assuming correct decoding, Terminal 3 knows w_{b-1} after transmission block b, and can encode the messages as shown in Fig. 3.

Block 1	Block 2	Block 3	Block 4
$\underline{x}_1(1,1,w_1)$	$\underline{x}_1(1,w_1,w_2)$	$\underline{x}_1(w_1,w_2,w_3)$	$\underline{x}_1(w_2,w_3,w_4)$
$\underline{x}_2(1,1)$	$\underline{x}_2(1,w_1)$	$\underline{x}_2(w_1,w_2)$	$\underline{x}_2(w_2,w_3)$
$\underline{x}_3(1)$	$\underline{x}_3(1)$	$\underline{x}_3(w_1)$	$\underline{x}_3(w_2)$

Block 5	Block 6	Block 7	Block 8
$\underline{x}_1(w_3,w_4,w_5)$	$\underline{x}_1(w_4,w_5,w_6)$	$\underline{x}_1(w_5,w_6,1)$	$\underline{x}_1(w_6,1,1)$
$\underline{x}_2(w_3,w_4)$	$\underline{x}_2(w_4,w_5)$	$\underline{x}_2(w_5,w_6)$	$\underline{x}_2(w_6,1)$
$\underline{x}_3(w_3)$	$\underline{x}_3(w_4)$	$\underline{x}_3(w_5)$	$\underline{x}_3(w_6)$

Fig. 3. A multi-hopping strategy for the two-relay network.

Finally, terminal 4 decodes w_{b-2} by using $\underline{y}_{4(b-2)}$, $\underline{y}_{4(b-1)}$, and $\underline{y}_{4b}$. This can be done reliably if n is large, its past message estimates $\hat{w}_{b-4}^{(4)}$, $\hat{w}_{b-3}^{(4)}$ were correct, and

$$0 \leq R < I(X_1 X_2 X_3; Y_4). \tag{13}$$

This mutual information has a contribution of $I(X_3; Y_4)$ from $\underline{y}_{4b}$, $I(X_2; Y_4|X_3)$ from $\underline{y}_{4(b-1)}$, and $I(X_1; Y_4|X_2 X_3)$ from $\underline{y}_{4(b-2)}$. The overall rate is $R \cdot B/(B+2)$, so by making B large we can get the rate as close to R as desired.

It is clear that window decoding generalizes to T-terminal relay networks, and one can prove the following theorem. Let $\pi(\cdot)$ be a permutation on $\mathcal{T}$, and define $\pi(1) = 1$, $\pi(T) = T$, and $\pi(i:j) = \{\pi(i), \pi(i+1), \ldots, \pi(j)\}$.

Theorem 1: The T-terminal relay network capacity is at least

$$R_{DF} = \max_{\pi(\cdot)} \min_{1 \leq t \leq T-1} I(X_{\pi(1:t)}; Y_{\pi(t+1)} \mid X_{\pi(t+1:T-1)}) \tag{14}$$

where one can choose any distribution on $(X_1, X_{\mathcal{T}})$.

Remark 7: Theorem 1 is essentially due to Xie and Kumar, and appeared for AWGN channels in [20]. The result appeared for more general classes of channels in [21], [22]. Proofs can be found in [20], [21].

Remark 8: Theorem 1 appeared for *degraded* relay networks in [5, p. 69], where degradation was defined as

$$\begin{aligned} &p(y_{\pi(t)}|x_{\pi(1:T-1)}, y_{\pi(2:t-1)}) \\ &\quad = p(y_{\pi(t)}|x_{\pi(t-1)}, x_{\pi(t)}, y_{\pi(t-1)}) \end{aligned} \tag{15}$$

for $t = 2, 3, \ldots, T$, some permutation $\pi(\cdot)$ on $\mathcal{T}$, and where $\pi(T) = T$ (see [5, p. 54], and also [4, eq. (10)] and [38]). Moreover, R_{DF} is the capacity region of such channels [5, p. 69]. One can, in fact, replace (15) by the more general

$$\begin{aligned} &p(y_{\pi(t:T)}|x_{\pi(1:T-1)}, y_{\pi(t-1)}) \\ &\quad = p(y_{\pi(t:T)}|x_{\pi(t-1:T-1)}, y_{\pi(t-1)}) \end{aligned} \tag{16}$$

for $t = 3, 4, \ldots, T$. The condition (16) simply makes the upper bound of Proposition 1 the same as R_{DF} [21].

Remark 9: We can apply Remark 1 to (14), i.e., convex optimization algorithms can efficiently perform the maximization over $p(x_1, \ldots, x_{T-1})$.

Block 1	Block 2	Block 3	Block 4
$\underline{u}_1(1)$	$\underline{u}_1(w_{11})$	$\underline{u}_1(w_{12})$	$\underline{u}_1(w_{13})$
$\underline{x}_1(1, w_{11})$	$\underline{x}_1(w_{11}, w_{12})$	$\underline{x}_1(w_{12}, w_{13})$	$\underline{x}_1(w_{13}, 1)$
$\underline{u}_2(1)$	$\underline{u}_2(w_{21})$	$\underline{u}_2(w_{22})$	$\underline{u}_2(w_{23})$
$\underline{x}_2(1, w_{21})$	$\underline{x}_2(w_{21}, w_{22})$	$\underline{x}_2(w_{22}, w_{23})$	$\underline{x}_2(w_{23}, 1)$
$\underline{x}_3(1, 1)$	$\underline{x}_3(w_{11}, w_{21})$	$\underline{x}_3(w_{12}, w_{22})$	$\underline{x}_3(w_{13}, w_{23})$

Fig. 4. A multi-hopping strategy for a MARC.

Block 1	Block 2	Block 3
$\underline{u}_0(1, w'_{01})$	$\underline{u}_0(w'_{01}, w'_{02})$	$\underline{u}_0(w'_{02}, w'_{03})$
$\underline{u}_1(1, w'_{01}, s_{11})$	$\underline{u}_1(w'_{01}, w'_{02}, s_{12})$	$\underline{u}_1(w'_{02}, w'_{03}, s_{13})$
$\underline{u}_2(1, w'_{01}, s_{21})$	$\underline{u}_2(w'_{01}, w'_{02}, s_{22})$	$\underline{u}_2(w'_{02}, w'_{03}, s_{23})$
$\underline{x}_1(1, w'_{01}, s_{11}, s_{21})$	$\underline{x}_1(w'_{01}, w'_{02}, s_{12}, s_{22})$	$\underline{x}_1(w'_{02}, w'_{03}, s_{13}, s_{23})$
$\underline{x}_2(1)$	$\underline{x}_2(w'_{01})$	$\underline{x}_2(w'_{02})$

Fig. 5. A multi-hopping strategy for a BRC.

Remark 10: We have expressed Theorem 1 using only permutations rather than the level sets of [19]–[21]. This is because one need consider only permutations to maximize the *rate*, i.e., one need consider only flowgraphs in that have *one* vertex per level set. However, observe that to minimize the *delay* for a given rate, one will need to consider level sets again. This occurs, e.g., if one relay is at the same location as another.

Remark 11: Backward decoding also achieves R_{DF}, but for multiple relays one must transmit using nested blocks to allow the intermediate relays (e.g., terminal 3 in Fig. 3) to decode before the destination. The result is an excessive decoding delay.

Remark 12: Equation (14) illustrates the multi-antenna transmission behavior: the mutual information (14) can be interpreted as the information between t transmit antennas and one receive antenna. The main limitation of the strategy is that only *one* antenna is used to decode. This deficiency is remedied to some extent by the second basic strategy of [4, Thm. 6].

Remark 13: One can generalize Theorem 1 and let the relays perform *partial* decoding (see [4, Thm. 7], [6]). We will not consider this possibility here, but later do consider a restricted form of partial decoding.

D. Multi-source Networks

We demonstrate that regular encoding and window decoding are useful for relay networks with multiple sources. Consider a MARC with two sources [34]. Such a network has four terminals: terminals 1 and 2 transmit the independent messages W_1 and W_2 at rates R_1 and R_2, respectively, terminal 3 acts as a relay, and terminal 4 is the destination for both messages. This model might fit a situation where there are two sensors that are too weak to cooperate, but they can send their data to more powerful terminals that are part of a "backbone" network.

A regular encoding structure is as follows. The message w_t is divided into B blocks $w_{t1}, w_{t2}, \ldots, w_{tB}$ of 2^{nR_t} bits each, $t = 1, 2$. Transmission is performed in $B + 1$ blocks by using codewords $\underline{u}_1(i_1)$, $\underline{x}_1(i_1, j_1)$, $\underline{u}_2(i_2)$, $\underline{x}_2(i_2, j_2)$, and $\underline{x}_3(i_1, i_2)$ of length n, where i_t and j_t range from 1 to 2^{nR_t}. The message-to-codeword mappings are organized as in Fig. 4. The details of the codebook construction, encoding and decoding are given in Appendix B. Summarizing the results, after block b terminal 3 can decode (w_{b1}, w_{b2}) reliably if n is large, its past estimate of $(w_{1(b-1)}, w_{2(b-1)})$ is correct, and

$$\begin{aligned} 0 &\le R_1 < I(X_1; Y_3 | U_1 U_2 X_2 X_3) \\ 0 &\le R_2 < I(X_2; Y_3 | U_1 U_2 X_1 X_3) \\ R_1 + R_2 &< I(X_1 X_2; Y_3 | U_1 U_2 X_3) \end{aligned} \tag{17}$$

where $p(u_1, u_2, x_1, x_2, x_3) = p(u_1, x_1)p(u_2, x_2)p(x_3|u_1, u_2)$. Suppose Terminal 4 uses *backward* decoding. One finds that this terminal can decode reliably if

$$\begin{aligned} 0 &\le R_1 < I(X_1 X_3; Y_4 | U_2 X_2) \\ 0 &\le R_2 < I(X_2 X_3; Y_4 | U_1 X_1) \\ R_1 + R_2 &< I(X_1 X_2 X_3; Y_4). \end{aligned} \tag{18}$$

The achievable rates of (17) and (18) were determined in [34] for AWGN channels by using irregular block Markov encoding, and with x_3 a deterministic function of u_1 and u_2. A cut-set outer bound on the capacity region was also given in [34].

Remark 14: We can use the flowgraphs of [19, Sec. IV] to define other strategies for the MARC. There are two different flowgraphs for each source, namely one that uses the relay and one that does not. Suppose that both sources use the relay. There are then 2 successive decoding orderings for the relay and 6 such orderings for the destination, for a total of 12 strategies. There are 6 more strategies if one source uses the relay and the other does not, and 2 more strategies if neither source uses the relay. There are even more possibilities if one splits each source into two colocated "virtual" sources and performs an optimization over the choice of flowgraphs and the decoding orderings, as was suggested in [19, p. 1883].

Consider next a BRC with four terminals and three independent messages W_0, W_1, W_2. Terminal 1 transmits W_0 at rate R_0 to both terminals 3 and 4, W_1 at rate R_1 to terminal 3, and W_2 at rate R_2 to terminal 4. Terminal 2 acts as a relay. Such a model might fit a scenario where a central node forwards instructions to a number of agents via a relay.

Several block Markov encoding strategies can be defined for BRCs, and one of them is depicted in Fig. 5. The messages w_0, w_1, w_2 are again divided into B blocks w_{0b}, w_{1b}, w_{2b}, respectively, for $b = 1, 2, \ldots, B$. However, to improve rates, for each b these blocks are reorganized into blocks $w'_{0b}, w'_{1b}, w'_{2b}$ such that (w'_{0b}, w'_{1b}) contains the bits of (w_{0b}, w_{1b}), and (w'_{0b}, w'_{2b}) contains the bits of (w_{0b}, w_{2b}). Finally, (w'_{1b}, w'_{2b}) is encoded into a pair of integers (s_{1b}, s_{2b}) such that s_{1b} uniquely determines w'_{1b}, and s_{2b} uniquely determines w'_{2b}.

Decoding proceeds as follows. Terminal 2 decodes w'_{0b} after block b, but terminal 3 waits until block $b + 1$ to decode both w'_{0b} and s_{1b} by using *window* decoding with its channel outputs from blocks b and $b + 1$. Similarly, terminal 4 decodes w'_{0b} and s_{2b} after block $b + 1$. Appendix C outlines an analysis for this decode-forward strategy that is closely related to the theory in [54], [55], [56, p. 391]. Summarizing the results, we have the following theorem.

Theorem 2: The non-negative rate triples (R_0, R_1, R_2) satisfying

$$\begin{aligned} &R_0 < \min(I_2, I_3, I_4) \\ &R_0 + R_1 < \min(I_2, I_3) + I(U_1; Y_3|U_0X_2) \\ &R_0 + R_2 < \min(I_2, I_4) + I(U_2; Y_4|U_0X_2) \\ &R_0 + R_1 + R_2 < \min(I_2, I_3, I_4) + I(U_1; Y_3|U_0X_2) \\ &\qquad + I(U_2; Y_4|U_0X_2) - I(U_1; U_2|U_0X_2) \end{aligned} \tag{19}$$

are in the capacity region of the BRC, where

$$I_2 = I(U_0; Y_2|X_2), I_3 = I(U_0X_2; Y_3), I_4 = I(U_0X_2; Y_4)$$

and where $p(u_0, u_1, u_2, x_1, x_2)$ is arbitrary up to the alphabet constraints on X_1 and X_2. For example, the choice $U_0 = X_1$ and $U_1 = U_2 = 0$ gives

$$\begin{aligned} &R_0 + R_1 + R_2 < \\ &\quad \min\left[I(X_1; Y_2|X_2), I(X_1X_2; Y_3), I(X_1X_2; Y_4)\right]. \end{aligned} \tag{20}$$

Remark 15: The region (19) includes Marton's region [54], [56, p. 391] by turning off the relay and making it colocated with the source. That is, we choose $X_2 = 0$ and $Y_2 = X_1$ so that I_2 is larger than I_3 and I_4.

Remark 16: One can generalize Theorem 2 by letting the relay perform *partial* decoding of U_0. More precisely, suppose that $\tilde{w}'_{0(b-1)}$ is some portion of the bits in $w'_{0(b-1)}$. We first generate a codebook of codewords $\underline{x}_2(\tilde{w}'_{0(b-1)})$; the size of this codebook is smaller than before. We next superpose on each $\underline{x}_2(\tilde{w}'_{0(b-1)})$ a codebook of codewords $\underline{v}(\tilde{w}'_{0(b-1)}, \tilde{w}'_{0b})$ generated by an auxiliary random variable V. Next, we superpose a codebook of codewords $\underline{u}_0(\tilde{w}'_{0(b-1)}, w'_{0b})$, and similarly for $\underline{u}_1$, $\underline{u}_2$ and $\underline{x}_1$. In block b, one thus replaces $w'_{0(b-1)}$ by $\tilde{w}'_{0(b-1)}$.

As yet another approach, the relay might choose to decode all new messages after each block. This seems appropriate if there is a high capacity link between the source and relay. The choice of strategy will, of course, depend on the channel.

IV. Compress-and-Forward

Consider the strategy of [4, Thm. 6] where the relay forwards a compressed version of its channel outputs to the destination. This approach lets one achieve gains related to multi-antenna *reception* [24], [28, p. 64]. The strategy has become known as "compress-and-forward", or simply compress-forward (other authors prefer "observe-and-forward" [28] or "quantize-forward"), and we label the corresponding rates R_{CF}. For a different approach named *collaborative decoding*, we refer to [39].

A. Single Relay

The strategy of [4, Thm. 6] achieves any rate up to

$$R_{CF} = I(X_1; \hat{Y}_2Y_3|X_2) \tag{21}$$

where

$$I(\hat{Y}_2; Y_2|Y_3X_2) \le I(X_2; Y_3) \tag{22}$$

and where where the joint probability distribution of the random variables factors as

$$p(x_1)\, p(x_2)\, p(\hat{y}_2|x_2, y_2)\, p(y_2, y_3|x_1, x_2). \tag{23}$$

Remark 17: Equation (21) illustrates the multi-antenna reception behavior: $I(X_1; \hat{Y}_2Y_3|X_2)$ can be interpreted as the rate for a channel with one transmit antenna and two receive antennas. The multi-antenna gain is limited by (22), which states that the rate used to compress Y_2 must be smaller than the rate from the relay to the destination. The compression uses techniques developed by Wyner and Ziv [23], i.e., it exploits the destination's side information Y_3.

B. Multiple Relays

For multiple relays a multi-access channel (MAC) appears because the relays transmit to the destination simultaneously. Furthermore, the signals observed by the relays are correlated. We thus have the problem of sending correlated sources over a MAC as treated in [57]. However, there are two additional features that do not arise in [57]. First, the destination has channel outputs that are correlated with the relay channel outputs. This situation also arose in [4], and we adopt the methodology of that paper. Second, the relays observe noisy versions of each other's symbols. This situation did not arise in [4] or [33], and we deal with it by using partial decoding.

The compress-forward scheme for two relays operates as follows. During block b terminal 2 receives symbols that depend on both X_1 and X_3. After block b, terminal 2 *partially* decodes terminal 3's codeword and "subtracts" this from Y_2. How much terminal 2 decodes is controlled by choosing auxiliary random variables $U_{\mathcal{T}}$. For example, if $U_3 = X_3$ then terminal 2 will completely decode terminal 3's codewords, while if $U_3 = 0$ then terminal 2 ignores terminal 3.

The symbols Y_2, modified by the "subtraction" using U_3, are next compressed to $\hat{Y}_2$ by using the correlation between Y_2, Y_3 and Y_4, i.e., the compression is performed using Wyner-Ziv coding [23]. However, the relays additionally have the problem of encoding in a distributed fashion [33]. After the compression, terminal 2 transmits a codeword in block $b+1$ from which the destination obtains $\hat{Y}_2$ from block b. Finally, the destination uses its Y_4 during block b together with the codewords it receives from the relays in block $b+1$ to estimate the X_1 of block b.

In Appendix D we derive the rates of such schemes for any number of relays. The result is the following Theorem. Let $\mathcal{S}^C$ be the complement of $\mathcal{S}$ in $\mathcal{T}$, and let $\mathcal{S}_1 \backslash \mathcal{S}_2$ denote $\{s : s \in \mathcal{S}_1, s \notin \mathcal{S}_2\}$.

Theorem 3: Compress-forward achieves any rate up to

$$R_{CF} = I(X_1; \hat{Y}_{\mathcal{T}} Y_T | U_{\mathcal{T}} X_{\mathcal{T}}) \tag{24}$$

where

$$\begin{aligned} &I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | U_{\mathcal{T}} X_{\mathcal{T}} \hat{Y}_{\mathcal{S}^C} Y_T) + \sum_{t \in \mathcal{S}} I(\hat{Y}_t; X_{\mathcal{T}\backslash\{t\}} | U_{\mathcal{T}} X_t) \\ &\le I(X_{\mathcal{S}}; Y_T | U_{\mathcal{S}} X_{\mathcal{S}^C}) + \sum_{m=1}^{M} I(U_{\mathcal{K}_m}; Y_{r(m)} | U_{\mathcal{K}_m^C} X_{r(m)}) \end{aligned} \tag{25}$$

for all $\mathcal{S} \subseteq \mathcal{T}$, all partitions $\{\mathcal{K}_m\}_{m=1}^M$ of $\mathcal{S}$, and all $r(m) \in \{2,3,\ldots,T\}$ such that $r(m) \notin \mathcal{K}_m$. For $r(m) = T$ we set $X_T = 0$. Furthermore, the joint probability distribution of the random variables factors as

$$p(x_1) \left[\prod_{t=2}^{T-1} p(u_t, x_t)\, p(\hat{y}_t | u_{\mathcal{T}}, x_t, y_t) \right] \cdot p(y_2, \ldots, y_T | x_1, \ldots, x_{T-1}). \tag{26}$$

Remark 18: Equation (26) implies that the X_t, $t = 1, \ldots, T-1$, are statistically independent. Equation (24) illustrates the multi-antenna reception behavior: the mutual information can be interpreted as the rate for a channel with one transmit antenna and $T-1$ receive antennas. The multi-antenna gain is limited by (25), which is a combination of source and channel coding bounds.

Remark 19: For $T = 3$ we recover (21)–(22).

Remark 20: For $T = 4$ there are already nine bounds of the form (25): two for $\mathcal{S} = \{2\}$ ($r(1) = 3$ and $r(1) = 4$), two for $\mathcal{S} = \{3\}$ ($r(1) = 2$ and $r(1) = 4$), and five bounds for $\mathcal{S} = \{2,3\}$ ($r(1) = 4$ for $\mathcal{K}_1 = \mathcal{S}$, and otherwise $(r(1), r(2))$ being $(2,3)$, $(2,4)$, $(3,4)$, or $(4,4)$). Clearly, computing the compress-forward rate for large T is tedious.

Remark 21: Suppose the relays perform no partial decoding, i.e., $U_t = 0$ for all t. The bounds (25) simplify because we need not partition $\mathcal{S}$. For example, for $T = 4$ we have

$$\begin{aligned} &I(\hat{Y}_2; Y_2 | X_2 X_3 \hat{Y}_3 Y_4) + I(\hat{Y}_2; X_3 | X_2) \le I(X_2; Y_4 | X_3) \\ &I(\hat{Y}_3; Y_3 | X_2 X_3 \hat{Y}_2 Y_4) + I(\hat{Y}_3; X_2 | X_3) \le I(X_3; Y_4 | X_2) \\ &I(\hat{Y}_2 \hat{Y}_3; Y_2 Y_3 | X_2 X_3 Y_4) + I(\hat{Y}_2; X_3 | X_2) + I(\hat{Y}_3; X_2 | X_3) \\ &\qquad \le I(X_2 X_3; Y_4) \end{aligned} \tag{27}$$

For Schein's parallel relay network, this region reduces to Theorem 3.3.2 in [33, p. 136].

Remark 22: Suppose the relays decode each other's codewords entirely, i.e., $U_t = X_t$ for all $t \in \mathcal{T}$. The relays thereby remove each other's interference before compressing their channel outputs. The bound (25) simplifies to

$$I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{T}} \hat{Y}_{\mathcal{S}^C} Y_T) \le \sum_{m=1}^{M} I(X_{\mathcal{K}_m}; Y_{r(m)} | X_{\mathcal{K}_m^C}). \tag{28}$$

However, there are still nine bounds of this form.

Remark 23: The left hand side of (25) describes rates that generalize results of [23] to distributed source coding with side information at the receiver (see [58]). The right hand side of (25) describes rates achievable on the multi-way channel between the relays and the destination. In other words, an (extended) Wyner-Ziv source coding region must intersect a channel coding region. This approach separates source and channel coding, which will be suboptimal in general (see [59, Ch. 1]).

V. Mixed Strategies

The strategies of Sections III and IV can be combined as in [4, Thm. 7]. However, we consider only the case where each relay chooses either decode-forward or compress-forward. We divide the relay indexes into the two sets

$$\mathcal{T}_1 = \{2, 3, \ldots, T_1 + 1\}, \quad \mathcal{T}_2 = \{T_1 + 2, \ldots, T - 1\}.$$

The relays in $\mathcal{T}_1$ use decode-forward while the relays in $\mathcal{T}_2$ use compress-forward. The result is the following theorem.

Theorem 4: Choosing either decode-forward or compress-forward achieves any rate up to

$$\begin{aligned} R_{DCF} = \min \Big\{ &\min_{1 \le t \le T_1} I(X_{\pi(1:t)}; Y_{\pi(t+1)} \mid X_{\pi(t+1:T_1+1)}) , \\ &I(X_1 X_{\mathcal{T}_1}; \hat{Y}_{\mathcal{T}_2} Y_T | U_{\mathcal{T}_2} X_{\mathcal{T}_2}) \quad \Big\} \end{aligned} \tag{29}$$

where $\pi(\cdot)$ is a permutation on $\mathcal{T}_1$, we set $\pi(1) = 1$, and

$$\begin{aligned} &I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | U_{\mathcal{T}_2} X_{\mathcal{T}_2} \hat{Y}_{\mathcal{S}^C} Y_T) + \sum_{t \in \mathcal{S}} I(\hat{Y}_t; X_{\mathcal{T}_2 \setminus \{t\}} | U_{\mathcal{T}_2} X_t) \\ &\le I(X_{\mathcal{S}}; Y_T | U_{\mathcal{S}} X_{\mathcal{S}^C}) + \sum_{m=1}^{M} I(U_{\mathcal{K}_m}; Y_{r(m)} | U_{\mathcal{K}_m^C} X_{r(m)}) \end{aligned} \tag{30}$$

for all $\mathcal{S} \subseteq \mathcal{T}_2$, all partitions $\{\mathcal{K}_m\}_{m=1}^M$ of $\mathcal{S}$, and all $r(m) \in \mathcal{T}_2 \cup \{T\}$ such that $r(m) \notin \mathcal{K}_m$. We here write $\mathcal{S}^C$ for the complement of $\mathcal{S}$ in $\mathcal{T}_2$. For $r(m) = T$ we set $X_T = 0$. Furthermore, the joint probability distribution of the random variables factors as

$$p(x_1, x_{\mathcal{T}_1}) \cdot \left[\prod_{t \in \mathcal{T}_2} p(u_t, x_t)\, p(\hat{y}_t | u_{\mathcal{T}_2}, x_t, y_t) \right] \cdot p(y_2, \ldots, y_T | x_1, \ldots, x_{T-1}). \tag{31}$$

As an example, consider $T = 4$, $T_1 = 1$, and $U_2 = X_2$ (or $U_2 = 0$). We find that Theorem 4 simplifies to

$$R_{DCF} = \min \left\{ I(X_1; Y_2 | X_2),\ I(X_1 X_2; \hat{Y}_3 Y_4 | X_3) \right\} \tag{32}$$

where

$$I(\hat{Y}_3; Y_3 | X_3 Y_4) \le I(X_3; Y_4) \tag{33}$$

and the joint probability distribution of the random variables factors as

$$p(x_1, x_2)\, p(x_3)\, p(\hat{y}_3 | x_3, y_3)\, p(y_2, y_3, y_4 | x_1, x_2, x_3). \tag{34}$$

The second information in (32) can be interpreted as the rate for a 2×2 multi-antenna system. Hence, when terminal 2 is close to the source and terminal 3 is close to the destination, we will achieve rates close to the capacities described in [60], [61].

We omit the proof of Theorem 4 because of its similarity to the proofs in [20], [21] and Appendix D. Instead, we supply a proof of Theorem 5 below. This theorem illustrates how one can improve on Theorem 4 by permitting partial decoding at one of the relays.

A. Two Relays and Partial Decoding

Suppose there are two relays, and that terminal 2 uses decode-forward while terminal 3 uses compress-forward. Terminal 3 further partially decodes the signal from terminal 2 before compressing its observation. However, as in Theorem 4, we make X_3 statistically independent of X_1 and X_2. In Appendix E we show that this strategy achieves the following rates.

Theorem 5: For the two-relay network, any rate up to

$$R_{DCF} = \min\{I(X_1;Y_2|U_2X_2), \quad I(X_1X_2;\hat{Y}_3Y_4|U_2X_3) + R_2'\} \tag{35}$$

is achievable, where for some R_2' and R_3 we have

$$R_3 \geq I(\hat{Y}_3;Y_3|U_2X_3Y_4) \tag{36}$$
$$0 \leq R_2' \leq \min\{I(U_2;Y_3|X_3),\ I(U_2;Y_4|X_3)\} \tag{37}$$
$$0 \leq R_3 \leq I(X_3;Y_4|U_2) \tag{38}$$
$$0 \leq R_2' + R_3 \leq I(U_2X_3;Y_4) \tag{39}$$

and where the joint probability distribution of the random variables factors as

$$p(x_1,u_2,x_2)\, p(x_3)\, p(\hat{y}_3|u_2,x_3,y_3)\, p(y_2,y_3,y_4|x_1,x_2,x_3). \tag{40}$$

Remark 24: We recover (32) and (33) with $U_2 = 0$.

Remark 25: We recover (5) by turning off Relay 3 with $U_2 = X_3 = \hat{Y}_3 = 0$ and $R_2' = R_3 = 0$.

Remark 26: We recover (21)–(22) by turning off and ignoring Relay 2 with $U_2 = X_2 = 0$ and $Y_2 = X_1$.

VI. Wireless Models

The wireless channels we will consider have the X_t and Y_t being *vectors*, and we emphasize this by underlining symbols. We further concentrate on channels (2) with

$$\underline{Y}_t = \underline{Z}_t + \sum_{s\neq t} \frac{A_{st}}{\sqrt{d_{st}^{\alpha}}}\, \underline{X}_s \tag{41}$$

where d_{st} is the distance between terminals s and t, α is an attenuation exponent, $\underline{X}_s$ is a $n_s \times 1$ complex vector, A_{st} is a $n_t \times n_s$ matrix whose i,j entry $A_{st}^{(i,j)}$ is a complex fading random variable, and $\underline{Z}_t$ is a $n_t \times 1$ vector whose entries are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean, unit variance, and whose real and imaginary parts are i.i.d. We impose the per-symbol power constraints $\mathrm{E}[\underline{X}_s^\dagger \underline{X}_s] \leq P_s$ for all s, where $\underline{X}_s^\dagger$ is the complex-conjugate transpose of $\underline{X}_s$.

We will consider several kinds of fading:

- No fading: $A_{st}^{(i,j)}$ is a constant for all s, t, i, and j.
- Phase fading: $A_{st}^{(i,j)} = e^{j\theta_{st}^{(i,j)}}$, where $\theta_{st}^{(i,j)}$ is uniformly distributed over $[0, 2\pi)$. The $\theta_{st}^{(i,j)}$ are jointly independent of each other and all other random variables.
- Rayleigh fading: $A_{st}^{(i,j)}$ is a complex, Gaussian random variable with zero mean, unit variance, and whose real and imaginary parts are i.i.d. The $A_{st}^{(i,j)}$ are jointly independent of each other and all other random variables.
- Single-bounce fading: $A_{st} = B_{st}D_{st}C_{st}$, where B_{st} is a random $n_t \times n_{st}$ matrix, D_{st} is a random $n_{st} \times n_{st}$ diagonal matrix whose entries are independent and have phases that are uniformly distributed over $[0, 2\pi)$, and C_{st} is a random $n_{st} \times n_s$ matrix. The B_{st}, D_{st} and C_{st} are jointly independent of each other and all other random variables. The matrices B_{st} and C_{st} might represent knowledge about the directions of arrival and departure, respectively, of plane waves [62].
- Rayleigh fading with directions: $A_{st} = B_{st}G_{st}C_{st}$, where B_{st} is a random $n_t \times n_{st}^{(2)}$ matrix, G_{st} is a $n_{st}^{(2)} \times n_{st}^{(1)}$ complex, Gaussian random matrix whose entries are independent, zero mean, unit variance, and have i.i.d. real and imaginary parts, and C_{st} is a random $n_{st}^{(1)} \times n_s$ matrix. The B_{st}, G_{st} and C_{st} are jointly independent of each other and all other random variables.

We will usually assume that terminal t knows only its *own* fading coefficients. That is, terminal t knows A_{st} for all s, but it does not know $A_{st'}$ for $t' \neq t$. The one exception is the no-fading case where the A_{st} are known by all terminals.

Remark 27: The above model lets the relays transmit and receive at the same time (and in the same bandwidth). This is possible, e.g., if the relay has two antennas: one receiving and one transmitting. When simultaneous transmission and reception is not possible, then one should modify (41) by, say, adding the constraints that $\underline{Y}_t = \underline{0}$ if $\underline{X}_t \neq \underline{0}$ for all t. The remaining analysis is then similar to that described below (cf. [37], [40], [44], [46]), but some of the the capacity results change (see Remark 32). On might also wish to use the strategy of [6] or the flowgraphs of [19] to improve rates.

A. No Fading and One Relay

Suppose we have a single relay, terminals with one antenna, and no fading. Let ρ be the correlation coefficient of X_1 and X_2. For a fixed covariance matrix of X_1 and X_2, a conditional maximum entropy theorem (see [64, Lemma 1]) ensures that both of the differential entropies $h(Y_2Y_3|X_2)$ and $h(Y_3)$ are maximized by making $p(x_1,x_2)$ zero-mean Gaussian. Since $h(Y_2Y_3|X_1X_2)$ and $h(Y_3|X_1X_2)$ are constants, the cut-set bound (4) is

$$C \leq \max_{0\leq\rho\leq 1} \min\left\{\log\left(1 + P_1\left(\frac{1}{d_{12}^{\alpha}} + \frac{1}{d_{13}^{\alpha}}\right)(1-|\rho|^2)\right),\right.$$
$$\left.\log\left(1 + \frac{P_1}{d_{13}^{\alpha}} + \frac{P_2}{d_{23}^{\alpha}} + \frac{2\rho\sqrt{P_1P_2}}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}}\right)\right\} \tag{42}$$

where ρ is real. Similarly, the best decode-forward rate (5) is

$$R_{DF} = \max_{0\leq\rho\leq 1} \min\left\{\log\left(1 + \frac{P_1}{d_{12}^{\alpha}}(1-|\rho|^2)\right),\right.$$
$$\left.\log\left(1 + \frac{P_1}{d_{13}^{\alpha}} + \frac{P_2}{d_{23}^{\alpha}} + \frac{2\rho\sqrt{P_1P_2}}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}}\right)\right\}. \tag{43}$$

Consider next the compress-forward strategy. We choose X_1 and X_2 to be Gaussian, and $\hat{Y}_2 = Y_2 + \hat{Z}_2$ where $\hat{Z}_2$ is a Gaussian random variable with zero-mean, variance $\hat{N}_2$, and that is

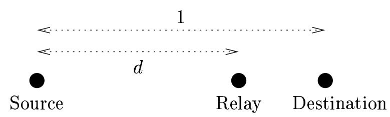

Fig. 6. A single relay on a line.

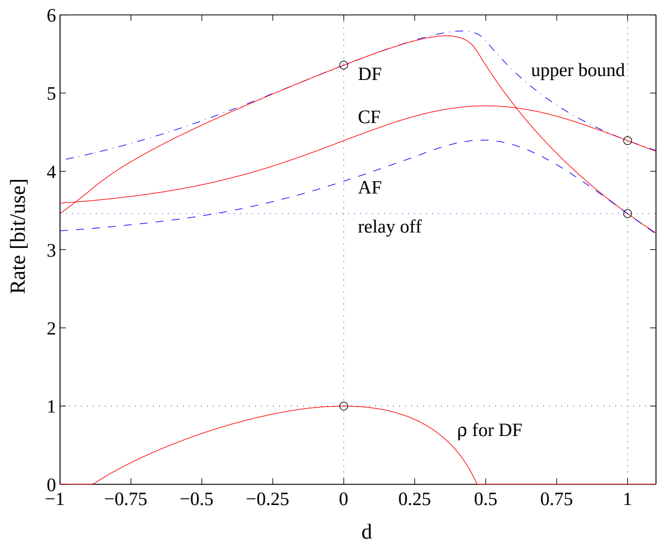

Fig. 7. Rates for a single-relay network with $P_1 = P_2 = 10$ and $\alpha = 2$.

independent of all other random variables. The rate (21) is then

$$R_{CF} = \log\left(1 + \frac{P_1}{d_{12}^{\alpha}(1+\hat{N}_2)} + \frac{P_1}{d_{13}^{\alpha}}\right) \tag{44}$$

where the choice

$$\hat{N}_2 = \frac{P_1(1/d_{12}^{\alpha} + 1/d_{13}^{\alpha}) + 1}{P_2/d_{23}^{\alpha}} \tag{45}$$

satisfies (22) with equality (see also [44, Sec. 3.2] for the same analysis).

As an example, suppose the source, relay and destination are aligned as in Fig. 6, where $d_{12} = d$, $d_{23} = 1 - d$, and $d_{13} = 1$. Fig. 7 plots various bounds for $P_1 = P_2 = 10$ and $\alpha = 2$. The curves labeled DF and CF give the respective decode-forward and compress-forward rates. Also shown are the rates when the relay is turned off, but now only half the power is being consumed as compared to the other cases. Finally, the figure plots rates for the strategy where the relay transmits $X_{2i} = c \cdot Y_{2(i-1)}$, where c is a scaling factor chosen so that $E[|X_{2i}|^2] = P_2$. This strategy is called "amplify-and-forward" in [28, p. 80] (see also [32], [33, p. 61]), and it turns the source to destination channel into a unit–memory intersymbol interference channel. The curve labeled AF shows the capacity of this channel.

Remark 28: As the relay moves toward the source ($d \to 0$), the rates (43) and (44) become

$$\begin{aligned} R_{DF} &= \log\left(1 + P_1 + P_2 + 2\sqrt{P_1 P_2}\right) \\ R_{CF} &= \log\left(1 + P_1 + P_2\right) \end{aligned} \tag{46}$$

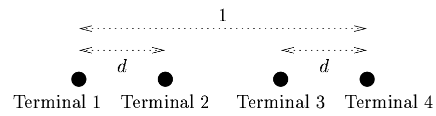

Fig. 8. Two relays on a line.

and R_{DF} is the capacity. Similarly, as the relay moves toward the destination ($d \to 1$), we have

$$\begin{aligned} R_{DF} &= \log\left(1 + P_1\right) \\ R_{CF} &= \log\left(1 + 2P_1\right) \end{aligned} \tag{47}$$

and R_{CF} is the capacity. These limiting capacity results extend to multiple relays, as discussed next.

B. No Fading and Two Relays

Suppose we have two relays, terminals with one antenna, and no fading. Suppose further that the relays are within a distance d of the source. The decode-forward rate of Theorem 1 becomes the capacity as $d \to 0$, which is

$$R_{DF} = \log\left(1 + \left[\sqrt{P_1} + \sqrt{P_2} + \sqrt{P_3}\right]^2\right).$$

This limiting capacity result generalizes to more relays and is called an *antenna clustering capacity* in [24].

Similarly, if the relays are within a distance d of the destination and $d \to 0$, the compress-forward rate of Theorem 3 becomes the capacity

$$R_{CF} = \log\left(1 + 3P_1\right).$$

This limiting capacity result again generalizes to many relays and is another type of antenna clustering capacity.

Finally, consider the geometry in Fig. 8. The mixed strategies of Theorems 4 or 5 achieve capacity as $d \to 0$, which is

$$R_{DCF} = \log\left(1 + 2\left[P_1 + P_2 + 2\sqrt{P_1 P_2}\right]\right).$$

This type of limiting capacity result generalizes to many relays if the T terminals form two closely spaced clusters.

Remark 29: In Fig. 8, we have $R_{DCF} \to 0$ as $d \to 0.5$. This is because terminal 2 does not decode what terminal 3 transmits, so that X_3 acts as interference for terminal 2. This problem could be fixed by adding partial or full decoding at terminal 2 to the strategy of Theorem 5.

C. Phase Fading and One Relay

Consider next phase fading where θ_{st} is known only to terminal t for all s. The result is that R_{DF} in (5) becomes

$$\max_{p(x_1,x_2)} \min\left\{I(X_1;Y_2|X_2\,\theta_{12}),\ I(X_1X_2;Y_3|\,\theta_{13}\,\theta_{23})\right\} \tag{48}$$

where the θ_{st} appear in the conditioning. We have

$$I(X_1;Y_2|X_2\,\theta_{12}) = \int_0^{2\pi} \frac{d\phi}{2\pi} I(X_1;Y_2|X_2,\theta_{12}=\phi) \tag{49}$$

and one can similarly express $I(X_1X_2;Y_3|\theta_{13}\,\theta_{23})$. Further, for a fixed covariance matrix of X_1 and X_2, the conditional maximum entropy theorem [64, Lemma 1] tells us that

$$h(Y_2\,|\,X_2,\theta_{12}=\phi_{12}) \quad \text{and}$$
$$h(Y_3\,|\,\theta_{13}=\phi_{13},\theta_{23}=\phi_{23})$$

are maximized by making $p(x_1,x_2)$ zero-mean Gaussian for any choice of the phases. The maximization (48) is thus

$$\max_{\rho}\ \min\left\{\log\left(1+\frac{P_1}{d_{12}^{\alpha}}(1-|\rho|^2)\right),\int_0^{2\pi}\frac{d\phi_{13}d\phi_{23}}{(2\pi)^2}\right.$$
$$\left.\log\left(1+\frac{P_1}{d_{13}^{\alpha}}+\frac{P_2}{d_{23}^{\alpha}}+\frac{2\,\Re(\rho\,e^{j(\phi_{13}-\phi_{23})})\sqrt{P_1P_2}}{d_{13}^{\alpha/2}\,d_{23}^{\alpha/2}}\right)\right\} \quad (50)$$

where $\Re(x)$ is the real part of x. Let $I(\rho)$ be the integral in (50). This integral does not depend on the phase of ρ, so we have $I(\rho)=[I(\rho)+I(-\rho)]/2$. But $\log(x)$ is concave in x and $\Re(x)$ is linear in x, so Jensen's inequality gives

$$I(\rho)+I(-\rho)]/2 \le I(0). \quad (51)$$

This shows that $\rho=0$ is best for both informations in (50).

Remark 30: An alternative proof is to observe that $I(\rho)=\mathrm{E}[\log(X)]$ for a random variable X with

$$\mathrm{E}[X]=1+P_1/d_{13}^{\alpha}+P_2/d_{23}^{\alpha}.$$

Jensen's inequality now directly gives the desired upper bound $I(\rho)\le I(0)$ (see [44, Lemma 1]). We instead used the proof with (51) because it seems easier to extend to the problems studied in Section VI-E.

Similar arguments show that $\rho=0$ is also best for the capacity upper bound (3) (see Sec. VI-E). This leads to the following theorem by combining (4) and (5).

Theorem 6: Decode-forward achieves capacity with phase fading if the relay is near the source. More precisely, if

$$P_1/d_{13}^{\alpha}+P_2/d_{23}^{\alpha}\le P_1/d_{12}^{\alpha} \quad (52)$$

then the capacity is

$$C=\log\left(1+P_1/d_{13}^{\alpha}+P_2/d_{23}^{\alpha}\right). \quad (53)$$

Remark 31: The optimality of $\rho=0$ for phase fading was also realized in [44, Lemma 1], [47, Sec. 2.3]. The geometric capacity result of Theorem 6 appeared in [22].

The condition (52) is satisfied for a range of d_{12} near zero. For example, for the geometry of Fig. 6 with $\alpha=2$ and $P_1=P_2$, the bound (52) is $-0.883\le d\le 0.469$. Fig. 9 plots the resulting decode-forward and upper bound rates for $P_1=P_2=10$ and a range of d. Fig. 9 also plots the rates of compress-forward when the relay uses $\hat{Y}_2=Y_2+\hat{Z}_2$ as for the no-fading case. In fact, these rates are given by (44) and (45), i.e., they are the same as for the no-fading case. We remark that compress-forward performs well for all d and even achieves capacity for $d=0$.

Consider next a two-dimensional geometry where the source is at the origin and the destination is a distance of 1 to the right of the source. For $P_1=P_2$ the condition (52) is

$$1/d_{13}^{\alpha}+1/d_{23}^{\alpha}\le 1/d_{12}^{\alpha}. \quad (54)$$

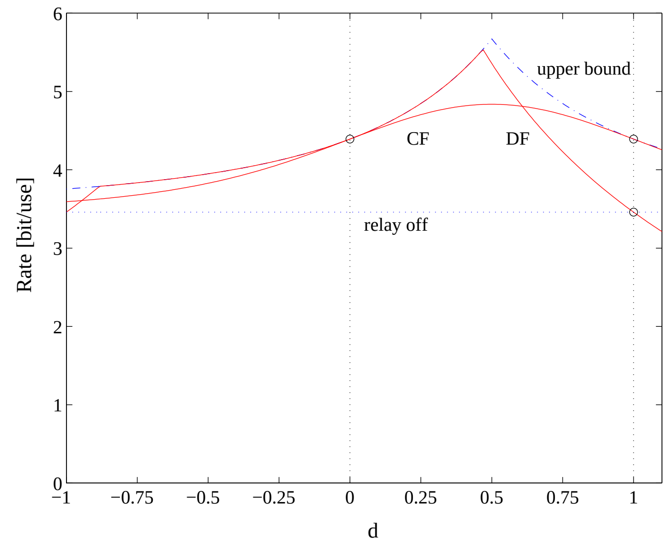

Fig. 9. Rates for a single-relay network with phase fading, $P_1=P_2=10$, and $\alpha=2$.

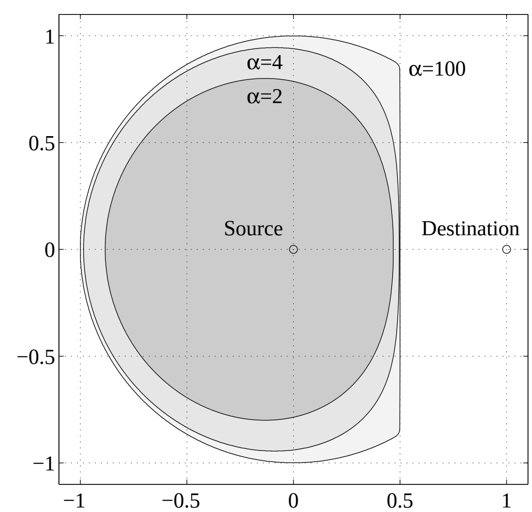

Fig. 10. Relay positions where decode-forward achieves capacity with phase-fading and $P_1=P_2$.

The relay positions that satisfy (54) for $\alpha=2$, $\alpha=4$, and $\alpha=100$ are drawn as the shaded regions in Fig. 10. As α increases, this region expands to become all points inside the circle of radius one around the origin, excepting those points that are closer to the destination than the source.

Remark 32: The capacity results of Theorem 6 (and Theorems 7-11) are no longer valid if the relays cannot transmit and receive at the same time. The reason is that one must introduce a time-sharing parameter, and this parameter will take on different values for the capacity lower and upper bounds. Similarly, the capacity results are not valid if the per-terminal power constraints are replaced by a *network* power constraint, e.g., $\sum_t P_t\le P$. However, we note that if the time-sharing parameter or power levels are restricted to certain ranges, then one can again derive capacity theorems. For instance, this occurs if protocols restrict time to be shared equally between transmission and reception.

D. Phase Fading and Many Relays

Suppose there are T terminals subject to phase fading. Evaluating (3) and (14), we find that it is best to make the X_t, $t = 1, 2, \ldots, T-1$, Gaussian and independent (see Sec. VI-E). We thus have the following generalization of Theorem 6.

Theorem 7: Decode-forward achieves capacity with phase fading if

$$\sum_{t=1}^{T-1} \frac{P_t}{d_{tT}^\alpha} \leq \max_{\pi(\cdot)} \min_{1 \leq s \leq T-2} \sum_{t \in \pi(1:s)} \frac{P_t}{d_{t\pi(s+1)}^\alpha} \tag{55}$$

and the resulting capacity is

$$C = \log\left(1 + \sum_{t=1}^{T-1} \frac{P_t}{d_{tT}^\alpha}\right). \tag{56}$$

Note that the minimization in (55) does not include $s = T-1$.

The condition (55) is satisfied if all relays are near the source. For example, consider a two-dimensional geometry and suppose the relays are in a circle of radius d around the source. Then if the destination is a distance of 1 from the source, we have $d_{tT} \geq 1 - d$. Suppose further that $P_t = P$ for all t. The bound (55) tells us that decode-forward achieves capacity if

$$d \leq \frac{1}{(T-1)^{1/\alpha} + 1}. \tag{57}$$

The relays must therefore be in a circle of radius about $T^{-1/\alpha}$ around the source for large T.

As another geometric example, consider a linear network as in Fig. 6 but with $T-2$ relays placed regularly to the right of the source at $d_{1t} = (t-1)d$, $2 \leq t \leq T-1$, where $0 \leq d < 1/(T-1)$ (see also [20, Sec. 2]). Suppose again that $P_t = P$ for all t. The bound (55) ensures that decode-forward achieves capacity if d satisfies

$$\sum_{t=1}^{T-1} \frac{1}{[1-(t-1)d]^\alpha} \leq \frac{1}{d^\alpha}. \tag{58}$$

Suppose we choose $d = 1/(T-1+\epsilon)$ where ϵ is a positive constant independent of T. We can upper bound the sum in (58) by integrating the function $1/(xd)^\alpha$ from ϵ to ∞. As a result, we find that for any $\alpha > 1$ there is an ϵ satisfying

$$0 < \epsilon \leq \left(\frac{1}{\alpha - 1}\right)^{\frac{1}{\alpha-1}} \tag{59}$$

such that (58) holds with equality. This choice of ϵ gives

$$C = \log\left(1 + (T-1+\epsilon)^\alpha\right) \underset{\text{large } T}{\approx} \alpha \log(T). \tag{60}$$

In other words, capacity grows logarithmically in the number of terminals (or relays). Other related logarithmic scaling laws were obtained in [20] and [36].

E. Phase Fading, Many Relays, and Multiple Antennas

Suppose we have phase fading, many relays, and multiple antennas. Gaussian distributed inputs are again optimal for (3) and (14). Letting A_{all} represent the vector of all matrices A_{st}, we can write and bound the information in (3) as

$$\begin{aligned} & I(\underline{X}_1 \underline{X}_{\mathcal{S}}; \underline{Y}_{\mathcal{S}^C} \underline{Y}_T \mid \underline{X}_{\mathcal{S}^C} A_{all}) \\ & \quad \leq \int_a p(a) \log\left(\left| Q_{\underline{Y}_{\mathcal{S}^C} \underline{Y}_T \mid \underline{X}_{\mathcal{S}^C}, A_{all}=a} \right|\right) da \end{aligned} \tag{61}$$

where $Q_{\underline{A}\,\underline{B}|\underline{C}, D=d}$ is the covariance matrix of the vector $\left[\underline{A}^T \underline{B}^T\right]$ conditioned on $\underline{C}$ and $D = d$, and $|Q|$ is the determinant of Q. The determinant in (61) evaluates to

$$\left| Q_{\underline{\tilde{Y}}_{\mathcal{S}^C} \underline{\tilde{Y}}_T | \underline{X}_{\mathcal{S}^C}} \right| \leq \left| Q_{\underline{\tilde{Y}}_{\mathcal{S}^C} \underline{\tilde{Y}}_T} \right| \tag{62}$$

where

$$\underline{\tilde{Y}}_t = \underline{Z}_t + \sum_{s \in \{1\} \cup \mathcal{S}} \frac{a_{st}}{\sqrt{d_{st}^\alpha}} \underline{X}_s \tag{63}$$

and where equality holds in (62) if $\underline{X}_{\{1\}\cup\mathcal{S}}$ and $\underline{X}_{\mathcal{S}^C}$ are independent.

Observe that we can replace A_{1t} with $-A_{1t}$ for all t, because it is immaterial what phase we begin integrating from for any entry in A_{1t}. Moreover, this is equivalent to using the same A_{1t} as originally, but replacing $\underline{X}_1$ with $-\underline{X}_1$. But this change makes all cross-correlation matrices $\mathrm{E}\left[\underline{X}_1 \underline{X}_s^\dagger\right]$ with $s \neq 1$ change sign. We can thus use the concavity of $\log(|A|)$ in positive semi-definite A of the same size, and apply Jensen's inequality to show that the mutual informations cannot decrease if we make $\mathrm{E}[\underline{X}_1 \underline{X}_s^\dagger] = 0$ for all $s \neq 1$. $\underline{X}_1$ and $\underline{X}_{\mathcal{T}}$ are therefore independent. Repeating these steps for all input vectors, the best input distribution has *independent* $\underline{X}_t$, $t = 1, 2, \ldots, T-1$. This implies that equality holds in (62).

We next determine the best $Q_{\underline{X}_1}$. Observe that, by the same argument as above, we can replace the first columns of the A_{1t} with their negative counterparts. This is equivalent to using the same A_{1t} as originally, but replacing the first entry $\underline{X}_1^{(1)}$ of $\underline{X}_1$ with $-\underline{X}_1^{(1)}$. This in turn makes the entries of the first row and column of $Q_{\underline{X}_1}$ change sign, with the exception of the diagonal element. Applying Jensen's inequality, we find that $\underline{X}_1^{(1)}$ should be independent of the remaining $\underline{X}_1^{(i)}$, $i = 2, 3, \ldots, n_1$. Repeating these steps for all the entries of all the $\underline{X}_t$, we find that the best $Q_{\underline{X}_t}$ are *diagonal*.

Finally, we can permute the diagonal elements of the $Q_{\underline{X}_t}$ without changing the mutual informations. Applying Jensen's inequality, we find that the best input distributions have

$$Q_{\underline{X}_t} = \sqrt{\frac{P_t}{n_t}}\, I \tag{64}$$

where I is the appropriately sized identity matrix. We find that (64) is also optimal for Rayleigh fading by repeating the above arguments. This gives the following theorem.

Theorem 8: Decode-forward achieves capacity with phase or Rayleigh fading if the choice $\mathcal{S} = \mathcal{T}$ minimizes (3), where the $\underline{X}_t$ are independent, Gaussian, and have covariance matrix (64). This requirement is satisfied, e.g., if all the relays are in a region near the source terminal. The resulting capacity is

$$C = \int_a p(a) \log \left(\left| I + \sum_{t=1}^{T-1} \frac{P_t}{n_t} \frac{a_{tT} a_{tT}^\dagger}{d_{tT}^\alpha} \right| \right) da. \quad (65)$$

Remark 33: Theorem 8 appeared in [51]. The capacity result was also derived in [52] for one relay ($T = 3$) and Rayleigh fading by using the above proof technique.

For example, suppose we have phase fading with one relay, $n_1 = n_2 = 1$, and $n_3 = 2$, i.e., the destination has two antennas. Equation (65) is then

$$\int_0^{2\pi} \int_0^{2\pi} \log \left(\left| I + \frac{P_1}{d_{13}^\alpha} \begin{bmatrix} 1 & e^{j\phi_1} \\ e^{-j\phi_1} & 1 \end{bmatrix} + \frac{P_2}{d_{23}^\alpha} \begin{bmatrix} 1 & e^{j\phi_2} \\ e^{-j\phi_2} & 1 \end{bmatrix} \right| \right) \frac{d\phi_1\, d\phi_2}{4\pi^2} = \log \left(\frac{a + \sqrt{a^2 - b^2}}{2} \right) \quad (66)$$

where

$$a = 1 + \frac{2P_1}{d_{13}^\alpha} + \frac{2P_2}{d_{23}^\alpha} + \frac{2P_1 P_2}{d_{13}^\alpha d_{23}^\alpha}, \quad b = \frac{2P_1 P_2}{d_{13}^\alpha d_{23}^\alpha}. \quad (67)$$

For the geometry of Fig. 6 with $\alpha = 2$ and $P_1 = P_2$, capacity is therefore achieved if

$$\frac{a + \sqrt{a^2 - b^2}}{2} \le 1 + \frac{P_1}{d_{12}^\alpha}. \quad (68)$$

Again, this condition is satisfied for a range of d_{12} near zero. Fig. 11 plots the decode-forward and upper bound rates for $P_1 = P_2 = 10$ and the same range of d as in Fig. 9. We have also plotted the decode-forward rates from Fig. 9, and the compress-forward rates when the relay again uses $\hat{Y}_2 = Y_2 + \hat{Z}_2$. The compress-forward rates are now

$$R_{CF} = \log \left(1 + \frac{P_1}{d_{12}^\alpha (1 + \hat{N}_2)} + \frac{2P_1}{d_{13}^\alpha} \right) \quad (69)$$

where the choice

$$\hat{N}_2 = \frac{P_1 (1/d_{12}^\alpha + 2/d_{13}^\alpha) + 1}{\left(a + \sqrt{a^2 - b^2} \right)/2 - (2P_1/d_{13}^\alpha + 1)} \quad (70)$$

satisfies (22) with equality. Compress-forward again performs well for all d and achieves capacity for $d = 0$ and $d = 1$.

Fig. 12 plots the relay positions that satisfy (68) for the same geometry as in Fig. 10. Again, as α increases this region expands to become all points inside the circle of radius one around the origin, except those points that are closer to the destination than the source. Observe further that the known-capacity regions are much smaller than for $n_3 = 1$ when α is small. At the same time, the rates with $n_3 = 2$ are much larger than with $n_3 = 1$.

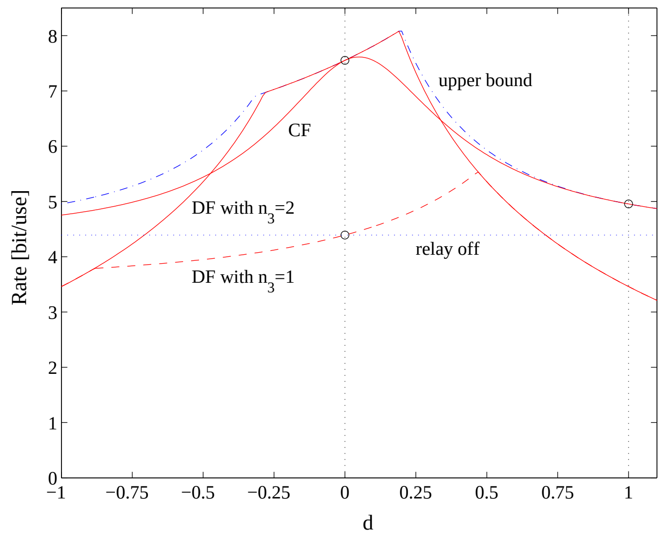

Fig. 11. Rates for a single-relay network with phase fading, $n_1 = n_2 = 1$, $n_3 = 2$, $P_1 = P_2 = 10$, and $\alpha = 2$.

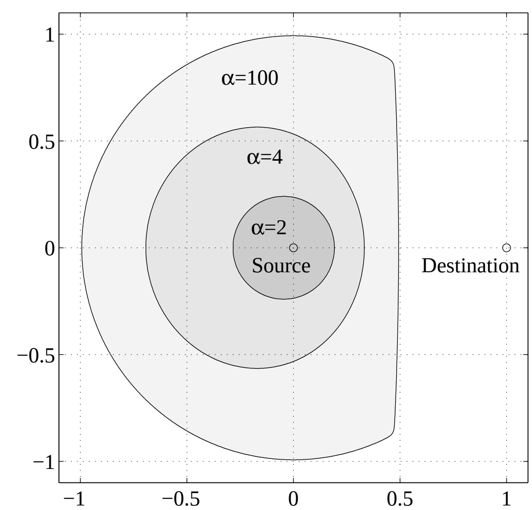

Fig. 12. Relay positions where decode-forward achieves capacity with phase-fading, $n_1 = n_2 = 1$, $n_3 = 2$, and $P_1 = P_2 = 10$.

F. Fading with Directions

Single-bounce and Rayleigh fading with directions can be dealt with as above, and the best $\underline{X}_t$ for (3) and (14) are again Gaussian and independent. However, now the best $Q_{\underline{X}_t}$ are not necessarily given by (64), and they might not be diagonal. For example, suppose we have Rayleigh fading where B_{st} and C_{st} are fixed (deterministic) matrices. Suppose the singular value decomposition of C_{st} is $V_{st} D_{st} U_{st}^\dagger$, where V_{st} and $U_{st}^\dagger$ are unitary, and where D_{st} is nonnegative with positive values along its diagonal only [65, p. 414]. We can absorb V_{st} into the matrix G_{st} by applying [61, Lemma 5]. One might now guess that the optimal $Q_{\underline{X}_s}$ has the form $Q_{\underline{X}_s} = U_{st} D_s$ where D_s is diagonal. However, since each $\underline{X}_s$ goes through multiple U_{st}, it is unclear what the best choice of $Q_{\underline{X}_s}$ should be. Nevertheless, the capacity is again achieved if all relays are in a region near the source terminal, because the best choice of $Q_{\underline{X}_s}$ will be the same for both (3) and (14).

There are, of course, some simple cases where we can say more. Suppose there is single-bounce fading, the nonzero entries of D_{st} are $e^{j\theta_{st}^{(i,i)}}$, $n_{st}=n_s$, $C_{st}=I$, and B_{st} is fixed (deterministic) for all s. We find that (64) is best and (65) simplifies to

$$C=\log\left(\left|I+\sum_{t=1}^{T-1}\frac{P_t}{n_t}\,\frac{B_{tT}B_{tT}^{\dagger}}{d_{tT}^{\alpha}}\right|\right). \tag{71}$$

We illustrate the behavior of C with the following simple generalization of Theorem 7.

Theorem 9: Consider single-bounce fading with $n_t=n_{st}=n_s=n_1$, $B_{st}=I$, $\left|D_{st}^{(i,i)}\right|=1$, and $C_{st}=I$ for all s,t,i. Decode-forward achieves capacity if

$$\sum_{t=1}^{T-1}\frac{P_t}{d_{tT}^{\alpha}}\le\max_{\pi(\cdot)}\min_{1\le s\le T-2}\sum_{t\in\pi(1:s)}\frac{P_t}{d_{t\pi(s+1)}^{\alpha}} \tag{72}$$

and the resulting capacity is

$$C=n_1\log\left(1+\frac{1}{n_1}\sum_{t=1}^{T-1}\frac{P_t}{d_{tT}^{\alpha}}\right). \tag{73}$$

The condition (72) is identical to (55), but the capacity (73) is increased if $n_1>1$.

Remark 34: The above capacity results remain valid for various practical extensions of our models. For example, suppose that phase information can be shared locally, i.e., terminals that are near each other can exchange knowledge of their A_{st}. If the destination is far away, and the transmitting terminals cannot determine the A_{st} between them and the destination, then the capacity remains the same. Similarly, if local cooperation is possible but the receiver is far away and there is phase uncertainty, the right hand side of (65) is still the capacity.

G. Quasistatic Fading

Quasistatic fading has the A_{st} chosen randomly at the beginning of time and held fixed for all channel uses [61, Sec. 5]. The information rates given $P(x_1,x_{\mathcal{T}})$ can therefore be viewed as random variables [66, p. 2631], and the situation is rather more complicated than for ergodic fading. To illustrate the differences, consider a single-relay, single-antenna terminals, and phase fading.

Suppose we use decode-forward with irregular encoding and successive decoding. This strategy will not work well because its intermediate decoding steps can fail. Consider instead regular encoding with either backward or window decoding. The destination will likely make errors if either $I(X_1X_2;Y_3|A_{13}A_{23})$ or $I(X_1;Y_2|X_2A_{12})$ is smaller than the code rate R, because in the second case the relay likely transmits the wrong codewords. We thus say that an outage occurs if either of these informations is too small.

The best input distribution $P(x_1,x_2)$ for all our bounds and for any realization of A_{12}, A_{13}, and A_{23} is again Gaussian, but one must now carefully adjust $\rho=\mathrm{E}[X_1X_2^*]/\sqrt{P_1P_2}$. Recall that $A_{st}=e^{j\theta_{st}}$, and let $\theta=\theta_{13}-\theta_{23}+\theta_\rho$ where θ_ρ is the phase of ρ. The information rate of (5) is the random variable

$$\Psi(\rho,\theta)=\min\left[\log\left(1+\frac{P_1}{d_{12}^{\alpha}}(1-|\rho|^2)\right),\right.$$
$$\left.\log\left(1+\frac{P_1}{d_{13}^{\alpha}}+\frac{P_2}{d_{23}^{\alpha}}+\frac{2\sqrt{P_1P_2}|\rho|\cos\theta}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}}\right)\right]. \tag{74}$$

This random variable depends on θ_ρ but, since θ is uniform over the interval $[0,2\pi)$, we can restrict attention to real and non-negative ρ. The decode-forward outage probability is thus

$$P_{out}^{DF}(R)=\min_{0\le\rho\le1}\Pr\left(\Psi(\rho,\theta)\le R\right). \tag{75}$$

We similarly denote the best possible outage probability at rate R by $P_{out}(R)$.

Continuing with decode-forward, observe that if

$$R\ge\log\left(1+\frac{P_1}{d_{12}^{\alpha}}\right) \tag{76}$$

then $P_{out}^{DF}(R)=1$. For smaller R, we infer from (74) that one should choose ρ as large as possible, i.e., choose the positive ρ satisfying

$$R=\log\left(1+\frac{P_1}{d_{12}^{\alpha}}(1-\rho^2)\right). \tag{77}$$

The random component of (74) is $\xi=\cos\theta$ that has the cumulative distribution

$$\Pr(\xi\le x)=\begin{cases}0, & x<-1\\ \frac{1}{2}+\frac{1}{\pi}\arcsin(x), & -1\le x\le 1\\ 1, & x>1.\end{cases} \tag{78}$$

Using (77) and (78), we compute

$$P_{out}^{DF}(R)=\begin{cases}0, & f(R)<0\\ \frac{1}{2}+\frac{1}{\pi}\arcsin(f(R)), & 0\le f(R)\le 1\\ 1, & f(R)>1\end{cases} \tag{79}$$

where

$$f(R)=\frac{\left(e^R-1-\frac{P_1}{d_{13}^{\alpha}}-\frac{P_2}{d_{23}^{\alpha}}\right)d_{13}^{\alpha/2}d_{23}^{\alpha/2}}{2\sqrt{P_2}\sqrt{P_1-d_{12}^{\alpha}(e^R-1)}}. \tag{80}$$

We remark that $P_{out}^{DF}(R)=0$ clearly implies $P_{out}(R)=0$. Also, if $P_{out}^{DF}(R)\neq 0$ then $P_{out}^{DF}(R)\ge 1/2$. However, decode-forward is not necessarily optimal when $P_{out}(R)\neq 0$. A lower bound on $P_{out}(R)$ can be computed using (3) and (78). We have

$$P_{out}(R)=1,\ \text{if } R\ge\log\left(1+P_1\left(\frac{1}{d_{12}^{\alpha}}+\frac{1}{d_{13}^{\alpha}}\right)\right) \tag{81}$$

and, if the rate is smaller than in (81),

$$P_{out}(R)\ge\begin{cases}0, & g(R)<0\\ \frac{1}{2}+\frac{1}{\pi}\arcsin(g(R)), & 0\le g(R)\le 1\\ 1, & g(R)>1\end{cases} \tag{82}$$

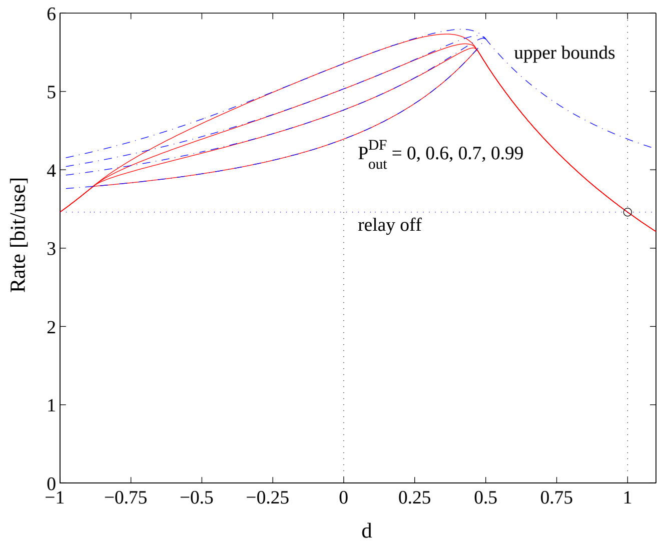

Fig. 13. Outage rates for a single-relay network with phase fading, $P_1 = P_2 = 10$, and $\alpha = 2$.

where

$$g(R) = \frac{\left(e^R - 1 - \frac{P_1}{d_{13}^\alpha} - \frac{P_2}{d_{23}^\alpha}\right) d_{13}^{\alpha/2} d_{23}^{\alpha/2}}{2\sqrt{P_2}\sqrt{P_1 - d_{12}^\alpha d_{13}^\alpha (e^R - 1)/(d_{12}^\alpha + d_{13}^\alpha)}}. \quad (83)$$

To illustrate the outage probability behavior, consider again the geometry of Fig. 6. We plot the decode-forward rates for $P_{out}^{DF} = 0, 0.6, 0.7, 0.99$ as the solid lines in Fig. 13 (the rates for P_{out}^{DF} satisfying $0 \le P_{out}^{DF} \le 0.5$ are the all same). Observe that as $P_{out}^{DF} \to 1$ the rates approach the decode-forward rates in Fig. 7. The dash-dotted curves in Fig. 13 are upper bounds on the best possible rates for $P_{out} = 0, 0.6, 0.7, 0.99$. These rates were computed using (81)–(83), and they approach the upper bound in Fig. 7 as $P_{out} \to 1$.

Remark 35: For Rayleigh fading the decode-forward information rate is the random variable

$$\Psi(\rho, A_{all}) = \min\left[\log\left(1 + \frac{P_1|A_{12}|^2}{d_{12}^\alpha}(1 - |\rho|^2)\right), \log\left(1 + \frac{P_1|A_{13}|^2}{d_{13}^\alpha} + \frac{P_2|A_{23}|^2}{d_{23}^\alpha} + \frac{2\sqrt{P_1P_2}\,\Re(\rho A_{13}A_{23}^*)}{d_{13}^{\alpha/2} d_{23}^{\alpha/2}}\right)\right]. \quad (84)$$

Note that the two random variables inside the minimization are independent, which helps to simplify the analysis somewhat. The outage statistics for the first random variable can be computed using the incomplete gamma function as in [61, Sec. 5.1].

Remark 36: Suppose that instead of quasistatic fading we have *block fading* where the A_{st} are chosen independently from block to block. The relay outage probability with decode-forward is then the same from block to block, but the destination outage probability depends on whether the relay made an error in the previous block. Suppose the relay outage probability is p_2, and the destination outage probability is p_3' if the relay sends the correct codeword. It seems natural to define the overall destination outage probability to be $p_2 + (1 - p_2)p_3'$. One should thus minimize this quantity rather than the probability on the right hand side of (75).

H. Multi-source Networks and Phase Fading

The capacity theorems derived above generalize to several multi-source networks. For instance, consider MARCs with phase fading (see Sec. III-D). We find that the best input distribution for (17) and (18) is Gaussian with $U_1 = U_2 = 0$, i.e., the $\underline{X}_t$ are independent. One again achieves capacity if the source and relay terminals are near each other, and we summarize this with the following theorem.

Theorem 10: The decode-forward strategy of Sec. III-D achieves all points inside the capacity region of MARCs with phase fading if

$$\begin{aligned} P_1/d_{14}^\alpha + P_3/d_{34}^\alpha &\le P_1/d_{13}^\alpha \\ P_2/d_{24}^\alpha + P_3/d_{34}^\alpha &\le P_2/d_{23}^\alpha \\ P_1/d_{14}^\alpha + P_2/d_{24}^\alpha + P_3/d_{34}^\alpha &\le P_1/d_{13}^\alpha + P_2/d_{23}^\alpha \quad (85) \end{aligned}$$

and the capacity region is the set of (R_1, R_2) satisfying

$$\begin{aligned} 0 \le R_1 &\le \log\left(1 + P_1/d_{14}^\alpha + P_3/d_{34}^\alpha\right) \\ 0 \le R_2 &\le \log\left(1 + P_2/d_{14}^\alpha + P_3/d_{34}^\alpha\right) \\ R_1 + R_2 &\le \log\left(1 + P_1/d_{14}^\alpha + P_2/d_{24}^\alpha + P_3/d_{34}^\alpha\right). \quad (86) \end{aligned}$$

Generalizations of Theorem 10 to include more sources and relays, as well as multiple antennas, are possible. Related capacity statements can also be made for BRCs. For example, suppose we broadcast a common message W_0 to two destinations with $R_1 = R_2 = 0$. We apply a cut-set upper bound and use (20) with independent X_1 and X_2 to prove the following theorem.

Theorem 11: The decode-forward strategy of Sec. III-D achieves the capacity of BRCs with phase fading and with a common message if

$$\min\left[\frac{P_1}{d_{13}^\alpha} + \frac{P_2}{d_{23}^\alpha}, \frac{P_1}{d_{14}^\alpha} + \frac{P_2}{d_{24}^\alpha}\right] \le P_1/d_{12}^\alpha \quad (87)$$

and the resulting capacity is

$$C = \min\left[\log\left(1 + \frac{P_1}{d_{13}^\alpha} + \frac{P_2}{d_{23}^\alpha}\right), \log\left(1 + \frac{P_1}{d_{14}^\alpha} + \frac{P_2}{d_{24}^\alpha}\right)\right]. \quad (88)$$

Theorem 11 generalizes to other fading models. However, some care is needed if $R_1 > 0$ or $R_2 > 0$ because the BRCs might not be degraded.

VII. Conclusions

We considered several coding strategies for relay networks. The decode-forward strategies are useful for relays that are close to the source, and the compress-forward strategies are useful for relays that are close to the destination (and sometimes even close to the source). A strategy that mixes decode-forward and compress-forward achieves capacity if the terminals form two closely-spaced clusters. It was further shown that decode-forward achieves the ergodic capacity of a number of wireless channels with phase fading if phase information is available only locally, and if all relays are near the source terminal. The capacity results extend to multi-source problems such as MARCs and BRCs.

There are many directions for further work on relay networks. For example, the fundamental problem of the capacity of the single-relay channel has been open for decades. In fact, even for the Gaussian single-relay channel without fading we know capacity only if the relay is colocated with either the source or destination. Another challenge is designing codes that approach the performance predicted by the theory. First results of this nature have already appeared in [67].

APPENDIX A
AUXILIARY LEMMA

The set $T_\epsilon^{(n)}(X)$ of ϵ-typical vectors $\underline{x}$ of length n with respect to $p_X(\cdot)$ is defined as (see [4, Defn. 1])

$$T_\epsilon^{(n)}(X) = \left\{ \underline{x} : \frac{|n_{\underline{x}}(a)/n - p_X(a)|}{|\mathcal{X}|} < \epsilon \text{ for all } a \in \mathcal{X} \right\}. \tag{89}$$

Here $\mathcal{X}$ is the alphabet of X and the entries of $\underline{x}$, and $n_{\underline{x}}(a)$ is the number of times the letter a occurs in $\underline{x}$. We will need the following simple extension of Theorem 14.2.3 in [17, p. 387].

Lemma 1: Consider the distribution $p(s_1, \ldots, s_M)$, and let $(\underline{s}'_1, \ldots, \underline{s}'_M)$ be a random vector of M-tuples $(s'_{1i}, \ldots, s'_{Mi})$, $i = 1, \ldots, n$, that are chosen i.i.d. with distribution $p(s_1) \prod_{m=2}^M p(s_m|s_1)$. Then the probability P that $(\underline{s}'_1, \ldots, \underline{s}'_M)$ is in $T_\epsilon^{(n)}(S_1, S_2, \ldots, S_M)$ is bounded by

$$(1-\epsilon)\, 2^{-n[2M\epsilon+\gamma]} \le P \le 2^{-n[-2M\epsilon+\gamma]} \tag{90}$$

where

$$\gamma = -H(S_2 \ldots S_M|S_1) + \sum_{m=2}^{M} H(S_m|S_1). \tag{91}$$

Lemma 1 automatically includes unconditioned bounds by making S_1 a constant. We omit the proof because of its similarity to [17, p. 387].

APPENDIX B
DECODE-FORWARD FOR MARCS

Consider MARCs with $T-2$ source terminals, a relay terminal $T-1$, and a destination terminal T. In what follows, we do not derive error probability bounds because the analysis is basically the same as for MAC decoding (see [17, p. 403]) and MAC backward decoding (see [16, Ch. 7]).

Let $\mathcal{T}' = \{1, 2, \ldots, T-2\}$ and $\underline{u}_\mathcal{S}(r_\mathcal{S}) = \{\underline{u}_t(r_t) : t \in \mathcal{S}\}$. For rates, we write $R_\mathcal{S} = \sum_{t\in S} R_t$. Each source terminal $t \in \mathcal{T}'$ sends B message blocks $w_{t1}, w_{t2}, \ldots, w_{tB}$ in $B+1$ transmission blocks. The overall rate is thus reduced by the factor $B/(B+1)$, but for large B this rate loss is negligible.

Random Code Construction:

1) For all $t \in \mathcal{T}'$ choose 2^{nR_t} i.i.d. $\underline{u}_t$ with $p(\underline{u}_t) = \prod_i p_{U_t}(u_{ti})$. Label these $\underline{u}_t(r_t)$, $r_t \in [1, 2^{nR_t}]$.
2) For $t \in \mathcal{T}'$ and for each $\underline{u}_t(r_t)$ choose 2^{nR_t} i.i.d. $\underline{x}_t$ with $p(\underline{x}_t|\underline{u}_t(r_t)) = \prod_i p_{X_t|U_t}(x_{ti}|u_{ti}(r_t))$. Label these $\underline{x}_t(r_t, s_t)$, $s_t \in [1, 2^{nR_t}]$.
3) For all $\underline{u}_{\mathcal{T}'}(r_{\mathcal{T}'})$ choose an $\underline{x}_{T-1}$ with $p(\underline{x}_{T-1}|\underline{u}_{\mathcal{T}'}(r_{\mathcal{T}'})) = \prod_i p_{X_t|U_{\mathcal{T}'i}}(x_{ti}|\underline{u}_{\mathcal{T}'i}(r_{\mathcal{T}'}))$. Label this vector $\underline{x}_{T-1}(r_{\mathcal{T}'})$.

Encoding: For block b encoding proceeds as follows.

1) Source t sends $\underline{x}_t(1, w_{tb})$ if $b = 1$, $\underline{x}_t(w_{t(b-1)}, w_{tb})$ if $1 < b < B+1$, and $\underline{x}_t(w_{t(b-1)}, 1)$ if $b = B+1$.
2) The relay knows $w_{\mathcal{T}'(b-1)}$ from decoding step (1) and transmits $\underline{x}_{T-1}(w_{\mathcal{T}'(b-1)})$.

Decoding: Decoding proceeds as follows.

1) *(At the relay)* The relay decodes the messages $w_{\mathcal{T}'b}$ after block b. The decoding problem is the same as for a MAC with side information $U_{\mathcal{T}'}$ and X_{T-1}. We can thus decode reliably if (see [17, p. 403])
$$R_\mathcal{S} < I(X_\mathcal{S}; Y_{T-1}|U_{\mathcal{T}'} X_{\mathcal{S}^C} X_{T-1}) \tag{92}$$
for all $\mathcal{S} \subseteq \mathcal{T}'$, where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{T}'$. For example, for $\mathcal{T}' = \{1,2\}$ we have the bounds (17).
2) *(At the destination)* The destination waits until all transmission is completed. It then decodes $w_{\mathcal{T}'b}$ for $b = B, B-1, \ldots, 1$ by using its output block $\underline{y}_{T(b+1)}$. The techniques of [16, Ch. 7] can be used to show that one can decode reliably if
$$R_\mathcal{S} < I(X_\mathcal{S} X_{T-1}; Y_T|U_{\mathcal{S}^C} X_{\mathcal{S}^C}) \tag{93}$$
for all $\mathcal{S} \subseteq \mathcal{T}'$, where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{T}'$. For example, for $\mathcal{T}' = \{1,2\}$ we obtain (18) .

APPENDIX C
DECODE-FORWARD FOR BRCS

Consider a BRC with four terminals as in Sec. III-D. We again do not give error probability bounds because the analysis is based on standard arguments (see [17, Sec. 14.2],[54], [55]).

The source terminal divides its messages W_0, W_1, W_2 into B blocks w_{0b}, w_{1b}, w_{2b} for $b = 1, 2, \ldots, B$, and transmits these in $B+1$ blocks. The w_{0b}, w_{1b}, w_{2b} have the respective rates R_0, R_1, R_2 for each b. The first encoding step is to map these blocks into new blocks $w'_{0b}, w'_{1b}, w'_{2b}$ with respective rates R'_0, R'_1, R'_2. The mapping is restricted to have the property that the bits of (w_{0b}, w_{1b}) are in (w'_{0b}, w'_{1b}), and that the bits of (w_{0b}, w_{2b}) are in (w'_{0b}, w'_{2b}). This can be done in one of two ways, as depicted in Fig. 14 (the bits w_{0b} in w'_{1b} and w'_{2b} must be the same on the right in Fig. 14). The corresponding restrictions on the rates R_0, R_1, R_2 are

$$\begin{aligned} R_0 + R_1 &\le R'_0 + R'_1 \\ R_0 + R_2 &\le R'_0 + R'_2 \\ R_0 + R_1 + R_2 &\le R'_0 + R'_1 + R'_2 \end{aligned} \tag{94}$$

The encoding also involves two *binning* steps, for which we need rates R''_1 and R''_2 satisfying

$$R''_1 \ge R'_1, \; R''_2 \ge R'_2. \tag{95}$$

Random Code Construction:

1) Choose $2^{nR'_0}$ i.i.d. $\underline{x}_2$ with $p(\underline{x}_2) = \prod_i p_{X_2}(x_{2i})$. Label these $\underline{x}_2(q)$, $q \in [1, 2^{nR'_0}]$.
2) For each $\underline{x}_2(q)$ choose $2^{nR'_0}$ i.i.d. $\underline{u}_0$ with $p(\underline{u}_0|\underline{x}_2(q)) = \prod_i p_{U_0}(u_{0i}|\underline{x}_{2i}(q))$. Label these $\underline{u}_0(q, r_0)$, $r_0 \in [1, 2^{nR'_0}]$.

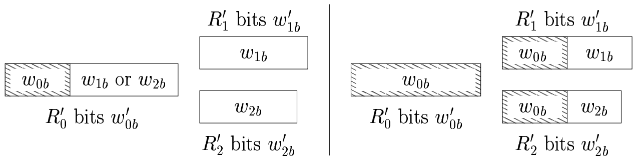

Fig. 14. Reorganization of the message bits for BRC coding.

3) For each $\underline{u}_0(q,r_0)$ choose $2^{nR''_1}$ i.i.d. $\underline{u}_1$ with $p(\underline{u}_1|\underline{x}_2(q),\underline{u}_0(q,r_0)) = \prod_i p_{U_1}(u_{1i}|\underline{x}_{2i}(q),\underline{u}_{0i}(q,r_0))$. Label these $\underline{u}_1(q,r_0,s_1)$, $s_1 \in [1, 2^{nR''_1}]$.
4) For each $\underline{u}_0(q,r_0)$ choose $2^{nR''_2}$ i.i.d. $\underline{u}_2$ with $p(\underline{u}_2|\underline{x}_2(q),\underline{u}_0(q,r_0)) = \prod_i p_{U_2}(u_{2i}|\underline{x}_{2i}(q),\underline{u}_{0i}(q,r_0))$. Label these $\underline{u}_2(q,r_0,s_2)$, $s_2 \in [1, 2^{nR''_2}]$.
5) Randomly partition the set $\{1,\cdots,2^{nR''_1}\}$ into $2^{nR'_1}$ cells S_{r_1} with $r_1 \in [1, 2^{nR'_1}]$.
6) Randomly partition the set $\{1,\cdots,2^{nR''_2}\}$ into $2^{nR'_2}$ cells S_{r_2} with $r_2 \in [1, 2^{nR'_2}]$. Note that we are abusing notation by not distinguishing between the r_1 and r_2 cells. However, the context will make clear which cells we are referring to.
7) For each q, r_0, s_1, s_2 choose an $\underline{x}_1$ with $p(\underline{x}_1|\underline{x}_2(q),\underline{u}_0(q,r_0),\underline{u}_1(q,r_0,s_1),\underline{u}_2(q,r_0,s_2)) = \prod_i p_{X_1|X_2U_0U_1U_2}(x_{1i}|\underline{x}_{2i}(q),\underline{u}_{0i}(q,r_0),\underline{u}_{1i}(q,r_0,s_1), \underline{u}_{2i}(q,r_0,s_2))$. Label this vector $\underline{x}_1(q,r_0,s_1,s_2)$.

Encoding: For block b encoding proceeds as follows.

1) Map w_{0b}, w_{1b}, w_{2b} into $w'_{0b}, w'_{1b}, w'_{2b}$ as discussed above. Set $w'_{00} = 1$.
2) The source terminal finds a pair

$$\left(\underline{u}_1(w'_{0(b-1)}, w'_{0b}, s_{1b}), \underline{u}_2(w'_{0(b-1)}, w'_{0b}, s_{2b})\right)$$

with $s_{1b} \in S_{w'_{1b}}$, $s_{2b} \in S_{w'_{2b}}$, and such that this pair is jointly typical with $\underline{x}_2(w'_{0(b-1)})$ and $\underline{u}_0(w'_{0(b-1)}, w'_{0b})$. The source terminal transmits $\underline{x}_1(w'_{0(b-1)}, w'_{0b}, s_{1b}, s_{2b})$.
3) The relay knows $w'_{0(b-1)}$ from decoding step (1) and transmits $\underline{x}_2(w'_{0(b-1)})$.

The second encoding step can be done only if there is a pair of codewords $(\underline{u}_1, \underline{u}_2)$ satisfying the desired conditions. Standard binning arguments (see [55]) guarantee that such a pair exists with high probability if n is large and

$$(R''_1 - R'_1) + (R''_2 - R'_2) > I(U_1; U_2|U_0X_2). \qquad (96)$$

Decoding: After block b decoding proceeds as follows.

1) *(At the relay)* The relay decodes w'_{0b} by using $\underline{y}_{2b}$, and this can be done reliably if

$$R'_0 < I(U_0; Y_2|X_2). \qquad (97)$$

2) *(At the destinations)* Terminal 3 decodes $w'_{0(b-1)}$ and $s_{1(b-1)}$ by using its past two output blocks $\underline{y}_{3(b-1)}$ and $\underline{y}_{3b}$ (see Fig. 5). Similarly, terminal 4 decodes $w'_{0(b-1)}$ and $s_{2(b-1)}$ by using $\underline{y}_{4(b-1)}$ and $\underline{y}_{4b}$. The techniques of [15], [21] can be used to show that both terminals can decode reliably if

$$\begin{aligned} R''_1 &< I(U_1; Y_3|U_0X_2) \\ R'_0 + R''_1 &< I(U_0U_1X_2; Y_3) \\ R''_2 &< I(U_2; Y_4|U_0X_2) \\ R'_0 + R''_2 &< I(U_0U_2X_2; Y_4). \end{aligned} \qquad (98)$$

Terminals 3 and 4 can recover w'_{1b} and w'_{2b} from the respective s_{1b} and s_{2b}. They can further recover (w_{0b}, w_{1b}) and (w_{0b}, w_{2b}) from the respective (w'_{0b}, w'_{1b}) and (w'_{0b}, w'_{2b}).

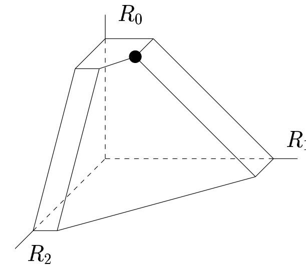

Fig. 15. Achievable rate region for the BRC.

The rate region of (19) has the form depicted in Fig. 15. We proceed as in [54, Sec. III] and show that one can approach the corner point

$$\begin{aligned} R_0 &= \min(I_2, I_3, I_4) \\ R_1 &= \min(I_2, I_3) + I(U_1; Y_3|U_0X_2) - \min(I_2, I_3, I_4) \\ R_2 &= \min(I_2, I_3, I_4) + I(U_2; Y_4|U_0X_2) \\ &\quad - I(U_1; U_2|U_0X_2) - \min(I_2, I_3) \end{aligned} \qquad (99)$$

where we recall that

$$I_2 = I(U_0; Y_2|X_2), I_3 = I(U_0X_2; Y_3), I_4 = I(U_0X_2; Y_4)$$

and $p(u_0, u_1, u_2, x_1, x_2)$ is fixed. We begin by choosing

$$\begin{aligned} R''_1 &= R'_1 \\ R''_2 &= R'_2 + I(U_1; U_2|U_0X_2) + \epsilon \end{aligned} \qquad (100)$$

for a small positive ϵ. This choice satisfies (95) and (96). We next consider two cases separately.

1) Suppose we have $\min(I_2, I_3, I_4) \neq I_4$. We choose

$$\begin{aligned} R_0 = R'_0 &= \min(I_2, I_3) - \epsilon \\ R_1 = R'_1 &= I(U_1; Y_3|U_0X_2) - \epsilon \\ R_2 = R'_2 &= I(U_2; Y_4|U_0X_2) - I(U_1; U_2|U_0X_2) - 2\epsilon \end{aligned} \qquad (101)$$

to satisfy (94), (97), and (98). We can thus approach the corner point (99) by letting $\epsilon \to 0$.
2) Suppose we have $\min(I_2, I_3, I_4) = I_4$. We choose

$$\begin{aligned} R_0 &= I_4 - \epsilon \\ R'_0 &= \min(I_2, I_3) - \epsilon \\ R_1 &= \min(I_2, I_3) + I(U_1; Y_3|U_0X_2) - I_4 - \epsilon \\ R'_1 &= I(U_1; Y_3|U_0X_2) - \epsilon \\ R_2 = R'_2 &= I(U_2; Y_4|U_0X_2) - I(U_1; U_2|U_0X_2) \\ &\quad + I_4 - \min(I_2, I_3) - 2\epsilon \end{aligned} \qquad (102)$$

which satisfies (94), (97), and (98) (note that we are using the mapping on the left in Fig. 14). We can again approach the corner point (99) by letting $\epsilon \to 0$.

One can thus approach both $R_0 = \min(I_2, I_3, I_4)$ corner points in Fig. 15. One can approach the $R_0 = 0$ corner points by operating at one of the $R_0 = \min(I_2, I_3, I_4)$ corner points and assigning all of the W_0 bits to either W_1 or W_2. The remaining points in the region of Fig. 15 are achieved by time-sharing.

APPENDIX D
PROOF OF THEOREM 3

Our proof follows closely the proof of Theorem 6 in [4]. Recall that $\mathcal{T} = \{2, 3, \ldots, T-1\}$ and $r_{\mathcal{S}} = \{r_t : t \in \mathcal{S}\}$. We write $\underline{u}_{\mathcal{S}}(r_{\mathcal{S}}) = \{\underline{u}_t(r_t) : t \in \mathcal{S}\}$. For rates, we write $R_{\mathcal{S}} = \sum_{t \in \mathcal{S}} R_t$. We send B message blocks $w_1, w_2, \ldots, w_B$ in $B+1$ transmission blocks. The overall rate is thus reduced by the factor $B/(B+1)$, but for large B this rate loss is negligible. The code construction is illustrated for 2 relays in Fig. 16.

Random Code Construction:

1) Choose 2^{nR_1} i.i.d. $\underline{x}_1$ with $p(\underline{x}_1) = \prod_i p_{X_1}(x_{1i})$. Label these $\underline{x}_1(w)$, $w \in [1, 2^{nR_1}]$.
2) For all $t \in \mathcal{T}$ choose $2^{nR'_t}$ i.i.d. $\underline{u}_t$ with $p(\underline{u}_t) = \prod_i p_{U_t}(u_{ti})$. Label these $\underline{u}_t(r_t)$, $r_t \in [1, 2^{nR'_t}]$.
3) For all $t \in \mathcal{T}$ and for each $\underline{u}_t(r_t)$, choose $2^{n(R_t - R'_t)}$ i.i.d. $\underline{x}_t$ with $p(\underline{x}_t|\underline{u}_t(r_t)) = \prod_i p_{X_t|U_t}(x_{ti}|u_{ti}(r_t))$. Label these $\underline{x}_t(s_t|r_t)$, $s_t \in [1, 2^{n(R_t - R'_t)}]$.
4) For all $t \in \mathcal{T}$ and for each $(\underline{x}_t(s_t|r_t), \underline{u}_{\mathcal{T}}(r_{\mathcal{T}}))$, choose $2^{n\hat{R}_t}$ i.i.d. $\underline{\hat{y}}_t$ with $p(\underline{\hat{y}}_t|\underline{x}_t(s_t|r_t), \underline{u}_{\mathcal{T}}(r_{\mathcal{T}})) = \prod_i p_{\hat{Y}_t|X_t U_{\mathcal{T}}}(\hat{y}_{ti}|x_{ti}(s_t|r_t), u_{\mathcal{T}i}(r_{\mathcal{T}}))$. Label these $\underline{\hat{y}}_t(z_t|s_t, r_{\mathcal{T}})$, $z_t \in [1, 2^{n\hat{R}_t}]$.
5) For all $t \in \mathcal{T}$ randomly partition the set $\{1, \cdots, 2^{n\hat{R}_t}\}$ into 2^{nR_t} cells S_{r_t, s_t}, $r_t \in [1, 2^{nR'_t}]$, $s_t \in [1, 2^{n(R_t - R'_t)}]$.

We remark that the $\underline{u}_t$ are decoded by the destination terminal *and* the relays, while the $\underline{x}_t$ are decoded by the destination terminal only. One can think of the $\underline{u}_t$ as cloud centers (coarse spacing) and the $\underline{x}_t$ as clouds (fine spacing).

Encoding:

1) Let w_b be the message in block b. The source terminal sends $\underline{x}_1(w_b)$.
2) For all $t \in \mathcal{T}$ terminal t knows $z_{t(b-1)}$ from decoding step (4), and chooses (r_{tb}, s_{tb}) so that $z_{t(b-1)} \in S_{r_{tb}, s_{tb}}$. Terminal t transmits $\underline{x}_t(s_{tb}|r_{tb})$.

Decoding: After block b decoding proceeds as follows.

1) *(At the destination)* The destination terminal first decodes the $T-2$ indexes $r_{\mathcal{T}b}$ using $\underline{y}_{Tb}$. This can be done reliably if (see [17, p. 403])
$$R'_{\mathcal{S}} < I(U_{\mathcal{S}}; Y_T | U_{\mathcal{S}^C}) \tag{103}$$
for all $\mathcal{S} \subseteq \mathcal{T}$, where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{T}$. The destination terminal next decodes $s_{\mathcal{T}b}$. This can be done reliably if
$$R_{\mathcal{S}} - R'_{\mathcal{S}} < I(X_{\mathcal{S}}; Y_T | U_{\mathcal{T}} X_{\mathcal{S}^C}) \tag{104}$$
for all $\mathcal{S} \subseteq \mathcal{T}$.

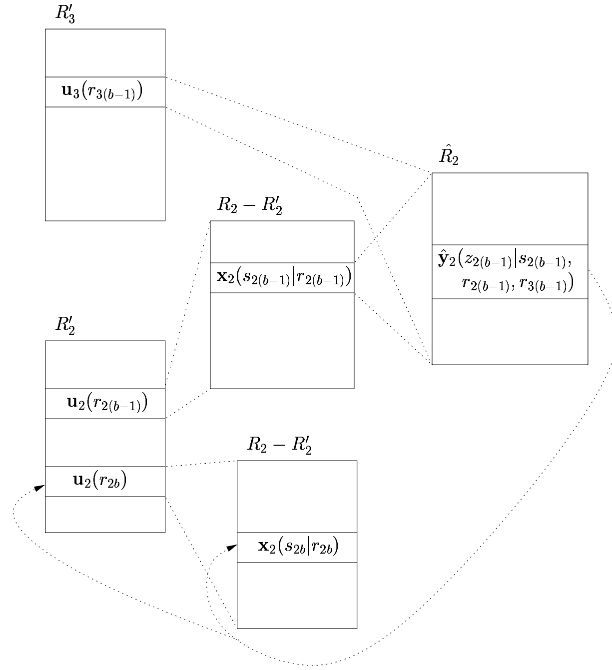

Fig. 16. The code used by terminal 2 (the first relay) for compress-forward.

2) *(At the destination)* The destination terminal determines the set $\mathcal{L}(\underline{y}_{T(b-1)})$ of $z_{\mathcal{T}}$ such that
$$\Big(\{\underline{u}_t(r_{t(b-1)}), \underline{x}_t(s_{t(b-1)}|r_{t(b-1)}), \underline{\hat{y}}_t(z_t|s_{t(b-1)}, r_{\mathcal{T}(b-1)}) : t \in \mathcal{T}\}, \underline{y}_{T(b-1)}\Big) \tag{105}$$
is jointly ϵ-typical, where the $r_{t(b-1)}$ and $s_{t(b-1)}$ were obtained from the first decoding step. The destination terminal declares that $z_{\mathcal{T}(b-1)}$ was sent in block $b-1$ if
$$z_{\mathcal{T}} \in \left(S_{s_{2b}, r_{2b}} \times \cdots \times S_{s_{(T-1)b}, r_{(T-1)b}}\right) \cap \mathcal{L}(\underline{y}_{T(b-1)}) \tag{106}$$
where $z_{\mathcal{T}} = z_{\mathcal{T}(b-1)}$.

We compute the error probability of this decoding step, i.e., the probability that $z_{\mathcal{T}(b-1)}$ was incorrectly chosen. We first partition $\mathcal{L}(\underline{y}_{T(b-1)})$ into 2^{T-2} sets:
$$\mathcal{L}_{\mathcal{S}} = \{z_{\mathcal{T}} : z_{\mathcal{T}} \in \mathcal{L}(\underline{y}_{T(b-1)}), z_t \neq z_{t(b-1)} \text{ for } t \in \mathcal{S}, \; z_t = z_{t(b-1)} \text{ for } t \notin \mathcal{S}\} \tag{107}$$
where $\mathcal{S} \subseteq \mathcal{T}$. These sets could be empty. We proceed to determine their average size. We follow [4] and write F^c_{b-1} for the event that all decisions in block $b-1$ were correct. We have
$$\mathrm{E}[||\mathcal{L}_{\mathcal{S}}|| \,|F^c_{b-1}] = \sum_{z_{\mathcal{T}} \in \mathcal{L}_{\mathcal{S}}} \mathrm{E}[\psi(Z_{\mathcal{T}}|\underline{y}_{T(b-1)})|F^c_{b-1}] \tag{108}$$
where $Z_{\mathcal{T}}$ is the random variable that maps to $z_{\mathcal{T}}$, and where
$$\psi(z_{\mathcal{T}}|\underline{y}_{T(b-1)}) = \begin{cases} 1, & \text{if (105) is jointly typical} \\ 0, & \text{otherwise.} \end{cases} \tag{109}$$

Now if $z_{\mathcal{T}} \in \mathcal{L}_{\mathcal{S}}$ then $(\hat{Y}_{\mathcal{S}^C}, Y_T)$ and the $\hat{Y}_t$ with $t \in \mathcal{S}$ are jointly independent given $(U_{\mathcal{T}}, X_{\mathcal{T}})$. Lemma 1 thus ensures that

$$\begin{aligned}
&\mathrm{E}[\psi(Z_{\mathcal{T}}|\underline{y}_{T(b-1)})|F_{b-1}^c] \\
&\le 2^{-n\left[-2(|\mathcal{S}|+2)\epsilon - H(\hat{Y}_{\mathcal{S}}\hat{Y}_{\mathcal{S}^C}Y_T|U_{\mathcal{T}}X_{\mathcal{T}}) + H(\hat{Y}_{\mathcal{S}^C}Y_T|U_{\mathcal{T}}X_{\mathcal{T}})\right.} \\
&\qquad {}^{\left. + \sum_{t\in\mathcal{S}} H(\hat{Y}_t|U_{\mathcal{T}}X_{\mathcal{T}})\right]} \\
&= 2^{-n\left[-2(|\mathcal{S}|+2)\epsilon - H(\hat{Y}_{\mathcal{S}}|U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{S}^C}Y_T) + \sum_{t\in\mathcal{S}} H(\hat{Y}_t|U_{\mathcal{T}}X_{\mathcal{T}})\right]}.
\end{aligned} \tag{110}$$

There are $2^{n\hat{R}_{\mathcal{S}}} - 1$ choices for $z_{\mathcal{S}} \neq z_{\mathcal{S}(b-1)}$. Hence, using the union bound, we upper bound $\mathrm{E}[||\mathcal{L}_{\mathcal{S}}|| \big| F_{b-1}^c]$ by

$$2^{n\hat{R}_{\mathcal{S}}} 2^{-n\left[-2(|\mathcal{S}|+2)\epsilon - H(\hat{Y}_{\mathcal{S}}|U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{S}^C}Y_T) + \sum_{t\in\mathcal{S}} H(\hat{Y}_t|U_{\mathcal{T}}X_{\mathcal{T}})\right]}. \tag{111}$$

As long as the s_{tb} of block b have been decoded correctly, an error is made only if there is a $z_{\mathcal{T}} \neq z_{\mathcal{T}(b-1)}$ in $\mathcal{L}(\underline{y}_{T(b-1)})$ which maps back to $s_{\mathcal{T}b}$. We thus compute

$$\begin{aligned}
&\Pr(\text{ error in step (2) }) \\
&\le \Pr\left(\bigcup_{\mathcal{S}\subseteq\mathcal{T},\mathcal{S}\neq\emptyset} \bigcup_{z_{\mathcal{T}}\in\mathcal{L}_{\mathcal{S}}} \text{event (106) occurs} \middle| F_{b-1}^c \right) \\
&\le \sum_{\mathcal{S}\subseteq\mathcal{T},\mathcal{S}\neq\emptyset} \Pr\left(\bigcup_{z_{\mathcal{T}}\in\mathcal{L}_{\mathcal{S}}} \text{event (106) occurs} \middle| F_{b-1}^c \right) \\
&\le \sum_{\mathcal{S}\subseteq\mathcal{T},\mathcal{S}\neq\emptyset} \Pr\left(\bigcup_{z_{\mathcal{T}}\in\mathcal{L}_{\mathcal{S}}} \{z_t \in S_{s_{tb},r_{tb}} \text{ for all } t \in \mathcal{S}\} \middle| F_{b-1}^c \right) \\
&= \sum_{\mathcal{S}\subseteq\mathcal{T},\mathcal{S}\neq\emptyset} \mathrm{E}\left[||\mathcal{L}_{\mathcal{S}}|| \big| F_{b-1}^c\right] 2^{-nR_{\mathcal{S}}}.
\end{aligned} \tag{112}$$

Inserting (111) into (112), we see that as long as

$$R_{\mathcal{S}} > \hat{R}_{\mathcal{S}} + H(\hat{Y}_{\mathcal{S}}|U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{S}^C}Y_T) - \sum_{t\in\mathcal{S}} H(\hat{Y}_t|U_{\mathcal{T}}X_{\mathcal{T}}) \tag{113}$$

for all $\mathcal{S} \subseteq \mathcal{T}$ then decoding step (2) can be made reliable.

Decoding (continued):

3) *(At the destination).* Assuming that the $r_{\mathcal{T}(b-1)}$, $s_{\mathcal{T}(b-1)}$ and $z_{\mathcal{T}(b-1)}$ were decoded correctly, the destination finally declares that w_{b-1} was sent in block $b-1$ if

$$\begin{aligned}
&\big(\underline{x}_1(w_{b-1}), \{\underline{u}_t(r_{t(b-1)}), \underline{x}_t(s_{t(b-1)}|r_{t(b-1)}), \\
&\quad \underline{\hat{y}}_t(z_{t(b-1)}|s_{t(b-1)}, r_{\mathcal{T}(b-1)}) : t \in \mathcal{T}\}, \underline{y}_{T(b-1)}\big)
\end{aligned} \tag{114}$$

is jointly ϵ-typical. Using Lemma 1, the probability that there exists a $w \neq w_{b-1}$ such that $\underline{x}_1(w)$ satisfies (114) is upper bounded by

$$2^{-n(I(X_1;U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{T}}Y_T)+6\epsilon)}. \tag{115}$$

Applying $I(X_1;U_{\mathcal{T}}X_{\mathcal{T}}) = 0$, we find that if

$$R_1 < I(X_1;\hat{Y}_{\mathcal{T}}Y_T|U_{\mathcal{T}}X_{\mathcal{T}}) \tag{116}$$

then this decoding step can be made reliable.

4) *(At the relays).* Relay t estimates the r_{sb} with $s \neq t$, which can be done reliably if

$$R'_{\mathcal{S}} < I(U_{\mathcal{S}};Y_t|U_{\mathcal{S}^C}X_t) \tag{117}$$

for all $\mathcal{S} \subseteq \mathcal{T}\backslash\{t\}$. The set $\mathcal{S}^C$ is again the complement of $\mathcal{S}$ in $\mathcal{T}$, but one can remove the U_t in the conditioning of (117). Suppose the $T-3$ estimates of the r_{sb} are correct. Relay t chooses any of the z_t so that

$$\left(\{\underline{u}_t(r_{tb}) : t \in \mathcal{T}\}, \underline{x}_t(s_{tb}|r_{tb}), \underline{\hat{y}}_t(z_t|s_{tb}, r_{\mathcal{T}b}), \underline{y}_{tb}\right)$$

is jointly ϵ-typical.

The probability P that there is *no* such z_t is bounded by

$$P \le \left(1 - (1-\epsilon)2^{-n[4\epsilon + I(\hat{Y}_t;Y_t|U_{\mathcal{T}}X_t)]}\right)^{2^{n\hat{R}_t}} \tag{118}$$

as follows from Lemma 1. We use $\log x \le (x-1)$ to bound

$$\begin{aligned}
\log P &\le 2^{n\hat{R}_t} \log\left(1 - (1-\epsilon)2^{-n[4\epsilon + I(\hat{Y}_t;Y_t|U_{\mathcal{T}}X_t)]}\right) \\
&\le -(1-\epsilon)2^{n\hat{R}_t} 2^{-n[4\epsilon + I(\hat{Y}_t;Y_t|U_{\mathcal{T}}X_t)]}
\end{aligned} \tag{119}$$

This expression can be made to approach $-\infty$ with large n if

$$\hat{R}_t = I(\hat{Y}_t;Y_t|U_{\mathcal{T}}X_t) + \delta. \tag{120}$$

for $\delta > 4\epsilon > 0$.

Finally, we combine (113) and (120), and use (see (26))

$$H(\hat{Y}_t|U_{\mathcal{T}}X_tY_t) = H(\hat{Y}_t|U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{S}^C}Y_{\mathcal{T}}Y_T)$$

to obtain the left hand side of (25). We combine (103), (104) and (117) to obtain the right hand side of (25).

Appendix E
Proof of Theorem 5

One could prove Theorem 5 by using backward or window decoding. Instead, we use the proof technique of [4]. The code construction is illustrated in Fig. 17.

Random Code Construction:

1) Choose $2^{nR'_2}$ i.i.d. $\underline{u}_2$ with $p(\underline{u}_2) = \prod_i p_{U_2}(u_{2i})$. Label these $\underline{u}_2(r_2)$, $r_2 \in [1, 2^{nR'_2}]$.
2) For each $\underline{u}_2(r_2)$ choose $2^{n(R_2-R'_2)}$ i.i.d. $\underline{x}_2$ with $p(\underline{x}_2|\underline{u}_2(r_2)) = \prod_i p_{X_2|U_2}(x_{2i}|u_{2i}(r_2))$. Label these $\underline{x}_2(s_2|r_2)$, $s_2 \in [1, 2^{n(R_2-R'_2)}]$.
3) For each $\underline{x}_2(s_2|r_2)$ choose 2^{nR_1} i.i.d. $\underline{x}_1$ with $p(\underline{x}_1|\underline{u}_2(r_2), \underline{x}_2(s_2|r_2))$ $= \prod_i p_{X_1|U_2X_2}(x_{1i}|u_{2i}(r_2), x_{2i}(s_2|r_2))$. Label these $\underline{x}_1(w|s_2, r_2)$, $w \in [1, 2^{nR_1}]$.
4) Choose 2^{nR_3} i.i.d. $\underline{x}_3$ with $p(\underline{x}_3) = \prod_i p_{X_3}(x_{3i})$. Label these $\underline{x}_3(s_3)$, $s_3 \in [1, 2^{nR_3}]$.
5) For each $(\underline{x}_3(s_3), \underline{u}_2(r_2))$ choose $2^{n\hat{R}_3}$ i.i.d. $\underline{\hat{y}}_3$ with $p(\underline{\hat{y}}_3|\underline{x}_3(s_3), \underline{u}_2(r_2))$ $= \prod_i p_{\hat{Y}_3|U_2X_3}(\hat{y}_{3i}|u_{2i}(r_2), x_{3i}(s_3))$. Label these $\underline{\hat{y}}_3(z_3|s_3, r_2)$, $z_3 \in [1, 2^{n\hat{R}_3}]$.
6) Randomly partition the set $\{1, \cdots, 2^{nR_1}\}$ into 2^{nR_2} cells S_{r_2,s_2}, $r_2 \in [1, 2^{nR'_2}]$, $s_2 \in [1, 2^{n(R_2-R'_2)}]$.

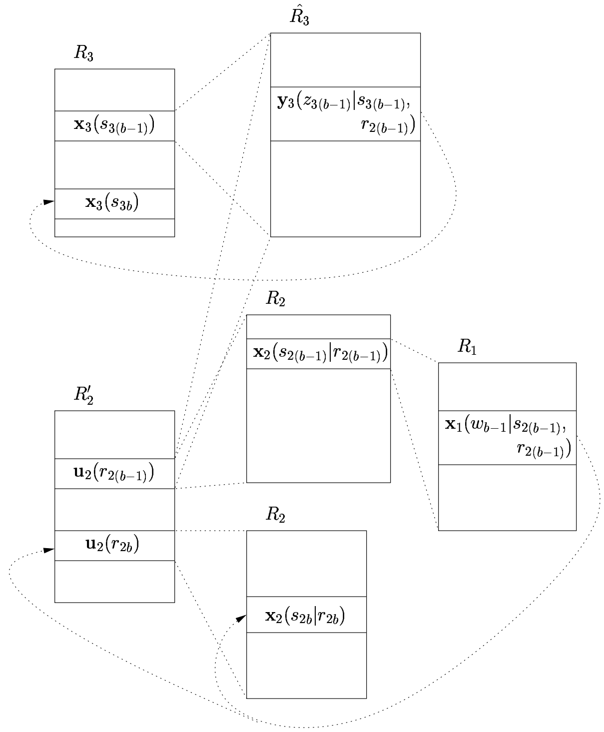

Fig. 17. The code construction for Theorem 5.

7) Randomly partition the set $\{1, \cdots, 2^{n\hat{R}_3}\}$ into 2^{nR_3} cells S_{s_3}, $s_3 \in [1, 2^{nR_3}]$.

Encoding:

1) The source transmits $\underline{x}_1(w_b|s_{2(b-1)}, r_{2(b-1)})$ in block b. It chooses (r_{2b}, s_{2b}) so that $w_{b-1} \in S_{r_{2b}, s_{2b}}$.
2) Terminal 2 knows w_{b-1} (decoding step (1)), and chooses (r_{2b}, s_{2b}) so that $w_{b-1} \in S_{r_{2b}, s_{2b}}$. It transmits $\underline{x}_2(s_{2b}|r_{2b})$.
3) Terminal 3 knows $z_{3(b-1)}$ (decoding step (4)), and chooses s_{3b} so that $z_{3(b-1)} \in S_{s_{3b}}$. It transmits $\underline{x}_3(s_{3b})$.

Decoding: After block b decoding proceeds as follows.

1) Terminal 2 chooses (one of) the message(s) w_b so that $(\underline{u}_2(r_{2b}), \underline{x}_1(w_b|s_{2b}), \underline{x}_2(s_{2b}|r_{2b}), \underline{y}_{2b})$ is jointly typical. This step can be made reliable if
$$R_1 < I(X_1; Y_2|U_2X_2). \tag{121}$$
2) The destination terminal decodes r_{2b} and s_{3b}. This step can be made reliable if
$$R_2' < I(U_2; Y_4|X_3) \tag{122}$$
$$R_3 < I(X_3; Y_4|U_2) \tag{123}$$
$$R_2' + R_3 < I(U_2X_3; Y_4). \tag{124}$$
3) The destination terminal determines the set] $\mathcal{L}(\underline{y}_{4(b-1)})$ of z_3 such that
$$\big(\underline{u}_2(r_{2(b-1)}), \underline{x}_3(s_{3(b-1)}), \\ \underline{\hat{y}}_3(z_3|s_{3(b-1)}, r_{2(b-1)}), \underline{y}_{4(b-1)}\big)$$
is jointly typical. The intersection of this set with the z_3 in $S_{s_{3b}}$ determines $z_{3(b-1)}$. The correct $z_{3(b-1)}$ can be found reliably if n is large and (see (113))
$$R_3 > \hat{R}_3 + H(\hat{Y}_3|U_2X_3Y_4) - H(\hat{Y}_3|U_2X_3) \\ = \hat{R}_3 - I(\hat{Y}_3; Y_4|U_2X_3). \tag{125}$$
This is identical to step (ii) of the proof of [4, Thm. 6].
4) The destination terminal chooses $s_{2(b-1)}$ so that
$$\big(\underline{u}_2(r_{2(b-1)}), \underline{x}_2(s_{2(b-1)}|r_{2(b-1)}), \underline{x}_3(s_{3(b-1)}), \\ \underline{\hat{y}}_3(z_{3(b-1)}|s_{3(b-1)}, r_{2(b-1)}), \underline{y}_{4(b-1)}\big)$$
is jointly typical. Using Lemma 1, this step can be made reliable if
$$R_2 - R_2' < I(X_2; \hat{Y}_3Y_4|U_2X_3). \tag{126}$$
5) The destination terminal determines the set $\mathcal{L}(\underline{y}_{4(b-1)})$ of w such that
$$\big(\underline{u}_2(r_{2(b-1)}), \underline{x}_1(w|s_{2(b-1)}, r_{2(b-1)}), \underline{x}_2(s_{2(b-1)}|r_{2(b-1)}), \\ \underline{x}_3(s_{3(b-1)}), \underline{\hat{y}}_3(z_{3(b-1)}|s_{3(b-1)}, r_{2(b-1)}), \underline{y}_{4(b-1)}\big)$$
is jointly typical. The destination terminal knows (s_{2b}, r_{2b}), and generates the intersection of $\mathcal{L}(\underline{y}_{4(b-1)})$ with those w in $S_{r_{2b}, s_{2b}}$. The correct w_b can be found reliably if
$$R_1 < I(X_1; \hat{Y}_3Y_4|U_2X_2X_3) + R_2. \tag{127}$$
This is the analog of step (iii) of the proof of [4, Thm. 6].
6) Terminal 3 decodes r_{2b}. This can be done reliably if n is large and
$$R_2' < I(U_2; Y_3|X_3). \tag{128}$$
Finally, terminal 3 tries to find a z_3 such that $(\underline{\hat{y}}_3(z_3|s_{3b}, r_{2b}, \underline{y}_{3b}, \underline{x}_3(s_{3b}), \underline{u}_2(r_{2b}))$ is jointly typical. Such a z_3 exists with high probability for large n if
$$\hat{R}_3 > I(\hat{Y}_3; Y_3|U_2X_3). \tag{129}$$
This is the analog of step (iv) of [4, Thm. 6].

In summary, R_1 is bounded by (121) and (127). Inserting (126) into (127), we have
$$R_1 < \min\{I(X_1X_2; \hat{Y}_3Y_4|U_2X_3) + R_2', \\ I(X_1; Y_2|U_2X_2)\}. \tag{130}$$
Combining (129) with (125) yields
$$R_3 > I(\hat{Y}_3; Y_3|U_2X_3Y_4) \tag{131}$$
where we have used (40). Furthermore, R_2' and R_3 have to satisfy (122)–(124) and (128), i.e., we have
$$R_2' < \min\{I(U_2; Y_3|X_3), I(U_2; Y_4|X_3)\} \tag{132}$$
$$R_3 < I(X_3; Y_4|U_2) \tag{133}$$
$$R_2' + R_3 < I(U_2X_3; Y_4). \tag{134}$$
Equations (130)–(134) are the same as (35)–(39).

REFERENCES

[1] E. C. van der Meulen, *Transmission of Information in a T-Terminal Discrete Memoryless Channel*. Ph.D. thesis, Univ. of California, Berkeley, CA, June 1968.

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

[3] E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory*, vol. IT–23, pp. 1–37, Jan. 1977.

[4] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT–25, pp. 572–584, Sept. 1979.

[5] M. R. Aref, *Information Flow in Relay Networks*. Ph.D. thesis, Stanford Univ., Stanford, CA, Oct. 1980.

[6] A. A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inform. Theory*, vol. IT–28, p. 536, May 1982.

[7] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. on Inform. Theory*, vol. IT–34, pp. 1106–1110, Sept. 1988.

[8] P. Vanroose and E. C. van der Meulen, "Uniquely decodable codes for deterministic relay channels," *IEEE Trans. on Inform. Theory*, vol. IT–38, pp. 1203–1212, July 1992.

[9] R. Ahlswede and A. H. Kaspi, "Optimal coding strategies for certain permuting channels," *IEEE Trans. on Inform. Theory*, vol. IT–33, pp. 310–314, May 1987.

[10] K. Kobayashi, "Combinatorial structure and capacity of the permuting relay channel," *IEEE Trans. on Inform. Theory*, vol. IT–33, pp. 813–826, Nov. 1987.

[11] R. C. King, *Multiple Access Channels with Generalized Feedback*. Ph.D. thesis, Stanford Univ., Stanford, CA, March 1978.

[12] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Sys. Tech. J.*, vol. 52, pp. 1037–1076, Sept. 1973.

[13] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inform. Theory*, vol. IT–21, no. 1, pp. 100–102, Jan. 1975.

[14] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inform. Theory*, vol. IT–27, no. 3, pp. 292–298, May 1981.

[15] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inform. Theory*, vol. IT–28, no. 6, pp. 841–850, Nov. 1982.

[16] F. M. J. Willems, *Informationtheoretical Results for the Discrete Memoryless Multiple Access Channel*. Doctor in de Wetenschappen Proefschrift, Katholieke Universiteit Leuven, Leuven, Belgium, Oct. 1992.

[17] T. M. Cover and J. A. Thomas, *Elements of Inform. Theory*. New York: Wiley, 1991.

[18] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 388–404, March 2000.

[19] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1877-1894, Aug. 2003.

[20] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[21] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple level relay channel," *IEEE Trans. Inform. Theory*, submitted Nov. 2003.

[22] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," *Proc. 41st Annual Allerton Conf. on Commun., Control, and Computing*, (Monticello, IL), pp. 1074-1083, Oct. 1–3, 2003.

[23] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. Inform. Theory*, vol. IT–22, pp. 1–11, Jan. 1976.

[24] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: coding and antenna-clustering capacity," *Proc. IEEE Int. Symp. Inform. Theory*, (Lausanne, Switzerland), p. 136, June 30-July 5, 2002.

[25] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part I: System description," and "User cooperation diversity-Part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.

[26] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[27] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, to appear.

[28] J. N. Laneman, *Cooperative Diversity in Wireless Networks: Algorithms and Architectures*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, Aug. 2002.

[29] A. Stefanov and E. Erkip, "Cooperative information transmission in wireless networks," *Proc. Workshop on Concepts in Inform. Theory*, (Breisach, Germany), pp. 90–93, June 26–28, 2002.

[30] T. E. Hunter and A. Nosratinia, "Cooperation diversity through coding," *Proc. IEEE Int. Symp. Inform. Theory*, (Lausanne, Switzerland), p. 220, June 30-July 5, 2002.

[31] P. Herhold, E. Zimmermann, and G. Fettweis, "On the performance of cooperative amplify-and-forward relay networks," *Proc. 5th Int. ITG Conf. on Source and Channel Coding*, (Erlangen-Nuremberg, Germany), pp. 451–458, Jan. 14–16, 2004.

[32] B. Schein and R. G. Gallager, "The Gaussian parallel relay network," in *Proc IEEE Int. Symp. Inform. Theory*, (Sorrento, Italy), p. 22, June 2000.

[33] B. E. Schein, *Distributed Coordination in Network Information Theory*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, Oct. 2001.

[34] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proc. IEEE Int. Symp. Inform. Theory*, (Sorrento, Italy), p. 40, June 2000.

[35] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[36] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inform. Theory*, to appear.

[37] A. Høst-Madsen, "On the capacity of wireless relaying," *Proc. IEEE Vehic. Techn. Conf., VTC 2002 Fall*, (Vancouver, BC), vol. 3, pp. 1333–1337, Sept. 24–28, 2002.

[38] A. Reznik, S. R. Kulkarni, and S. Verdú, "Capacity and optimal resource allocation in the degraded Gaussian relay channel with multiple relays," *Proc. 40th Annual Allerton Conf. on Commun., Control, and Computing*, (Monticello, IL), Oct. 2–4, 2002.

[39] J. M. Shea, T. F. Wong, A. Avudainayagam, and X. Li, "Collaborative decoding on block-fading channels," *IEEE Trans. Commun.*, submitted Feb. 2003.

[40] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "On the capacity of 'cheap' relay networks," *Proc. 37th Annual Conf. on Information Sciences and Systems (CISS)*, (Baltimore, MD), March 12–14, 2003.

[41] M. O. Hasna and M.-S. Alouini, "Optimal power allocation for relayed transmissions over Rayleigh fading channels," *Proc. IEEE Vehic. Techn. Conf., VTC 2003 Spring*, vol. 4, pp. 2461–2465, April 22–25, 2003.

[42] M. O. Hasna and M.-S. Alouini, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 216–218, May 2003.

[43] S. Toumpis and A. J. Goldsmith, "Capacity regions for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 736–748, July 2003.

[44] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for the wireless relay channel," *IEEE Trans. Inform. Theory*, submitted Sept. 2003.

[45] R. U. Nabar, Ö. Oyman, H. Bölcskei, and A. J. Paulraj, "Capacity scaling laws in MIMO wireless networks," *Proc. 41st Annual Allerton Conf. on Commun., Control, and Computing*, (Monticello, IL), pp. 378-389, Oct. 1–3, 2003.

[46] U. Mitra and A. Sabharwal, "On achievable rates of complexity constrained relay channels," *Proc. 41st Annual Allerton Conf. on Commun., Control, and Computing*, (Monticello, IL), pp. 551-560, Oct. 1–3, 2003.

[47] B. Wang and J. Zhang, "MIMO relay channel and its application for cooperative communication in ad hoc networks," *Proc. 41st Annual Allerton Conf. on Commun., Control, and Computing*, (Monticello, IL), pp. 1556-1565, Oct. 1–3, 2003.

[48] M. O. Hasna and M.-S. Alouini, "End-to-end performance of transmission systems with relays over Rayleigh-fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1126–1131, Nov. 2003.

[49] R. U. Nabar and H. Bölcskei, "Space-time signal design for fading relay channels," *Proc. IEEE Global Telecommun. Conf. (GLOBECOM '03)*, (San Francisco, CA, USA), vol. 4, pp. 1952–1956, Dec. 1–5, 2003.

[50] Z. Dawy and H. Kamoun, "The general Gaussian relay channel: analysis and insights," *Proc. 5th Int. ITG Conf. on Source and Channel Coding*, (Erlangen-Nuremberg, Germany), pp. 469–476, Jan. 14–16, 2004.

[51] G. Kramer, M. Gastpar, and P. Gupta, "Information-theoretic multi-hopping for relay networks," *Proc. 2004 Int. Zurich Seminar*, (Zurich, Switzerland), pp. 192-195, Feb. 18-20, 2004.

[52] B. Wang, J. Zhang, and A. Høst-Madsen, "On the ergodic capacity of MIMO relay channel," *Proc. 38th Annual Conf. on Information Sciences and Systems (CISS)*, (Princeton, NJ), pp. 603–608, March 17–19, 2004.

[53] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

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

[55] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 27, pp. 120–122, Jan. 1981.

[56] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*, Budapest: Akadémiai Kiadó, 1981.

[57] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. 26, pp. 648–657, Nov. 1980.

[58] M. Gastpar, "The Wyner-Ziv problem with multiple sources," submitted to *IEEE Trans. Inform. Theory*, June 2002.

[59] M. Gastpar, *To Code or Not To Code*. Doctoral Thesis, Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland, Dec. 2002.

[60] G. J. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi–element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[61] I. E. Telatar, "Capacity of multi–antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[62] M. Debbah, "MIMO channel modeling and the principle of maximum entropy, Part I: Model construction and Part II: Model validation," *IEEE Trans. Inform. Theory*, submitted June 2003.

[63] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

[64] J. A. Thomas, "Feedback can at most double Gaussian multiple access channel capacity," *IEEE Trans. Inform. Theory*, vol. 33, no. 5, pp. 711–716, Sept. 1987.

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

[66] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[67] B. Zhao and M. C. Valenti, "Distributed turbo coded diversity for relay channel," *Electron. Lett.*, vo. 39, no. 10, pp. 786–787, May 2003.