Cooperative Strategies and Capacity Theorems for Relay Networks

Gerhard Kramer, *Member, IEEE*, Michael Gastpar, *Member, IEEE*, and Piyush Gupta, *Member, IEEE*

***Abstract*—Coding strategies that exploit node cooperation are developed for relay networks. Two basic schemes are studied: the relays decode-and-forward the source message to the destination, or they compress-and-forward their channel outputs to the destination. The decode-and-forward scheme is a variant of multihopping, but in addition to having the relays successively decode the message, the transmitters cooperate and each receiver uses several or all of its past channel output blocks to decode. For the compress-and-forward scheme, the relays take advantage of the statistical dependence between their channel outputs and the destination's channel output. The strategies are applied to wireless channels, and it is shown that decode-and-forward achieves the ergodic capacity with phase fading if phase information is available only locally, and if the relays are near the source node. The ergodic capacity coincides with the rate of a distributed antenna array with full cooperation even though the transmitting antennas are not colocated. The capacity results generalize broadly, including to multiantenna transmission with Rayleigh fading, single-bounce fading, certain quasi-static 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 multisource and multidestination networks such as multiaccess and broadcast relay channels.**

***Index Terms*—Antenna arrays, capacity, coding, multiuser 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 multihop wireless network or a sensor network where nodes have limited power to transmit data. We 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 in pioneering work by van der Meulen [1], [2] (see also [3, Sec IX]). Substantial advances in the theory were made by Cover and El Gamal, who developed two fundamental coding strategies for one relay [4, Theorems 1 and 6]. A combination of these strategies [4, Theorem 7] achieves capacity for several classes of channels, as discussed in [4]–[10]. Capacity-achieving codes appeared in [11] for deterministic relay channels, and in [12], [13] for permuting relay channels with states or memory. We will consider only random coding and concentrate on generalizing the two basic strategies of [4].

A. Decode-and-Forward

The first strategy achieves the rates in [4, Theorem 1], and it is one of a class of schemes now commonly called *decode-and-forward* ([14, p. 64], [15, p. 82], [16, p. 76]). We consider three different decode-and-forward strategies, and we call these

- irregular encoding/successive decoding,
- regular encoding/sliding-window decoding,
- regular encoding/backward decoding.

The first strategy is that of [4], where the authors use block Markov superposition encoding, random partitioning (binning), and *successive* decoding. The encoding is done using codebooks of different sizes, hence the name *irregular* encoding.

The other two strategies were developed in the context of the *multiaccess channel with generalized feedback* (MAC-GF) studied by King [5]. King's MAC-GF has three nodes like a single-relay channel, but now two of the nodes transmit messages to the third node, e.g., nodes 1 and 2 transmit the respective messages M_1 and M_2 to node 3. Nodes 1 and 2 further receive a *common* channel output Y^* that can be different than node 3's output Y. King developed an achievable rate region for this channel that generalizes results of Slepian and Wolf [17], Gaarder and Wolf [18], and Cover and Leung [19].

Carleial extended King's model by giving the transmitters *different* channel outputs Y_1 and Y_2 [20]. We follow Willems' convention and refer to this extended model as a MAC-GF [21]. Carleial further derived an achievable rate region with 17 bounds [20, eqs. (7a)–(7q)]. Although this region can be difficult to evaluate, there are several interesting features of the approach. First, his 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 $Y_1 = Y_2$ [5, p. 36]). Second, Carleial achieves the same rates as in [4, Theorem 1] by an appropriate choice of the random variables in [20, eq.(7)] (see Remark 6 below). 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 destination employs a *sliding-window* decoding technique that uses two consecutive blocks of channel outputs [20, p. 842]. Hence, the name *regular* encoding/*sliding-window* decoding.

Manuscript received March 2, 2004; revised June 2005. The work of G. Kramer and P. Gupta was supported in part by the Board of Trustees of the University of Illinois Subaward no. 04-217 under National Science Foundation Grant CCR-0325673. The work of M. Gastpar was supported in part by the National Science Foundation under Awards CCF-0347298 (CAREER) and CNS-0326503. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002; the 41st Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2003; and the 2004 International Zurich Seminar, Zurich, Switzerland.

G. Kramer and P. Gupta are with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: 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 USA (e-mail: gastpar@eecs.berkeley.edu).

Communicated by A. Lapidoth, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2005.853304

0018-9448/$20.00 © 2005 IEEE

المنارة للاستشارات
www.manaraa.com

The third decode-and-forward strategy is based on work for the MAC-GF by Willems [21, Ch. 7]. Willems introduced a *backward* decoding technique that is better than sliding-window decoding in general [22]–[24], but for the relay channel regular encoding/backward decoding achieves the same rates as irregular encoding/successive decoding and regular encoding/sliding-window decoding.

Subsequent work focused on generalizing the strategies to multiple relays. Aref extended irregular encoding/successive decoding to degraded relay networks [7, Ch. 4], [8]. He further developed capacity-achieving binning strategies for deterministic broadcast relay networks and deterministic relay networks without interference. The capacity proofs relied on a (then new) *cut-set* bound [4, Theorem 4], [7, p. 23] that has become a standard tool for bounding capacity regions (see [25, p. 445]).

Recent work on decode-and-forward for multiple relays appeared in [14], [26]–[35]. In particular, Gupta and Kumar [31] applied irregular encoding/successive decoding to multiple-relay networks in a manner similar to [7]. The methodology was further extended to multisource 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 [31, p. 1883]). We interpret the relaying approach of [7], [31] as a variant of *multihopping*, i.e., the relays successively decode the source messages before these arrive at the destinations. However, in addition to the usual multihopping, 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.

Next, Xie and Kumar [32], [33] developed regular encoding/sliding-window decoding for multiple relays, and showed that their scheme achieves better rates than those of [7], [31]. One can similarly generalize regular encoding/backward decoding [34]. The achievable *rates* of the two regular encoding strategies turn out to be the same. However, the *delay* of sliding-window decoding is much less than that of backward decoding. Regular encoding/sliding-window decoding is therefore currently the preferred variant of multihopping in the sense that it achieves the best rates in the simplest way.

B. Compress-and-Forward

The second strategy of Cover and El Gamal is now often called *compress-and-forward*, although some authors prefer the names estimate-and-forward (based on [4, p. 581]), observe-and-forward [14], or quantize-and-forward [36]. King [5, p. 33] attributes van der Meulen with motivating the approach (see [2, Example 4]). The idea is that a relay, say node t, transmits a quantized and compressed version $\hat{Y}_t$ of its channel output Y_t to the destination, and the destination decodes by combining $\hat{Y}_t$ with its own output. The relays and destination further take advantage of the statistical dependence between the different outputs. More precisely, the relays use Wyner–Ziv source coding to exploit side information at the destination [37].

The compress-and-forward strategy was generalized to the MAC-GF by King [5, Ch. 3], to parallel relay networks by Schein and Gallager [27], [28], and to multiple relays by Gastpar *et al.* [29] (see [38], [39]). One can also mix the decode-and-forward and compress-and-forward strategies. This was done in [4, Theorem 7], for example.

C. Outline

This paper develops decode-and-forward and compress-and-forward strategies for relay networks with many relays, antennas, sources, and destinations. Section II summarizes our main results.

The technical sections are divided into two main parts. The first part, comprising Sections III to VI, deals with general relay channels. In Section III, we define several models and review a capacity upper bound. In Section IV, we develop decode-and-forward strategies for relay channels, multiaccess relay channels (MARCs), and broadcast relay channels (BRCs). Section V develops compress-and-forward strategies for multiple relays. Section VI describes mixed strategies where each relay uses either decode-and-forward or compress-and-forward.

The second part of the paper is Section VII that specializes the theory to wireless networks with geometries (distances) and fading. We show that one approaches capacity when the network nodes form two closely spaced clusters. We then study channels with phase fading, and where phase information is available only locally. We show that decode-and-forward achieves the ergodic capacity when all relays are near the source node. The capacity results generalize to certain quasi-static models, and to MARCs and BRCs. Section VIII 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, e.g., [40]–[49]. We also do not consider the popular *amplify-and-forward* strategies in much detail (see Fig. 16 and, e.g., [14]–[16], [27], [28], [36], [50], [51]). Many other recent results can be found in [26]–[36], [38]–[76], and references therein (see especially [72]–[74] for new capacity theorems).

II. Summary of Main Results

We summarize the main contributions of this paper. These results were reported in the conference papers [29], [34], [67].

- Theorem 1 gives an achievable rate for multiple relays. The main idea behind this theorem is due to Xie and Kumar [32] who studied regular encoding and sliding-window decoding for Gaussian channels. We extended their result to discrete memoryless and fading channels in [34], as did [33]. Our description of the rate was somewhat more compact than that of [32], and it showed that the level sets of [31], [32] are not necessary to maximize the *rate*. However, level sets can reduce the *delay* and they might improve rates for multiple source–destination pairs.
- We introduce the BRC as the natural counterpart of the MARC described in [24], [52]. Theorem 2 gives an achievable rate region for BRCs that combines regular encoding and sliding-window decoding.
- Theorem 3 generalizes the compress-and-forward strategy of [4, Theorem 6] to any number of relays. As pointed out in [29], [38], [39], the Wyner–Ziv

www.manaraa.com

problem with multiple sources appears naturally in this framework.

- Theorem 4 gives an achievable rate when the relays use either decode-and-forward or compress-and-forward.
- Theorem 5 adds partial decoding to Theorem 4 when there are two relays.

For Gaussian channels, the following general principle is established in a series of theorems: decode-and-forward achieves the ergodic capacity when there is phase fading (i.e., any fading process such as Rayleigh fading where the phase varies uniformly over time and the interval $[0, 2\pi)$), the transmitting nodes are close to each other (but not necessarily colocated), and phase information is available at the receivers only (or is available from nearby nodes). Perhaps unexpectedly, the capacity is the rate of a distributed antenna array even if the transmitting antennas are not colocated. The paper [75] further argues that internode phase and symbol synchronization is not needed to achieve the capacity (see Remark 42).

- Section VII-C points out that one achieves capacity if the network nodes form two closely spaced clusters.
- Theorems 6 and 7 establish the ergodic capacity claim for the simplest kind of phase fading, and for one or more relays. For example, decode-and-forward achieves capacity for $T-2$ relays if these are all located within a circle of radius about $T^{-1/\alpha}$ about the source, where α is the channel attenuation exponent. As another example, if $T-2$ relays are placed at regular intervals between the source and destination, then one can achieve a capacity of about $\alpha \log_2 T$ bits per use for large T.
- Theorem 8 establishes the capacity claim for Rayleigh fading, multiple antennas, and many relays. We remark that a special case of Theorem 8 appeared in parallel work by Wang, Zhang, and Høst-Madsen [68], [76], where one relay is considered. As kindly pointed out by these authors in [76, p. 30], the proof of their Theorem 4.1, and hence their Proposition 5.1, is based on our results in [67] that generalize [34, Theorem 2].
- Section VII-G outlines extensions of the theory to fading with directions.
- Theorems 9 and 10 establish the capacity claim for certain MARCs and BRCs, and Section VII-H describes generalizations to more complex cases.
- Section VII-I derives results for quasi-static fading.

We here develop the capacity theorems for *full-duplex* relays, i.e., relays that can transmit and receive at the same time and in the same frequency band. However, we emphasize that the theory does extend to *half-duplex* relays, i.e., relays that cannot transmit and receive at the same time, by using the memoryless models described in [75]. For example, all of the theory of Cover and El Gamal [4] and all of the theory developed in Sections III–VI applies to half-duplex relays. Moreover, our geometric capacity results apply to half-duplex relays if practical constraints are placed on the coding strategies (see Remarks 29 and 38). More theory for such relays can be found in [14], [36], [51]–[55], [59]–[62], [75], and references therein.

A third type of relay is one that can transmit and receive at the same time, but only in different frequency bands. Such relays

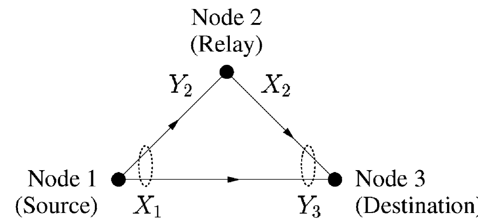

Fig. 1. A one-relay network.

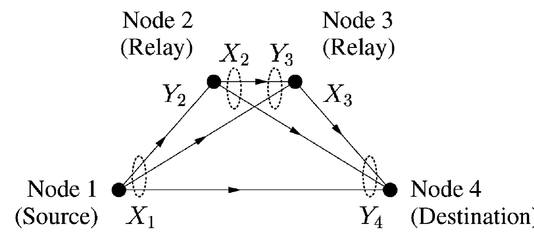

Fig. 2. A two-relay network.

are treated in [73], [74], where the authors constrain the source node to use only one of the two frequency bands available to it.

III. Models and Preliminaries

A. Relay Network Model

Consider the network model of [1, Fig. 1]. The T-node relay network has a source terminal (node 1), $T-2$ relays (nodes t with $t \in \{2, 3, \ldots, T-1\}$), and a destination terminal (node T). For example, relay networks with $T=3$ and $T=4$ are shown as graphs in Figs. 1 and 2. The network random variables are

- a message W,
- channel inputs $X_{ti}, t = 1, 2, \ldots, T-1, i = 1, 2, \ldots N$,
- channel outputs $Y_{ti}, t = 2, 3, \ldots, T, i = 1, 2, \ldots N$,
- a message estimate $\hat{W}$.

The X_{1i} are functions of W, and the $X_{ti}, t \neq 1$, are functions of node t's past outputs $Y_t^{i-1} = Y_{t1}, Y_{t2}, \ldots, Y_{t(i-1)}$. We write $p_{A\,|\,B}(a\,|\,b)$ for the conditional probability that $A = a$ given $B = b$, or simply $p(a\,|\,b)$ when the arguments are lower case versions of the corresponding random variables. The networks we consider are *memoryless* and *time invariant* in the sense that

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

for all i, where the X_t and $Y_t, t = 1, \ldots, T$, are random variables representing the respective channel inputs and outputs. The condition (1) lets one focus on the channel distribution

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

for further analysis.

The destination computes its message estimate $\hat{W}$ as a function of Y_T^N (recall that $Y_T^N = Y_{T1}, Y_{T2}, \ldots, Y_{TN}$). Suppose that W has B_W bits. The *capacity* C is the supremum of rates $R_W = B_W/N$ at which the destination's message estimate $\hat{W}$ can be made to satisfy $\Pr(\hat{W} \neq W) < \epsilon$ for any positive ϵ.

For example, consider the $T = 3$ node network of Fig. 1, and suppose the source (node 1) is *wired* to the relay and des-

المنارة للاستشارات
www.manaraa.com

tination (nodes 2 and 3, respectively). We write the channel inputs as $X_1 = [X_{11}, X_{12}]$ and X_2, and the outputs as Y_2 and $Y_3 = [Y_{31}, Y_{32}]$. An appropriate channel distribution for wired communication might be

$$p(y_2, y_3 \,|\, x_1, x_2) = p(y_2 \,|\, x_{11}) \cdot p(y_{31} \,|\, x_{12}) \cdot p(y_{32} \,|\, x_2). \quad (3)$$

Consider next the $T = 4$ node network of Fig. 2, and suppose the source (node 1) is wired to two relays (nodes 2 and 3) and the destination (node 4). The channel inputs are $X_1 = [X_{11}, X_{12}, X_{13}]$, $X_2 = [X_{21}, X_{22}]$ and X_3, and the outputs are $Y_2, Y_3 = [Y_{31}, Y_{32}]$ and $Y_4 = [Y_{41}, Y_{42}, Y_{43}]$. An appropriate channel distribution might be

$$\begin{aligned} &p(y_2, y_3, y_4 \,|\, x_1, x_2, x_3) \\ &\quad = p(y_2 \,|\, x_{11}) \cdot p(y_{31} \,|\, x_{12}) \cdot p(y_{32} \,|\, x_{21}) \\ &\qquad \cdot p(y_{41} \,|\, x_{13}) \cdot p(y_{42} \,|\, x_{22}) \cdot p(y_{43} \,|\, x_3). \end{aligned} \quad (4)$$

B. Capacity Upper Bound

We use the common notation $H(X), H(X \,|\, Y), I(X;Y)$, and $I(X;Y \,|\, Z)$ for the respective entropy of X, the entropy of X conditioned on Y, the mutual information between X and Y, and the mutual information between X and Y conditioned on Z [77, Ch. 2] (cf. [78, Ch. 1.1]). Let $X_{\mathcal{S}} = \{X_t : t \in \mathcal{S}\}$ and $\mathcal{T} = \{2, 3, \ldots, T-1\}$. A capacity upper bound is given by the cut-set bound in [7, p. 23] (see also [8, Theorem 1] and [25, p. 445]).

Proposition 1: The relay network capacity satisfies

$$C \le \max_{p(x_1, x_2, \ldots, x_{T-1})} \min_{\mathcal{S} \subseteq \mathcal{T}} I\left(X_1 X_{\mathcal{S}}; Y_{\mathcal{S}^C} Y_T \,|\, X_{\mathcal{S}^C}\right) \quad (5)$$

where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{T}$.

For example, for $T = 3$ the bound (5) is

$$C \le \max_{p(x_1, x_2)} \min\{I(X_1; Y_2 Y_3 \,|\, X_2), I(X_1 X_2; Y_3)\}. \quad (6)$$

Applying (6) to the wired networks (3), one recovers a standard flow cut-set bound

$$C \le \min\{C_{12} + C_{13}, C_{13} + C_{23}\} \quad (7)$$

where C_{ij} is the capacity of the channel from node i to node j [79, p. 179].

Remark 1: The set of $p_{X_1 \ldots X_{T-1}}(\cdot)$ is convex, and the mutual information expressions in (5) are concave in $p_{X_1 \ldots X_{T-1}}(\cdot)$ [25, p. 31]. Furthermore, the point-wise minimum of a collection of concave functions is concave [80, p. 35]. One can thus perform the maximization in (5) efficiently with convex optimization algorithms.

C. MARC Model

The relay networks described above have one message W. We will consider two networks with several messages: MARCs and BRCs. A MARC with two sources has four nodes: nodes 1 and 2 transmit the independent messages W_1 and W_2 at rates R_1 and R_2, respectively, node 3 acts as a relay, and node 4 is the destination for both messages [24], [52]. This model might fit a situation where sensors (the sources) are too weak to cooperate, but they can send their data to more powerful nodes that form a "backbone" network. The MARC has three inputs X_1, X_2, X_3 and two outputs Y_3, Y_4 (see Fig. 3). The channel distribution is

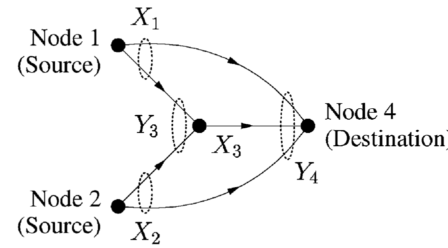

Fig. 3. A MARC with two sources. Node 3 is the relay.

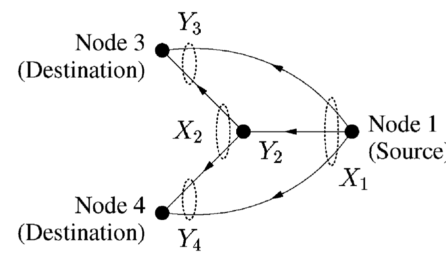

Fig. 4. A BRC with two destinations. Node 2 is the relay.

$$p(y_3, y_4 \,|\, x_1, x_2, x_3) \quad (8)$$

where X_1 and X_2 are independent. The capacity *region* is the closure of the set of rate pairs (R_1, R_2) at which the two sources can transmit W_1 and W_2 reliably to the destination.

Suppose, for example, that the sources (nodes 1 and 2) are wired to the relay (node 3) and the destination (node 4). The channel inputs are $X_1 = [X_{11}, X_{12}]$ and $X_2 = [X_{21}, X_{22}]$, and the outputs are $Y_3 = [Y_{31}, Y_{32}]$ and $Y_4 = [Y_{41}, Y_{42}, Y_{43}]$. An appropriate channel distribution might be

$$\begin{aligned} &p(y_3, y_4 \,|\, x_1, x_2, x_3) \\ &\quad = p(y_{31} \,|\, x_{11}) \cdot p(y_{32} \,|\, x_{21}) \\ &\qquad \cdot p(y_{41} \,|\, x_{12}) \cdot p(y_{42} \,|\, x_{22}) \cdot p(y_{43} \,|\, x_3). \end{aligned} \quad (9)$$

D. BRC Model

A BRC with two sinks and three independent messages W_0, W_1, W_2 is depicted in Fig. 4. Node 1 transmits W_0 at rate R_0 to both nodes 3 and 4, W_1 at rate R_1 to node 3, and W_2 at rate R_2 to node 4. Node 2 acts as a relay. Such a model might fit a scenario where a central node forwards instructions to a number of agents (the destinations) via a relay [67], [71]. The channel distribution has the form

$$p(y_2, y_3, y_4 \,|\, x_1, x_2). \quad (10)$$

The capacity region is the closure of the set of rate triples (R_0, R_1, R_2) at which the source can transmit (W_0, W_1) and (W_0, W_2) reliably to nodes 3 and 4, respectively.

Suppose, e.g., that in Fig. 4 the channel inputs are $X_1 = [X_{11}, X_{12}, X_{13}]$ and $X_2 = [X_{21}, X_{22}]$, and the outputs

المنارة للاستشارات
www.manaraa.com

are $Y_2, Y_3 = [Y_{31}, Y_{32}]$ and $Y_4 = [Y_{41}, Y_{42}]$. An appropriate channel distribution for wired communication might be

$$p(y_2, y_3, y_4 \,|\, x_1, x_2) = p(y_2 \,|\, x_{11}) \cdot p(y_{31} \,|\, x_{12}) \cdot p(y_{32} \,|\, x_{21}) \cdot p(y_{41} \,|\, x_{13}) \cdot p(y_{42} \,|\, x_{22}). \tag{11}$$

IV. Decode-and-Forward

The decode-and-forward strategies have as a common feature that the source controls what the relays transmit. For wireless networks, one consequently achieves gains related to multiantenna *transmission*. We label the corresponding rates $R_{\rm DF}$.

A. Rates for One Relay

The decode-and-forward strategy of Cover and El Gamal [4, Theorem 1] achieves any rate up to

$$R_{\rm DF} = \max_{p(x_1,x_2)} \min\{I(X_1; Y_2 \,|\, X_2), I(X_1 X_2; Y_3)\}. \tag{12}$$

The difference between (6) and (12) is that Y_3 is included in the first information expression on the right-hand side of (6).

Remark 2: One can apply Remark 1 to (12), 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 expression in (12) can be interpreted as the information between two transmit antennas X_1 and X_2, and one receive antenna Y_3 [28, p. 15], [29]. Decode-and-forward achieves the cooperative gain reflected by the maximization over all joint distributions $p(x_1, x_2)$.

Remark 4: The rate (12) requires the relay to decode the source message, and this can be a rather severe constraint. For example, suppose Fig. 1 represents a network of discrete memoryless channels (DMCs) defined by (3). Suppose that X_{11}, X_{12} and X_2 are binary, and that $Y_2 = X_{11}, Y_{31} = X_{12}$, and $Y_{32} = X_2$. The capacity is clearly 2 bits per use, but (12) gives only 1 bit per use.

Remark 5: One can generalize (12) by allowing the relay to *partially decode* the message. This is done in [4, Theorem 7] and [9] by introducing a random variable, say U, that represents the information decoded by the relay. The strategy of [9] (which is a special case of [4, Theorem 7]) achieves rates up to

$$R_{\rm PDF} = \min\{I(U; Y_2 \,|\, X_2) + I(X_1; Y_3 \,|\, X_2 U), I(X_1 X_2; Y_3)\} \tag{13}$$

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 (12). Moreover, the rate (13) is the capacity of the network (3) by choosing U, X_{12}, and X_2 as being independent, and $X_{11} = U$.

B. Regular Encoding for One Relay

The rate (12) has in the past been achieved with three different methods, as discussed in the Introduction. We refer to [4] for a description of the irregular encoding/successive decoding strategy. We instead review the regular encoding approach of [20], [21] that is depicted in Figs. 5 and 6.

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. 5. Decode-and-forward for one relay. The source and relay transmit the respective codewords $\underline{x}_1(\cdot)$ and $\underline{x}_2(\cdot)$.

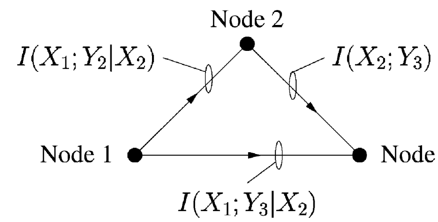

Fig. 6. The information transfer for one relay and regular encoding/sliding window decoding.

The message w is divided into B blocks $w_1, w_2, \ldots, w_B$ of 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}. We thus have $B_W = BnR, N = (B+1)n$, and $R_W = R \cdot B/(B+1)$, where we recall that B_W is the number of message bits, N is the total number of channel uses, and R_W is the overall rate. 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

$$\underline{x}_1(i,j) = \alpha \underline{x}_2(i) + \beta \underline{x}_1'(j) \tag{14}$$

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

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

$$R < I(X_1; Y_2 \,|\, X_2) \tag{15}$$

where we assume that $I(X_1; Y_2 \,|\, X_2)$ is positive (we similarly assume that the other mutual information expressions below are positive). The information transfer is depicted in Fig. 6, where the edge between nodes 1 and 2 is labeled $I(X_1; Y_2 \,|\, X_2)$. So suppose node 2 correctly obtains w_1. In the second block, node 1 transmits $\underline{x}_1(w_1, w_2)$ and node 2 transmits $\underline{x}_2(w_1)$. Node 2 can decode w_2 reliably as long as n is large and (15) is true. One continues in this way until block $B+1$. In this last block, node 1 transmits $\underline{x}_1(w_B, 1)$ and node 2 transmits $\underline{x}_2(w_B)$.

Consider now node 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. Node 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 [21, Ch. 7]) that node 3 can decode reliably as long as n is large and

$$R < I(X_1 X_2; Y_3). \tag{16}$$

Suppose node 3 has correctly decoded w_B. Node 3 next decodes w_{B-1} from $\underline{y}_{3B}$ that depends on $\underline{x}_1(w_{B-1}, w_B)$ and $\underline{x}_2(w_{B-1})$. Since node 3 knows w_B, it can again decode reliably as long as (16) is true. One continues in this fashion until all message blocks have been decoded. The overall rate is $R \cdot B/(B+1)$

www.manaraa.com

bits per use, and by making B large one can get the rate as close to R as desired.

Finally, we describe Carleial's *sliding-window* decoding technique [20]. Consider again Fig. 5, but suppose node 3 decodes w_1 after block 2 by using a sliding window of the two past received blocks $\underline{y}_{31}$ and $\underline{y}_{32}$. One can show (see [20], [32], [33]) that node 3 can decode reliably as long as n is large and

$$\begin{aligned} R &< I(X_2;Y_3) + I(X_1;Y_3\,|\,X_2) \\ &= I(X_1X_2;Y_3). \end{aligned} \tag{17}$$

The contribution $I(X_2;Y_3)$ in (17) is from $\underline{y}_{32}$, and the contribution $I(X_1;Y_3\,|\,X_2)$ is from $\underline{y}_{31}$. The information transfer is shown in Fig. 6 by the labels of the incoming edges of node 3. After receiving $\underline{y}_{3b}, b \geq 3$, node 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 one can get the rate as close to R as desired.

Remark 6: One recovers (12) from Carleial's region [20, eq. (7)] as follows. The relay channel problem corresponds to $Y_1 = 0$ and $R_{20} = R_{22} = R_2 = 0$, and we obtain (12) by choosing $R_{10} = R_{\text{DF}}, R_{11} = 0, U_1 = X_1, W_1 = X_2$, and $Q = V_1 = U_2 = V_2 = W_2 = 0$.

Remark 7: Carleial's strategy enjoys the advantages of both the Cover/El Gamal strategy (two block decoding delay) and the Carleial/Willems strategy (regular encoding). Furthermore, as discussed later, regular encoding and sliding-window decoding extend to multiple relays in a natural way.

Remark 8: The papers [32], [33] derive the bound (17) by generating different random code books for consecutive blocks. This is done to create statistical independence between the blocks. We use the same approach in Appendix B.

C. Multiple Relays

One approach to decode-and-forward with several relays is to generalize the irregular encoding/successive decoding strategy. This was done in [7], [31]. However, we here consider only the improved strategy of [32]–[34] that is depicted in Figs. 7 and 8 for two relays.

Consider two relays. We divide the message w into B blocks of 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. 7. Node 2 can reliably decode w_b after transmission block b if n is large, its past message estimates $(\hat{w}_{b-2}^{(2)}, \hat{w}_{b-1}^{(2)})$ of (w_{b-2}, w_{b-1}) are correct, and

$$R < I(X_1;Y_2\,|\,X_2X_3). \tag{18}$$

The information transfer is depicted in Fig. 8, where the edge between nodes 1 and 2 is labeled $I(X_1;Y_2\,|\,X_2X_3)$.

Node 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 $\hat{w}_{b-3}^{(3)}, \hat{w}_{b-2}^{(3)}$ are correct, and

$$\begin{aligned} R &< I(X_2;Y_3\,|\,X_3) + I(X_1;Y_3\,|\,X_2X_3) \\ &= I(X_1X_2;Y_3\,|\,X_3). \end{aligned} \tag{19}$$

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. 7. Decode-and-forward for two relays. The source, first relay, and second relay transmit the respective codewords $\underline{x}_1(\cdot)$, $\underline{x}_2(\cdot)$, and $\underline{x}_3(\cdot)$.

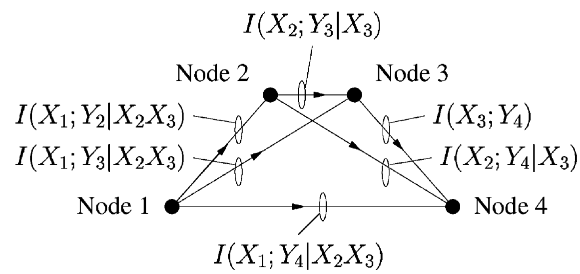

Fig. 8. The information transfer for two relays and regular encoding/sliding window decoding.

The contribution $I(X_2;Y_3\,|\,X_3)$ is from $\underline{y}_{3b}$, and the contribution $I(X_1;Y_3\,|\,X_2X_3)$ is from $\underline{y}_{3(b-1)}$. The information transfer is depicted in Fig. 8 by the labels of the incoming edges of node 3. Assuming correct decoding, node 3 knows w_{b-1} after transmission block b, and can encode the messages as shown in Fig. 7.

Finally, node 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 $\hat{w}_{b-4}^{(4)}, \hat{w}_{b-3}^{(4)}$ are correct, and

$$\begin{aligned} R &< I(X_3;Y_4) + I(X_2;Y_4\,|\,X_3) + I(X_1;Y_4\,|\,X_2X_3) \\ &= I(X_1X_2X_3;Y_4). \end{aligned} \tag{20}$$

The contribution $I(X_3;Y_4)$ is from $\underline{y}_{4b}$, $I(X_2;Y_4\,|\,X_3)$ is from $\underline{y}_{4(b-1)}$, and $I(X_1;Y_4\,|\,X_2X_3)$ is from $\underline{y}_{4(b-2)}$. The information transfer is shown in Fig. 8 by the labels of the incoming edges of node 4. 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 sliding-window decoding generalizes to T-node relay networks, and one can prove the following theorem. Let $\pi(\cdot)$ be a permutation on $\{1,2,\ldots,T\}$ with $\pi(1) = 1$ and $\pi(T) = T$, and let $\pi(i:j) = \{\pi(i), \pi(i+1), \ldots, \pi(j)\}$.

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

$$R_{\text{DF}} = \max_{\pi(\cdot)} \min_{1 \leq t \leq T-1} I\left(X_{\pi(1:t)}; Y_{\pi(t+1)} \middle| X_{\pi(t+1:T-1)}\right) \tag{21}$$

where one can choose any distribution on $X_1, X_2, \ldots, X_{T-1}$.

Proof: Theorem 1 is essentially due to Xie and Kumar, and appeared for additive white Gaussian noise (AWGN) channels in [32]. The result appeared for more general classes of channels in [33], [34]. Proofs can be found in [32], [33] where the authors show that (21) improves on the rate of [33, Theorem 3.1]. □

Remark 9: We have expressed Theorem 1 using only permutations rather than the level sets of [31]–[33]. This is because one need consider only permutations to maximize the *rate*, i.e., one

المنارة للاستشارات
www.manaraa.com

can restrict attention to flowgraphs that have only *one* vertex per level set. However, to minimize the *delay* for a given rate, one might need to consider level sets again. This occurs, e.g., if one relay is at the same location as another.

Remark 10: Theorem 1 appeared for *degraded* relay networks in [7, p. 69], where degradation was defined as

$$p\left(y_{\pi(t)}\,\middle|x_{\pi(1:T-1)}, y_{\pi(2:t-1)}\right) = p\left(y_{\pi(t)}\,\middle|x_{\pi(t-1)}, x_{\pi(t)}, y_{\pi(t-1)}\right) \quad (22)$$

for $t = 2, 3, \ldots, T$, where $\pi(\cdot)$ is some permutation on $\{1, 2, 3, \ldots, T\}$ with $\pi(1) = 1$ and $\pi(T) = T$ (see [7, p. 54], and also [4, eq.(10)] and [35]). Moreover, $R_{\rm DF}$ is the capacity region of such channels [7, p. 69]. One can, in fact, replace (22) by the more general

$$p\left(y_{\pi(t:T)}\,\middle|x_{\pi(1:T-1)}, y_{\pi(t-1)}\right) = p\left(y_{\pi(t:T)}\,\middle|x_{\pi(t-1:T-1)}, y_{\pi(t-1)}\right) \quad (23)$$

for $t = 3, 4, \ldots, T$. The condition (23) simply makes the cut-set bound of Proposition 1 the same as $R_{\rm DF}$ [33].

Remark 11: We can apply Remark 1 to (21), i.e., convex optimization algorithms can efficiently perform the maximization over $p(x_1, \ldots, x_{T-1})$.

Remark 12: Backward decoding also achieves $R_{\rm DF}$, but for multiple relays one must transmit using nested blocks to allow the intermediate relays (e.g., node 3 in Fig. 7) to decode before the destination. The result is an excessive decoding delay.

Remark 13: Theorem 1 illustrates the multiantenna transmission behavior: the rate (21) can be interpreted as the mutual 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 somewhat by the compress-and-forward strategy developed below.

Remark 14: One can generalize Theorem 1 and let the relays perform *partial* decoding (see [4, Theorem 7], [9]). We will not consider this possibility here, but later do consider a restricted form of partial decoding.

D. MARCs

Consider a MARC with two sources (see Fig. 3). A regular encoding strategy is depicted in Figs. 9 and 10, and it proceeds as follows.

The message w_t is divided into B blocks $w_{t1}, w_{t2}, \ldots, w_{tB}$ of 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 details of the codebook construction, encoding and decoding are given in Appendix A. It turns out that node 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

$$\begin{aligned} 0 &\le R_1 < I(X_1; Y_3 \,|\, U_1 U_2 X_2 X_3) \\ 0 &\le R_2 < I(X_2; Y_3 \,|\, U_1 U_2 X_1 X_3) \\ R_1 + R_2 &< I(X_1 X_2; Y_3 \,|\, U_1 U_2 X_3) \end{aligned} \quad (24)$$

where

$$p(u_1, u_2, x_1, x_2, x_3) = p(u_1, x_1) p(u_2, x_2) p(x_3 \,|\, u_1, u_2).$$

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. 9. Decode-and-forward for a MARC. The first and second source transmit $\underline{x}_1(\cdot)$ and $\underline{x}_2(\cdot)$, respectively. The relay's codeword $\underline{x}_3(\cdot)$ is statistically dependent on $\underline{x}_1(\cdot)$ and $\underline{x}_2(\cdot)$ through $\underline{u}_1(\cdot)$ and $\underline{u}_2(\cdot)$, which are auxiliary codewords at the respective sources 1 and 2.

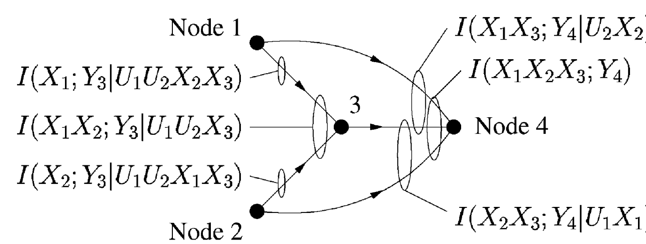

Fig. 10. The information transfer between the nodes of a MARC when using regular encoding/backward decoding.

The information transfer is depicted in Fig. 10, where the incoming edges of node 3 are labeled by the mutual information expressions in (24).

Suppose node 4 uses backward decoding. One finds that this node can decode reliably if

$$\begin{aligned} 0 &\le R_1 < I(X_1 X_3; Y_4 \,|\, U_2 X_2) \\ 0 &\le R_2 < I(X_2 X_3; Y_4 \,|\, U_1 X_1) \\ R_1 + R_2 &< I(X_1 X_2 X_3; Y_4). \end{aligned} \quad (25)$$

The information transfer is shown in Fig. 10, where the incoming edges of node 4 are labeled by the mutual information expressions in (25). The achievable rates of (24) and (25) were derived in [52, eq. (5)].

Remark 15: We can use the flowgraphs of [31, 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 two successive decoding orderings for the relay and six such orderings for the destination, for a total of 12 strategies. There are six more strategies if one source uses the relay and the other does not, and two 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 [31, p. 1883].

E. BRCs

Consider a BRC with four nodes and three messages (see Fig. 4). Several block Markov superposition encoding strategies can be defined, and one of them is depicted in Figs. 11 and 12. Encoding proceeds as follows.

The messages w_0, w_1, w_2 are divided into B blocks w_{0b}, w_{1b}, w_{2b}, respectively, for $b = 1, 2, \ldots, B$. However, before mapping these blocks into codewords, we take note of

المنارة للاستشارات
www.manaraa.com

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. 11. Decode-and-forward for a BRC. The source and relay transmit $\underline{x}_1(\cdot)$ and $\underline{x}_2(\cdot)$, respectively. The $\underline{u}_0(\cdot)$, $\underline{u}_1(\cdot)$, and $\underline{u}_2(\cdot)$ are auxiliary codewords at the source. In block b, the relay decodes w'_{0b}, as do the first and second destinations. The first destination then decodes s_{1b}, and the second decodes s_{2b}.

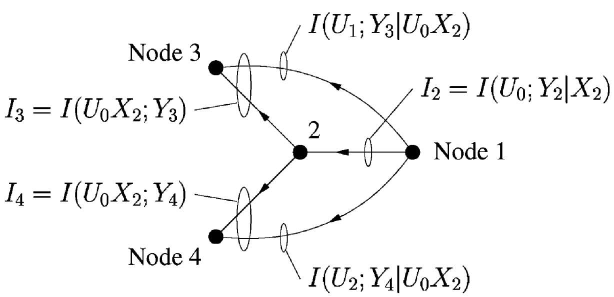

Fig. 12. The information transfer between the nodes of a BRC when using regular encoding/sliding window decoding.

the following result. One sometimes improves rates by letting a "stronger" receiver decode the bits of a "weaker" receiver, and having the former receiver subsequently discard these bits. This type of strategy achieves the capacity of *degraded* broadcast channels, for example [81] (see also [82, Sec. III], [83], and [84, Lemma 1]). In order to enable such an approach, for each b we convert some of the bits of w_{1b} and w_{2b} into bits of w_{0b} that are decoded by *both* destinations. More precisely, we reorganize the w_{0b}, w_{1b}, w_{2b} into blocks $w'_{0b}, w'_{1b}, w'_{2b}$ such that (w'_{0b}, w'_{1b}) includes the bits of (w_{0b}, w_{1b}), and (w'_{0b}, w'_{2b}) includes 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} (see Appendix B for more details).

Decoding proceeds as follows. Node 2 decodes w'_{0b} after block b, but node 3 waits until block $b+1$ to decode both w'_{0b} and s_{1b} by using sliding-window decoding with its channel outputs from blocks b and $b+1$. Similarly, node 4 decodes w'_{0b} and s_{2b} after block $b+1$. The resulting rates are given by the following theorem. The proof of this theorem uses the theory developed in [78, p. 391], [82]–[85].

Theorem 2: The nonnegative rate triples (R_0, R_1, R_2) satisfying

$$\begin{aligned} R_0 &\le \min(I_2, I_3, I_4) \\ R_0 + R_1 &\le \min(I_2, I_3) + I(U_1; Y_3 \,|\, U_0X_2) \\ R_0 + R_2 &\le \min(I_2, I_4) + I(U_2; Y_4 \,|\, U_0X_2) \\ R_0 + R_1 + R_2 &\le \min(I_2, I_3, I_4) + I(U_1; Y_3 \,|\, U_0X_2) \\ &\quad + I(U_2; Y_4 \,|\, U_0X_2) - I(U_1; U_2 \,|\, U_0X_2) \end{aligned} \tag{26}$$

are in the capacity region of the BRC, where

$$\begin{aligned} I_2 &= I(U_0; Y_2 \,|\, X_2) \\ I_3 &= I(U_0X_2; Y_3) \\ I_4 &= I(U_0X_2; Y_4) \end{aligned} \tag{27}$$

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.

Proof: See Appendix B. The information transfer across nodes is shown in Fig. 12, where the edges are labeled by mutual information expressions in (26). The last bound in (26) is due to binning at node 1. □

For example, suppose we choose $U_0 = X_1$ and $U_1 = U_2 = 0$ in Theorem 2. The result is that (R_0, R_1, R_2) is achievable if

$$\begin{aligned} &R_0 + R_1 + R_2 \\ &\quad \le \min[I(X_1; Y_2 \,|\, X_2), I(X_1X_2; Y_3), I(X_1X_2; Y_4)] \end{aligned} \tag{28}$$

for any choice of $p(x_1, x_2)$. We use (28) later in Theorem 10.

Remark 16: The region (26) includes Marton's region [78, p. 391], [82] by viewing the relay as having no input and as being 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 17: 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 new 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 $\underline{u}_0(\tilde{w}'_{0(b-1)}, w'_{0b})$ on each $\underline{v}(\tilde{w}'_{0(b-1)}, \tilde{w}'_{0b})$, and similarly for $\underline{u}_1, \underline{u}_2$, and $\underline{x}_1$. In block b, the cooperation between the source and the relay thus takes place through $\tilde{w}'_{0(b-1)}$ rather than $w'_{0(b-1)}$.

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.

V. Compress-and-Forward

The compress-and-forward strategy of [4, Theorem 6] has the relay forwarding a quantized and compressed version of its channel outputs to the destination. We label the corresponding rates R_{CF}. This approach lets one achieve gains related to multiantenna *reception* [14, p. 64], [29].

A. One Relay

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

$$R_{\text{CF}} = I(X_1; \hat{Y}_2Y_3 \,|\, X_2) \tag{29}$$

where

$$I(\hat{Y}_2; Y_2 \,|\, Y_3X_2) \le I(X_2; Y_3) \tag{30}$$

المنارة للاستشارات
www.manaraa.com

Block 1	Block 2	Block 3	Block 4
$\underline{x}_1(w_1)$	$\underline{x}_1(w_2)$	$\underline{x}_1(w_3)$	$\underline{x}_1(w_4)$
$\underline{u}_2(1)$	$\underline{u}_2(r_{22})$	$\underline{u}_2(r_{23})$	$\underline{u}_2(r_{24})$
$\underline{x}_2(1\|1)$	$\underline{x}_2(s_{22}\|r_{22})$	$\underline{x}_2(s_{23}\|r_{23})$	$\underline{x}_2(s_{24}\|r_{24})$
$\hat{\underline{y}}_2(z_{21})$	$\hat{\underline{y}}_2(z_{22})$	$\hat{\underline{y}}_2(z_{23})$	$\hat{\underline{y}}_2(z_{24})$
$\underline{u}_3(1)$	$\underline{u}_3(r_{32})$	$\underline{u}_3(r_{33})$	$\underline{u}_3(r_{34})$
$\underline{x}_3(1\|1)$	$\underline{x}_3(s_{32}\|r_{32})$	$\underline{x}_3(s_{33}\|r_{33})$	$\underline{x}_3(s_{34}\|r_{34})$
$\hat{\underline{y}}_3(z_{31})$	$\hat{\underline{y}}_3(z_{32})$	$\hat{\underline{y}}_3(z_{33})$	$\hat{\underline{y}}_3(z_{34})$

Fig. 13. Compress-and-forward for two relays. The source, first relay, and second relay transmit $\underline{x}_1(\cdot), \underline{x}_2(\cdot)$, and $\underline{x}_3(\cdot)$, respectively. In block b, the first relay decodes the index r_{3b} carried by $\underline{u}_3(\cdot)$, and quantizes its channel output $\underline{y}_2$ to $\hat{\underline{y}}_2(\cdot)$ by using its knowledge of $\underline{u}_2(\cdot), \underline{u}_3(\cdot), \underline{x}_2(\cdot)$, and the statistical dependence between these vectors and the channel outputs $\underline{y}_3$ and $\underline{y}_4$. This relay then transmits $\hat{\underline{y}}_2(\cdot)$ to the destination via $\underline{x}_2(\cdot)$. The second relay performs similar steps.

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

$$p(x_1)p(x_2)p(\hat{y}_2 \,|\, x_2, y_2)p(y_2, y_3 \,|\, x_1, x_2). \tag{31}$$

Remark 18: Equation (29) illustrates the multiantenna reception behavior: $I(X_1; \hat{Y}_2 Y_3 \,|\, X_2)$ can be interpreted as the rate for a channel with one transmit antenna and two receive antennas. The multiantenna gain is limited by (30), which states that the rate used to compress Y_2 must be smaller than the rate used to transmit data from the relay to the destination. The compression uses techniques developed by Wyner and Ziv [37], i.e., it exploits the destination's side information Y_3.

B. Multiple Relays

For multiple relays, we view the relays-to-destination channel as being a MAC. Note that the signals $Y_2, Y_3, \ldots, Y_{T-1}$ observed by the relays are statistically dependent, so we have the problem of sending "correlated" sources over a MAC as treated in [86]. However, there are two additional features that do not arise in [86]. 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 [28], and we deal with it by using partial decoding.

The compress-and-forward scheme for two relays is depicted in Figs. 13 and 14, and we outline its operation. One chooses random variables X_1, U_t, X_t, and $\hat{Y}_t$ for $t = 2, 3$ that are related as specified later on in (34). During block b, node 2 receives symbols $\underline{y}_{2b}$ that depend on both $\underline{x}_1(w_b)$ and $\underline{x}_3(s_{3b} \,|\, r_{3b})$, where r_{3b} and s_{3b} are indices that are defined in Appendix D. Node 2 decodes r_{3b} and "subtracts" $\underline{u}_3(r_{3b})$ from $\underline{y}_{2b}$. How much node 2 "subtracts" is controlled by how U_3 is related to X_3. For example, if $U_3 = X_3$ then node 2 completely decodes node 3's codewords, while if $U_3 = 0$ then node 2 ignores node 3.

The symbols $\underline{y}_{2b}$, modified by the "subtraction" using $\underline{u}_{3b}(r_{3b})$, are compressed to $\hat{\underline{y}}_2(z_{2b})$ by using the correlation between Y_2, Y_3, and Y_4, i.e., the compression is performed using Wyner–Ziv coding [37]. However, the relays additionally have the problem of encoding in a distributed fashion [28].

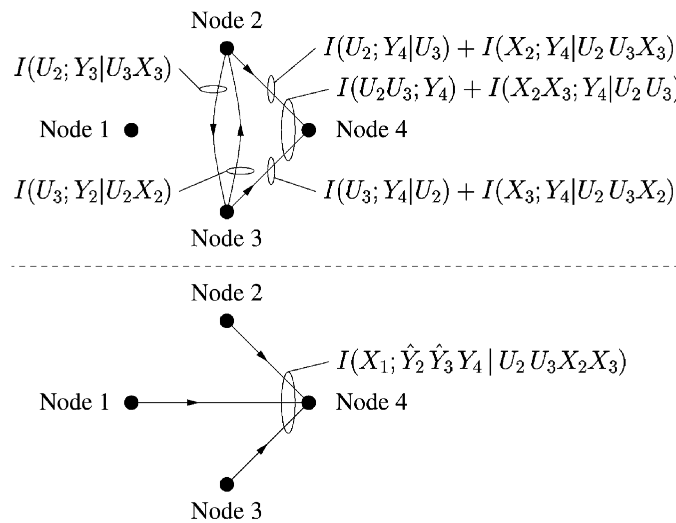

Fig. 14. The information transfer between the nodes for compress-and-forward. Two decoding stages are shown. The first stage (top of the figure) has the relays and destination decoding the indices r_{tb} carried by $\underline{u}_2(\cdot)$ and $\underline{u}_3(\cdot)$, followed by the destination decoding the indices s_{tb} carried by $\underline{x}_2(\cdot)$ and $\underline{x}_3(\cdot)$. The second stage (bottom of the figure) has the destination decoding the messages w_b of $\underline{x}_1(\cdot)$.

After compression, node 2 determines indices $r_{2(b+1)}$ and $s_{2(b+1)}$ from z_{2b} and transmits $\underline{x}_2(s_{2(b+1)} \,|\, r_{2(b+1)})$ in block $b+1$. Node 3 operates in a similar fashion as node 2. The destination uses $\underline{y}_{4b}$ to decode r_{2b}, r_{3b}, s_{2b}, and s_{3b} to obtain $z_{2(b-1)}$ and $z_{3(b-1)}$. Finally, the destination uses $\underline{y}_{4(b-1)}$ together with $\hat{\underline{y}}_2(z_{2(b-1)})$ and $\hat{\underline{y}}_3(z_{3(b-1)})$ to decode w_{b-1}.

We derive the rates of such schemes, and summarize the result with the following theorem. Let $\mathcal{S}^C$ be the complement of $\mathcal{S}$ in $\mathcal{T} = \{2, 3, \ldots, T{-}1\}$, and let $\mathcal{S}_1 \backslash \mathcal{S}_2$ denote $\{s : s \in \mathcal{S}_1, s \notin \mathcal{S}_2\}$.

Theorem 3: Compress-and-forward achieves any rate up to

$$R_{\text{CF}} = I(X_1; \hat{Y}_{\mathcal{T}} Y_T \,|\, U_{\mathcal{T}} X_{\mathcal{T}}) \tag{32}$$

where

$$\begin{aligned} I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} \,|\, U_{\mathcal{T}} X_{\mathcal{T}} \hat{Y}_{\mathcal{S}^C} Y_T) + \sum_{t \in \mathcal{S}} I\left(\hat{Y}_t; X_{\mathcal{T} \backslash \{t\}} \middle| U_{\mathcal{T}} X_t\right) \\ \leq I(X_{\mathcal{S}}; Y_T \,|\, U_{\mathcal{T}} X_{\mathcal{S}^C}) + \sum_{m=1}^{M} I\left(U_{\mathcal{K}_m}; Y_{r(m)} \middle| U_{\mathcal{K}_m^C} X_{r(m)}\right) \end{aligned} \tag{33}$$

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 $r(m) \notin \mathcal{K}_m$ (recall that $\mathcal{S}^C$ and $\mathcal{K}_m^C$ are the complements of the respective $\mathcal{S}$ and $\mathcal{K}_m$ in $\mathcal{T}$). For $r(m) = T$ we set $X_T = 0$. Furthermore, the joint probability distribution of the random variables factors as

$$p(x_1) \left[\prod_{t=2}^{T-1} p(u_t, x_t) p(\hat{y}_t \,|\, u_t, x_t, y_t) \right] \cdot p(y_2, \ldots, y_T \,|\, x_1, \ldots, x_{T-1}). \tag{34}$$

Proof: See Appendix D. □

Remark 19: The left-hand side of (33) has rates that generalize results of [37] to distributed source coding with side information at the destination (see [38], [39]). The right-hand side of (33) describes rates achievable on the multiway channel between the relays and the destination. In other words, an

المنارة للاستشارات
www.manaraa.com

(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 [87, Ch. 1]). This last remark also applies to (29)–(31).

Remark 20: Equation (34) specifies that the $X_t, t = 1,\dots,T-1$, are statistically independent. Equation (32) illustrates the multiantenna reception behavior: the mutual information expression can be interpreted as the rate for a channel with one transmit antenna and $T-1$ receive antennas. The multiantenna gain is limited by (33), which is a combination of source and channel coding bounds.

Remark 21: For $T=3$ we recover (29)–(30).

Remark 22: For $T=4$, there are nine bounds of the form (33): 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-and-forward rate for large T is tedious.

Remark 23: 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 (33) simplifies to

$$I(\hat{Y}_{\mathcal{S}};Y_{\mathcal{S}}\,|\,X_{\mathcal{T}}\hat{Y}_{\mathcal{S}^C}Y_T) \le \sum_{m=1}^{M} I\left(X_{\mathcal{K}_m};Y_{r(m)}\left|X_{\mathcal{K}_m^C}\right.\right). \quad (35)$$

However, for $T=4$ there are still nine bounds, namely

$$\begin{aligned}
&I(\hat{Y}_2;Y_2\,|\,X_2X_3\hat{Y}_3Y_4)\\
&\quad\le \min\{I(X_2;Y_3\,|\,X_3), I(X_2;Y_4\,|\,X_3)\}\\
&I(\hat{Y}_3;Y_3\,|\,X_2X_3\hat{Y}_2Y_4)\\
&\quad\le \min\{I(X_3;Y_2\,|\,X_2), I(X_3;Y_4\,|\,X_2)\}\\
&I(\hat{Y}_2\hat{Y}_3;Y_2Y_3\,|\,X_2X_3Y_4)\\
&\quad\le \min\{I(X_2X_3;Y_4), I(X_2;Y_3\,|\,X_3)+I(X_3;Y_2\,|\,X_2),\\
&\qquad\quad I(X_2;Y_3\,|\,X_3)+I(X_3;Y_4\,|\,X_2),\\
&\qquad\quad I(X_2;Y_4\,|\,X_3)+I(X_3;Y_2\,|\,X_2),\\
&\qquad\quad I(X_2;Y_4\,|\,X_3)+I(X_3;Y_4\,|\,X_2)\}. \qquad (36)
\end{aligned}$$

Remark 24: Suppose the relays perform no partial decoding, i.e., $U_t=0$ for all t. The bounds (33) simplify because we need not partition $\mathcal{S}$. For example, for $T=4$ we have

$$\begin{aligned}
&I(\hat{Y}_2;Y_2\,|\,X_2X_3\hat{Y}_3Y_4)+I(\hat{Y}_2;X_3\,|\,X_2)\le I(X_2;Y_4\,|\,X_3)\\
&I(\hat{Y}_3;Y_3\,|\,X_2X_3\hat{Y}_2Y_4)+I(\hat{Y}_3;X_2\,|\,X_3)\le I(X_3;Y_4\,|\,X_2)\\
&I(\hat{Y}_2\hat{Y}_3;Y_2Y_3\,|\,X_2X_3Y_4)+I(\hat{Y}_2;X_3\,|\,X_2)+I(\hat{Y}_3;X_2\,|\,X_3)\\
&\quad\le I(X_2X_3;Y_4). \qquad (37)
\end{aligned}$$

For Schein's parallel relay network, this region reduces to Theorem 3.3.2 in [28, p. 136].

VI. Mixed Strategies

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

$$\mathcal{T}_1=\{2,3,\dots,T_1+1\},\quad \mathcal{T}_2=\{T_1+2,\dots,T-1\}. \quad (38)$$

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

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

$$\begin{aligned}
R_{\text{DCF}} = \min\Bigg\{ &\min_{1\le t\le T_1} I\left(X_{\pi(1:t)};Y_{\pi(t+1)}\left|X_{\pi(t+1:T_1+1)}\right.\right),\\
&\qquad I\left(X_1X_{\mathcal{T}_1};\hat{Y}_{\mathcal{T}_2}Y_T\left|U_{\mathcal{T}_2}X_{\mathcal{T}_2}\right.\right)\Bigg\} \qquad (39)
\end{aligned}$$

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

$$\begin{aligned}
&I\left(\hat{Y}_{\mathcal{S}};Y_{\mathcal{S}}\left|U_{\mathcal{T}_2}X_{\mathcal{T}_2}\hat{Y}_{\mathcal{S}^C}Y_T\right.\right)+\sum_{t\in\mathcal{S}} I\left(\hat{Y}_t;X_{\mathcal{T}_2\setminus\{t\}}\left|U_{\mathcal{T}_2}X_t\right.\right)\\
&\quad\le I\left(X_{\mathcal{S}};Y_T\,|U_{\mathcal{S}}X_{\mathcal{S}^C}\right)+\sum_{m=1}^{M} I\left(U_{\mathcal{K}_m};Y_{r(m)}\,|U_{\mathcal{K}_m^C}X_{r(m)}\right)
\end{aligned} \quad (40)$$

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

$$\begin{aligned}
p(x_1,x_{\mathcal{T}_1})\cdot&\left[\prod_{t\in\mathcal{T}_2} p(u_t,x_t)p\left(\hat{y}_t\,|u_{\mathcal{T}_2},x_t,y_t\right)\right]\\
&\cdot p(y_2,\dots,y_T\,|\,x_1,\dots,x_{T-1}). \qquad (41)
\end{aligned}$$

Proof: We omit the proof because of its similarity to the proofs in [32], [33] and Appendix D. Summarizing the idea of the proof, nodes 1 to T_1+1 operate as if they were using decode-and-forward for a (T_1+2)-node relay network. Similarly, nodes T_1+2 to $T-1$ operate as if they were using compress-and-forward for a $(T-T_1)$-node relay network. Instead of proving Theorem 4, we supply a proof of Theorem 5 below. Theorem 5 illustrates how to improve Theorem 4 for $T=4$ by permitting partial decoding at one of the relays. □

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 (cf. [29, Theorem 2])

$$R_{\text{DCF}}=\min\{I(X_1;Y_2\,|\,X_2), I(X_1X_2;\hat{Y}_3Y_4\,|\,X_3)\} \quad (42)$$

where

$$I(\hat{Y}_3;Y_3\,|\,X_3Y_4)\le I(X_3;Y_4) \quad (43)$$

and the joint probability distribution of the random variables factors as

$$p(x_1,x_2)p(x_3)p(\hat{y}_3\,|\,x_3,y_3)p(y_2,y_3,y_4\,|\,x_1,x_2,x_3). \quad (44)$$

www.manaraa.com

The second mutual information expression in (42) can be interpreted as the rate for a 2×2 multiantenna system. Hence, when node 2 is near the source and node 3 is near the destination, one will achieve rates close to those described in [88], [89].

A. Two Relays and Partial Decoding

Suppose there are two relays, node 2 uses decode-and-forward, and node 3 uses compress-and-forward. Suppose node 3 further partially decodes the signal from node 2 before compressing its observation. However, for simplicity we make X_3 statistically independent of X_1 and X_2. We show that this strategy achieves the following rates.

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

$$R_{\rm DCF} = \min\{I(X_1;Y_2\,|\,U_2X_2),\; I(X_1X_2;\hat{Y}_3Y_4\,|\,U_2X_3) + R_2'\} \quad (45)$$

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

$$\begin{aligned} &R_3 \geq I(\hat{Y}_3;Y_3\,|\,U_2X_3Y_4)\\ &0 \leq R_2' \leq \min\{I(U_2;Y_3\,|\,X_3), I(U_2;Y_4\,|\,X_3)\}\\ &0 \leq R_3 \leq I(X_3;Y_4\,|\,U_2)\\ &R_2' + R_3 \leq I(U_2X_3;Y_4) \end{aligned} \quad (46)$$

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

$$p(x_1,u_2,x_2)p(x_3)p(\hat{y}_3\,|\,u_2,x_3,y_3)p(y_2,y_3,y_4\,|\,x_1,x_2,x_3). \quad (47)$$

Proof: See Appendix E. □

Remark 25: We recover (42) and (43) with $U_2 = 0$.

Remark 26: We recover (12) by turning off Relay 3 with $U_2 = X_3 = \hat{Y}_3 = 0$ and $R_2' = R_3 = 0$.

Remark 27: We recover (29)–(30) by turning off and ignoring Relay 2 with $U_2 = X_2 = 0$ and $Y_2 = X_1$.

VII. 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

$$\underline{Y}_t = \underline{Z}_t + \sum_{s \neq t} \frac{H_{st}}{\sqrt{d_{st}^{\alpha}}} \underline{X}_s \quad (48)$$

where d_{st} is the distance between nodes s and t, α is an attenuation exponent, $\underline{X}_s$ is an $n_s \times 1$ complex vector, H_{st} is an $n_t \times n_s$ matrix whose i,j entry $H_{st}^{(i,j)}$ is a complex fading random variable, and $\underline{Z}_t$ is an $n_t \times 1$ vector whose entries are independent and identically distributed (i.i.d.), proper [90], complex, Gaussian random variables with zero mean and unit variance. 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: $H_{st}^{(i,j)}$ is a constant for all s, t, i, and j.
- Phase fading: $H_{st}^{(i,j)} = e^{j\theta_{st}^{(i,j)}}$, 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 $\underline{X}_t$ and $\underline{Z}_t$.
- Rayleigh fading: $H_{st}^{(i,j)}$ is a proper, complex, Gaussian random variable with zero mean and unit variance. The $H_{st}^{(i,j)}$ are jointly independent of each other and all $\underline{X}_t$ and $\underline{Z}_t$.
- Single-bounce fading: $H_{st} = B_{st}D_{st}C_{st}$, 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 have independent 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 $\underline{X}_t$ and $\underline{Z}_t$. The matrices B_{st} and C_{st} might represent knowledge about the directions of arrival and departure, respectively, of plane waves [91].
- Fading with directions: $H_{st} = B_{st}G_{st}C_{st}$, where B_{st} is a random $n_t \times n_{st}^{(2)}$ matrix, G_{st} is an $n_{st}^{(2)} \times n_{st}^{(1)}$ random matrix whose entries have independent phases that are uniformly distributed over $[0, 2\pi)$, 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 $\underline{X}_t$ and $\underline{Z}_t$.

For the no-fading case, the H_{st} are known by all nodes. For the other cases, we assume that node t knows only its *own* fading coefficients. That is, node t knows H_{st} for all s, but it does not know $H_{st'}$ for $t' \neq t$. One can model this as in [89] and write the full channel output of node t as

$$\underline{Y}_t' = \left[\underline{Y}_t, H_{1t}, H_{2t}, \ldots, H_{(T-1)t}\right]. \quad (49)$$

Remark 28: The above model has *continuous* random variables, and one should therefore exercise caution when applying the theorems of Sections IV to VI (cf. [77, Sec. 2.5]). We will assume that the theorems are valid for Gaussian random variables. This poses no difficulty for the decode-and-forward theorems that can be derived by using entropy-typical sequences [25, p. 51], [78, p. 40]. However, the compress-and-forward scheme is trickier to deal with (see Remark 30).

Remark 29: The above model applies to *full-duplex* relays, i.e., relays that can transmit and receive at the same time and in the same frequency band. This might be possible, e.g., if each relay has two antennas: one receiving and one transmitting. For half-duplex devices, one should modify (48) by adding the constraints that $\underline{Y}_t = \underline{0}$ if $\underline{X}_t \neq \underline{0}$ for all t. The analysis is then similar to that described below if one assumes that every node knows at all times which mode (listen or talk) the other nodes are using (cf. [53], [55]). However, if the nodes do not know each others operating modes, then one might adopt an approach such as that described in [75].

A. Optimizing the Cut-Set and Decode-and-Forward Rates

Let $\underline{X}_{\mathcal{S}} = \{\underline{X}_t : t \in \mathcal{S}\}$ and $H_{\mathcal{S}\mathcal{U}} = \{H_{st} : s \in \mathcal{S}, t \in \mathcal{U}\}$. We write $\mathcal{S}^C$ for the complement of $\mathcal{S}$ in the set of transmitting nodes $\mathcal{T} = \{1, 2, \ldots, T-1\}$. We further write $h(A)$ and $h(A\,|\,B)$ for the differential entropies of A without and with conditioning on B, respectively [25, Ch. 9]. We prove the following proposition.

www.manaraa.com

Proposition 2: Consider the channels (48) and (49) where every node t knows only its own channel coefficients $H_{\mathcal{T}t}$. The maxima in the cut-set bound (5) and the decode-and-forward rate (21) are attained by Gaussian $\underline{X}_{\mathcal{T}}$.

Proof: All of the mutual information expressions in (5) and (21) can be written as

$$\begin{aligned} I\left(\underline{X}_{\mathcal{S}};\underline{Y}'_{\mathcal{U}}\,|\underline{X}_{\mathcal{S}^C}\right) &= I\left(\underline{X}_{\mathcal{S}};\underline{Y}_{\mathcal{U}}H_{\mathcal{T}\mathcal{U}}\,|\underline{X}_{\mathcal{S}^C}\right)\\ &= I\left(\underline{X}_{\mathcal{S}};\underline{Y}_{\mathcal{U}}\,|\underline{X}_{\mathcal{S}^C}H_{\mathcal{T}\mathcal{U}}\right)\\ &= h\left(\underline{Y}_{\mathcal{U}}\,|\underline{X}_{\mathcal{S}^C}H_{\mathcal{T}\mathcal{U}}\right) - h(\underline{Z}_{\mathcal{U}}) \end{aligned} \tag{50}$$

for some $\mathcal{S}$ and $\mathcal{U}$. The first step in (50) follows by (49), the second because the H_{st} are independent of the $\underline{X}_t$, and the third by (48). Observe that $h(\underline{Z}_{\mathcal{U}})$ cannot be influenced by $\underline{X}_{\mathcal{T}}$, so we are left with a *maximum entropy* problem.

The best $\underline{X}_{\mathcal{T}}$ has zero mean because every node t uses less power by sending $\underline{X}_t - \mathrm{E}[\underline{X}_t]$ rather than $\underline{X}_t$, and this change does not affect the mutual information expressions (50). Suppose the best $\underline{X}_{\mathcal{T}}$ has covariance matrix $Q_{\underline{X}_{\mathcal{T}}}$ for either (5) or (21). This choice of $Q_{\underline{X}_{\mathcal{T}}}$ fixes $Q_{\underline{Y}_{\mathcal{U}}\underline{X}_{\mathcal{S}^C}\,|\,H_{\mathcal{T}\mathcal{U}}=h}$ for all $\mathcal{U},\mathcal{S}$, and h, where $Q_{\underline{A}\,\underline{B}\,|\,C=c}$ is the covariance matrix of the vector $[\underline{A}^T\,\underline{B}^T]$ conditioned on $C=c$. But if $Q_{\underline{A}\,\underline{B}}$ is fixed, then one maximizes $h(\underline{A}\,|\,\underline{B})$ by making $\underline{A}$ and $\underline{B}$ jointly Gaussian (see [92, Lemma 1]). Hence, choosing $\underline{X}_{\mathcal{T}}$ to be Gaussian with covariance matrix $Q_{\underline{X}_{\mathcal{T}}}$ maximizes $h(\underline{Y}_{\mathcal{U}}\,|\,\underline{X}_{\mathcal{S}^C}, H_{\mathcal{T}\mathcal{U}}=h)$ for every h and *every* mutual information expression in (5) and (21). The theorem is proved by averaging over h. □

Choosing Gaussian $\underline{X}_{\mathcal{T}}$, we find that (50) simplifies to

$$I\left(\underline{X}_{\mathcal{S}};\underline{Y}'_{\mathcal{U}}\,|\underline{X}_{\mathcal{S}^C}\right) = \int_h p(h)\log\left(\left|Q_{\underline{Y}_{\mathcal{U}}\,|\,\underline{X}_{\mathcal{S}^C},H_{\mathcal{T}\mathcal{U}}=h}\right|\right)dh \tag{51}$$

where $|Q|$ is the determinant of Q. The final step is to optimize $Q_{\underline{X}_{\mathcal{T}}}$. This optimization depends on the type of fading, so we perform it case-by-case.

B. No Fading and One Relay

Suppose we have one relay, nodes with one antenna, and no fading. Let

$$\rho = \frac{\mathrm{E}[(X_1-\mathrm{E}[X_1])(X_2-\mathrm{E}[X_2])^*]}{\sqrt{\mathrm{E}[|X_1-\mathrm{E}[X_1]|^2]\mathrm{E}[|X_2-\mathrm{E}[X_2]|^2]}} \tag{52}$$

be the correlation coefficient of X_1 and X_2, where X_2^* is the complex conjugate of X_2. Using (51), the cut-set bound (6) is

$$\begin{aligned} C \le \max_{0\le\rho\le 1}\min\Bigg\{&\log\left(1+P_1\left(\frac{1}{d_{12}^{\alpha}}+\frac{1}{d_{13}^{\alpha}}\right)(1-|\rho|^2)\right),\\ &\log\left(1+\frac{P_1}{d_{13}^{\alpha}}+\frac{P_2}{d_{23}^{\alpha}}+\frac{2\rho\sqrt{P_1P_2}}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}}\right)\Bigg\} \end{aligned} \tag{53}$$

where ρ is real. Similarly, the best decode-and-forward rate (12) is

$$\begin{aligned} R_{\mathrm{DF}} = \max_{0\le\rho\le 1}\min\Bigg\{&\log\left(1+\frac{P_1}{d_{12}^{\alpha}}(1-|\rho|^2)\right),\\ &\log\left(1+\frac{P_1}{d_{13}^{\alpha}}+\frac{P_2}{d_{23}^{\alpha}}+\frac{2\rho\sqrt{P_1P_2}}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}}\right)\Bigg\}. \end{aligned} \tag{54}$$

Consider next the compress-and-forward strategy. Gaussian input distributions are not necessarily optimal, but for simplicity

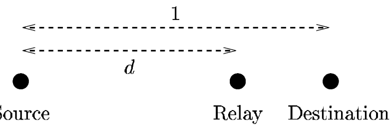

Fig. 15. A single relay on a line.

we choose X_1 and X_2 to be Gaussian. We further choose $\hat{Y}_2 = Y_2 + \hat{Z}_2$, where $\hat{Z}_2$ is a proper, complex, Gaussian, random variable with zero mean and variance $\hat{N}_2$, and $\hat{Z}_2$ is independent of all other random variables. The rate (29) is then

$$R_{\mathrm{CF}} = \log\left(1+\frac{P_1}{d_{12}^{\alpha}(1+\hat{N}_2)}+\frac{P_1}{d_{13}^{\alpha}}\right) \tag{55}$$

where the choice

$$\hat{N}_2 = \frac{P_1\left(1/d_{12}^{\alpha}+1/d_{13}^{\alpha}\right)+1}{P_2/d_{23}^{\alpha}} \tag{56}$$

satisfies (30) with equality (see also [29] and [60, Sec. 3.2] for the same analysis).

Remark 30: The proof of the achievability of (29) and its generalization (32) is presented in Appendix D. The proof requires strong typicality to invoke the *Markov lemma* of [93, Lemma 4.1] (cf. [78, p. 370, eq. (4.38)]). However, strong typicality does not apply to continuous random variables, and hence the achievability of (29) might not imply the achievability of (55). However, for Gaussian input distributions, one can generalize the Markov lemma along the lines of [94], [95] and thereby show that (55) is achievable.

As an example, suppose the source, relay, and destination are aligned as in Fig. 15, where $d_{12}=|d|$, $d_{23}=|1-d|$, and $d_{13}=1$. Fig. 16 plots various bounds for $P_1=P_2=10$ and $\alpha=2$. The curves labeled DF and CF give the respective decode-and-forward and compress-and-forward rates. Also shown are the rates when the relay is turned off, but now only half the overall 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 $E[|X_{2i}|^2]\le P_2$. This strategy is called *amplify-and-forward* in [14, p. 80] (see also [27], [28, p. 61], and [36], [50]). Amplify-and-forward turns the source-to-destination channel into a unit-memory intersymbol interference channel. The curve labeled AF shows the capacity of this channel after optimizing c.

Remark 31: As the relay moves toward the source ($d\to 0$), the rates (54) and (55) become

$$\begin{aligned} R_{\mathrm{DF}} &= \log(1+P_1+P_2+2\sqrt{P_1P_2})\\ R_{\mathrm{CF}} &= \log(1+P_1+P_2) \end{aligned} \tag{57}$$

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

$$\begin{aligned} R_{\mathrm{DF}} &= \log(1+P_1)\\ R_{\mathrm{CF}} &= \log(1+2P_1) \end{aligned} \tag{58}$$

and R_{CF} is the capacity. These limiting capacity results extend to multiple relays, as discussed next. Note that the right-hand side of (53) is the same as R_{DF} and R_{CF} only if $d=0$ and $d=1$, respectively.

www.manaraa.com

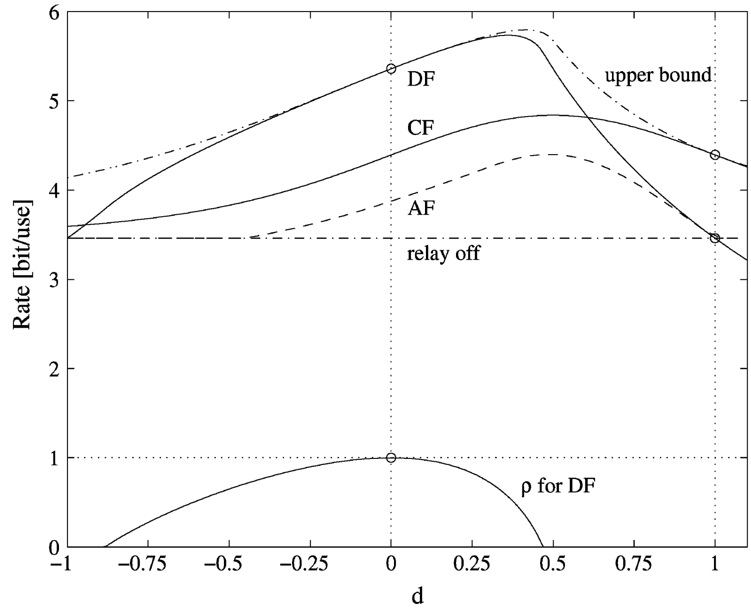

Fig. 16. Rates for one relay with $P_1 = P_2 = 10$ and $\alpha = 2$.

Remark 32: Traditional multihopping has the source transmitting the message W to the relay in one time slot, and then the relay forwarding W to the destination in a second time slot. This scheme can be used with both full-duplex and half-duplex relays. Suppose one assigns a fraction β of the time to the first hop. One then achieves the rate

$$R_{\text{MH}} = \min\left[\beta \log\left(1 + \frac{P_1}{\beta d_{12}^{\alpha}}\right), (1-\beta)\log\left(1 + \frac{P_2}{(1-\beta) d_{23}^{\alpha}}\right)\right]. \quad (59)$$

However, even after optimizing β one always performs worse than using no relay for any d in Fig. 16. This happens because $\alpha = 2$ is here too small to make multihopping useful.

C. No Fading and Many Relays

Suppose there are $T-2$ relays, all nodes have one antenna each, and there is no fading. Suppose further that the relays are within a distance d of the source. The decode-and-forward rate of Theorem 1 becomes the capacity as $d \to 0$, which is

$$R_{\text{DF}} = \log\left(1 + \left[\sum_{t=1}^{T-1} \sqrt{P_t}\right]^2\right). \quad (60)$$

This limiting capacity result is called an *antenna clustering capacity* in [29].

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

$$R_{\text{CF}} = \log(1 + (T-1)P_1). \quad (61)$$

This limiting capacity result is another type of antenna clustering capacity (see [29]).

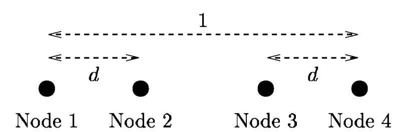

Fig. 17. Two relays on a line.

Finally, consider the geometry in Fig. 17. The mixed strategy of Theorem 4 achieves capacity as $d \to 0$, which is

$$R_{\text{DCF}} = \log\left(1 + (T - T_1 - 1)\left[\sum_{t=1}^{T_1+1} \sqrt{P_t}\right]^2\right) \quad (62)$$

with $T = 4$ and $T_1 = 1$.

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

D. Phase Fading and One Relay

Consider phase fading where θ_{st} is known only to node t for all s. We prove the following geometric capacity result.

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

$$P_1/d_{13}^{\alpha} + P_2/d_{23}^{\alpha} \le P_1/d_{12}^{\alpha} \quad (63)$$

then the capacity is

$$C = \log(1 + P_1/d_{13}^{\alpha} + P_2/d_{23}^{\alpha}). \quad (64)$$

Proof: Using (51), R_{DF} in (12) is

$$\max_{\rho} \min\left\{\log\left(1 + \frac{P_1}{d_{12}^{\alpha}}(1-|\rho|^2)\right), \int_0^{2\pi} \frac{d\phi_{13} d\phi_{23}}{(2\pi)^2} \log\left(1 + \frac{P_1}{d_{13}^{\alpha}} + \frac{P_2}{d_{23}^{\alpha}} + \frac{2\Re(\rho\, e^{j(\phi_{13}-\phi_{23})})\sqrt{P_1 P_2}}{d_{13}^{\alpha/2} d_{23}^{\alpha/2}}\right)\right\} \quad (65)$$

www.manaraa.com

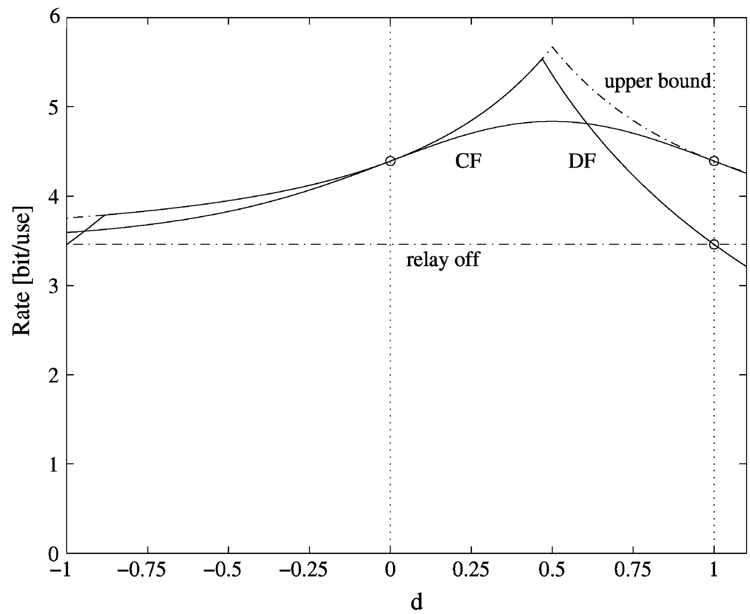

Fig. 18. Rates for one relay with phase fading, $P_1 = P_2 = 10$, and $\alpha = 2$.

where $\Re(x)$ is the real part of x. Let $I(\rho)$ be the integral in (65), and note that $I(\rho) = \mathrm{E}[\log(X)]$ for a random variable X with

$$\mathrm{E}[X] = 1 + P_1/d_{13}^{\alpha} + P_2/d_{23}^{\alpha}. \tag{66}$$

Jensen's inequality [25, p. 25] thus gives $I(\rho) \leq I(0)$, and this implies that $\rho = 0$ maximizes (65). The same arguments show that $\rho = 0$ is also best for the cut-set bound (5) (see Section VII-F for a more general proof of this claim). Summarizing, the capacity upper and lower bounds are (53) and (54), respectively, both with $\rho = 0$. The bounds meet if (63) is satisfied. □

Remark 33: Phase-fading relay channels are not degraded in the sense of [4].

Remark 34: The optimality of $\rho = 0$ for phase fading was independently stated in [60, Lemma 1], [63, Sec. 2.3]. The fact that decode-and-forward can achieve capacity with distributed nodes first appeared in [34].

Remark 35: The capacity (64) is the rate of a distributed antenna array with full cooperation even though the transmitting antennas are not colocated.

Remark 36: Theorem 6 generalizes to Rayleigh fading, and to other fading processes where the phase varies uniformly over time and $[0, 2\pi)$ (see Theorem 8).

Remark 37: Theorem 6 (and Theorems 7–10 below) remains valid for various practical extensions of our models. For example, suppose that the source and relays can share extra information due to their proximity. Nevertheless, if the destination is far away and there is phase uncertainty in the $H_{sT}, s = 1, 2, \ldots, T-1$, then the capacity for this problem is the same as when no extra information can be shared.

Remark 38: The capacity claim of Theorem 6 (and Theorems 7–10) is not valid for half-duplex relays even if each node is aware of the operating modes of all other nodes. The reason is that one must introduce a time-sharing parameter, and this parameter takes on different values for the capacity lower and upper bounds. However, one can derive capacity theorems if the time-sharing parameter is restricted to certain ranges [75]. For instance, this occurs if protocols restrict time to be shared equally between transmission and reception.

The condition (63) is satisfied for a range of d_{12} near zero. For example, for the geometry of Fig. 15 with $\alpha = 2$ and $P_1 = P_2$, the bound (63) is $-0.883 \leq d \leq 0.469$. Fig. 18 plots the resulting cut-set and decode-and-forward rates for $P_1 = P_2 = 10$ and a range of d. Fig. 18 also plots the rates of compress-and-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 (55) and (56), i.e., they are the same as for the no-fading case. We remark that compress-and-forward performs well for all d and even achieves capacity for $d = 0$ in addition to $d = 1$.

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 (63) is

$$1/d_{13}^{\alpha} + 1/d_{23}^{\alpha} \leq 1/d_{12}^{\alpha}\,. \tag{67}$$

The relay positions that satisfy (67) for $\alpha = 2, \alpha = 4$, and $\alpha = 100$ are drawn as the shaded regions in Fig. 19. 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 to the source.

E. Phase Fading and Many Relays

Suppose there are T nodes subject to phase fading. We have the following generalization of Theorem 6.

Theorem 7: Decode-and-forward achieves capacity with $T-2$ relays and phase fading if

$$\sum_{t=1}^{T-1} \frac{P_t}{d_{tT}^{\alpha}} \leq \max_{\pi(\cdot)} \min_{1 \leq s \leq T-2} \sum_{t \in \pi(1:s)} \frac{P_t}{d_{t\pi(s+1)}^{\alpha}} \tag{68}$$

www.manaraa.com

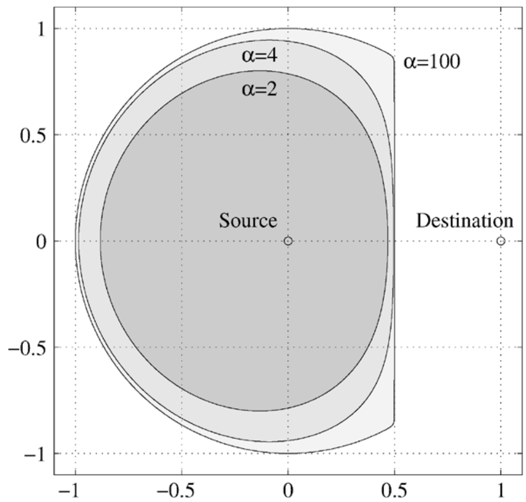

Fig. 19. Positions of the relay where decode-and-forward achieves capacity with phase fading and $P_1 = P_2$.

and the resulting capacity is

$$C = \log\left(1 + \sum_{t=1}^{T-1} \frac{P_t}{d_{tT}^{\alpha}}\right). \tag{69}$$

Note that the minimization in (68) does not include $s = T-1$.

Proof: The optimal input covariance matrix $Q_{\underline{X}_T}$ is diagonal for both the cut-set and decode-and-forward versions of (51). We defer a proof of this claim to the proof of Theorem 8. Using (51), the resulting decode-and-forward rates (21) are

$$I\left(X_{\pi(1:s)}; Y_{\pi(s+1)} \,\middle|\, X_{\pi(s+1:T-1)}\right) = \log\left(1 + \sum_{t \in \pi(1:s)} \frac{P_t}{d_{t\pi(s+1)}^{\alpha}}\right) \tag{70}$$

for $s = 1, 2, \ldots, T-1$. But the rate (70) with $s = T-1$ is also in the cut-set bound (5), and (68) ensures that every other mutual information expression in the the cut-set bound is larger than one of the rates of (70). The condition (68) thus ensures that both the cut-set and decode-and-forward rates are (69). □

The bound (68) is satisfied if all relays are near the source. For example, for $T = 4$ the expression (68) is

$$\frac{P_1}{d_{14}^{\alpha}} + \frac{P_2}{d_{24}^{\alpha}} + \frac{P_3}{d_{34}^{\alpha}} \le \max\left[\min\left(\frac{P_1}{d_{12}^{\alpha}}, \frac{P_1}{d_{13}^{\alpha}} + \frac{P_2}{d_{23}^{\alpha}}\right), \min\left(\frac{P_1}{d_{13}^{\alpha}}, \frac{P_1}{d_{12}^{\alpha}} + \frac{P_3}{d_{23}^{\alpha}}\right)\right]. \tag{71}$$

The bound (71) is satisfied for fixed P_1, P_2, and P_3 if both d_{12} and d_{13} are small as compared to $d_{t4}, t = 1, 2, 3$.

As a second 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 $d_{tT} \ge 1 - d$. Suppose further that $P_t = P$ for all t. The left-hand side of (68) is then at most $(T-1)P/(1-d)^{\alpha}$, while the right-hand side of (68) is at least P/d^{α} (cf. (71)). The bound (68) is thus satisfied if

$$d \le \frac{1}{(T-1)^{1/\alpha} + 1}. \tag{72}$$

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

As a third geometric example, consider a linear network as in Fig. 15 but with $T-2$ relays placed regularly to the right of the source at positions $d_{1t} = (t-1)d, 2 \le t \le T-1$, where $0 \le d < 1/(T-1)$ (see also [32, Fig. 9]). Suppose again that $P_t = P$ for all t, and note that the right-hand side of (68) is P/d^{α} (cf. (71)). The bound (68) is thus satisfied if

$$\sum_{t=1}^{T-1} \frac{P}{[1-(t-1)d]^{\alpha}} \le \frac{P}{d^{\alpha}}. \tag{73}$$

Suppose we choose $d = 1/(T-1+\epsilon)$ where ϵ is a positive constant independent of T. Multiplying both sides of (73) by d^{α}/P, we obtain

$$\sum_{t=1}^{T-1} \frac{1}{(t+\epsilon)^{\alpha}} \le 1. \tag{74}$$

The expression (74) is clearly not satisfied if $\epsilon = 0$, but the left-hand side of (74) is continuous and decreasing in ϵ. We upper-bound the sum in (74) by the integral

$$\int_0^{\infty} \frac{1}{(t+\epsilon)^{\alpha}}\, dt = \frac{1}{(\alpha-1)\epsilon^{\alpha-1}}. \tag{75}$$

The bound (74) is thus satisfied if $\alpha > 1$ and

$$\epsilon \ge \left(\frac{1}{\alpha-1}\right)^{\frac{1}{\alpha-1}}. \tag{76}$$

But this implies there is an ϵ between 0 and the right-hand side of (76) such that (73) holds with equality. This choice for ϵ makes (69) become

$$C = \log(1 + P/d^{\alpha}) \underset{\text{large } T}{\approx} \alpha \log(T). \tag{77}$$

In other words, the capacity grows logarithmically in the number of nodes (or relays). Other types of logarithmic scaling laws were obtained in [32] and [51].

F. Phase Fading, Many Relays, and Multiple Antennas

Suppose we have phase or Rayleigh fading, many relays, and multiple antennas. We prove the following general result.

Theorem 8: The best $\underline{X}_t$ for both the cut-set bound (5) and the decode-and-forward rate (21) are independent and Gaussian if there is phase or Rayleigh fading with multiple antennas. The best covariance matrices are $Q_{\underline{X}_t} = \sqrt{P_t/n_t} \cdot I$. Moreover, decode-and-forward achieves capacity if R_{DF} in (21) is $I(X_1 X_2 \ldots X_{T-1}; Y_T)$. The resulting capacity is

$$C = \int_{h_{\mathcal{T}T}} p(h_{\mathcal{T}T}) \log\left(\left| I + \sum_{t=1}^{T-1} \frac{P_t}{n_t} \frac{h_{tT} h_{tT}^{\dagger}}{d_{tT}^{\alpha}} \right|\right) dh_{\mathcal{T}T} \tag{78}$$

where the integral is over all channel matrices

$$h_{\mathcal{T}T} = \{h_{sT} : s = 1, 2, \ldots, T-1\}.$$

Proof: Using (51), we write the mutual information expression in (5) as

$$\int_h p(h) \log\left(\left| Q_{\underline{Y}_{\mathcal{S}^C} \underline{Y}_T | \underline{X}_{\mathcal{S}^C}, H_{\mathcal{T}\mathcal{U}}=h} \right|\right) dh \tag{79}$$

www.manaraa.com

where $\mathcal{U} = \mathcal{S}^C \cup \{T\}$. For $H_{st} = h_{st}$, we define

$$\underline{\tilde{Y}}_t = \underline{Z}_t + \sum_{s \in \{1\} \cup \mathcal{S}} \frac{h_{st}}{\sqrt{d_{st}^\alpha}} \underline{X}_s. \quad (80)$$

The determinant in (79) evaluates to

$$\left| Q_{\underline{\tilde{Y}}_{\mathcal{S}^C} \underline{\tilde{Y}}_T \,|\, \underline{X}_{\mathcal{S}^C}, H_{\mathcal{TU}}=h} \right| \le \left| Q_{\underline{\tilde{Y}}_{\mathcal{S}^C} \underline{\tilde{Y}}_T \,|\, H_{\mathcal{TU}}=h} \right| \quad (81)$$

where the inequality follows because conditioning cannot increase entropy (see [25, Sec. 16.8]). Equality holds in (81) if $\underline{X}_{\{1\}\cup\mathcal{S}}$ and $\underline{X}_{\mathcal{S}^C}$ are statistically independent.

Observe that, in (79), we can replace H_{1t} with $-H_{1t}$ for all t because it is immaterial what phase we begin integrating from for any $H_{st}^{(i,j)}$. This step is valid for any fading process where the $H_{st}^{(i,j)}$ have independent and uniform phases. Moreover, this step is equivalent to using the same H_{1t} as originally, but replacing $\underline{X}_1$ with $-\underline{X}_1$. But this change makes all cross-correlation matrices $\mathrm{E}[\underline{X}_1 \underline{X}_s^\dagger]$ with $s \neq 1$ change sign. We can thus use the concavity of $\log(|A|)$ in positive-semidefinite A, and apply Jensen's inequality to show that the mutual information expressions cannot decrease if we make $\mathrm{E}[\underline{X}_1 \underline{X}_s^\dagger] = 0$ for all $s \neq 1$. $\underline{X}_1$ and $\underline{X}_{\{2,3,\dots,T-1\}}$ are therefore independent. Repeating these steps for all input vectors, the best input distribution has *independent* $\underline{X}_t, t = 1, 2, \dots, T-1$. This implies that equality holds in (81).

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 H_{1t} with their negative counterparts. This is equivalent to using the same H_{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 $\underline{X}_1^{(i)}, i = 2, 3, \dots, n_1$. 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 replace H_{tu} with $H_{tu}\Pi_t$ for $u = 2, \dots, T$, where Π_t is an $n_t \times n_t$ permutation matrix. Equivalently, we can permute the diagonal elements of the $Q_{\underline{X}_t}$ without changing the mutual information expressions. Applying Jensen's inequality, we find that the best input distributions are

$$Q_{\underline{X}_t} = \sqrt{\frac{P_t}{n_t}} I \quad (82)$$

where I is the appropriately sized identity matrix. We have thus proved that the optimal $\underline{X}_t$ are independent and satisfy (82) for the cut-set bound (5). The same claim can be made for the decode-and-forward rate (21) by using the same arguments. Using (51), the decode-and-forward rates (21) are thus

$$I\left(X_{\pi(1:s)}; Y_{\pi(s+1)} \,\middle|\, X_{\pi(s+1:T-1)}\right) = \int_h p(h) \log\left(\left| I + \sum_{t \in \pi(1:s)} \frac{P_t}{n_t} \frac{h_{t\pi(s+1)} h_{t\pi(s+1)}^\dagger}{d_{t\pi(s+1)}^\alpha} \right|\right) dh \quad (83)$$

for $s = 1, 2, \dots, T-1$, where the integral is over all channel matrices h_{st}, $s = 1, 2, \dots, T-1$, $t = 2, 3, \dots, T$. The rest of the proof is similar to the proof of Theorem 7, i.e., if the rate with $s = T-1$ in (83) is the smallest of the $T-1$ rates, then this rate is the capacity. □

Remark 39: Theorem 8 appeared in [67]. The theorem was also derived in [68], [76] for one relay by using the proof technique of [67].

Remark 40: It is clear that Theorem 8 generalizes to many other fading processes where the phase varies uniformly over time and the interval $[0, 2\pi)$, including fading processes with memory. The main insight is that, because the transmitting nodes do not have knowledge about their channel phases to the destination, their best (per-letter) inputs are independent when the destination is far from all of them.

Remark 41: The geometries for which one achieves capacity can be computed, but they do depend on the type of fading, the number of antennas, and the attenuation exponent. However, one always achieves capacity if all the relays are near the source node (but are not necessarily colocated with it).

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. The rate (78) is then

$$\int_0^{2\pi}\int_0^{2\pi} \frac{1}{4\pi^2} \log\left(\left| I + \frac{P_1}{d_{13}^\alpha}\begin{bmatrix} 1 & e^{j\phi_1} \\ e^{-j\phi_1} & 1 \end{bmatrix} + \frac{P_2}{d_{23}^\alpha}\begin{bmatrix} 1 & e^{j\phi_2} \\ e^{-j\phi_2} & 1 \end{bmatrix} \right|\right) d\phi_1\, d\phi_2 = \log\left(\frac{a + \sqrt{a^2 - b^2}}{2}\right) \quad (84)$$

where ϕ_1 and ϕ_2 represent phase differences and

$$a = 1 + \frac{2P_1}{d_{13}^\alpha} + \frac{2P_2}{d_{23}^\alpha} + \frac{2P_1P_2}{d_{13}^\alpha d_{23}^\alpha}, \quad b = \frac{2P_1P_2}{d_{13}^\alpha d_{23}^\alpha}. \quad (85)$$

Recall from the proof of Theorem 8 that (84) is $I(X_1X_2; Y_3)$. For the geometry of Fig. 15, capacity is therefore achieved if (84) is less than $I(X_1; Y_2 \,|\, X_2) = \log(1 + P_1/d_{12}^\alpha)$. That is, one achieves capacity if

$$\frac{a + \sqrt{a^2 - b^2}}{2} \le 1 + \frac{P_1}{d_{12}^\alpha}. \quad (86)$$

Again, this condition is satisfied for a range of d_{12} near zero. Fig. 20 plots the cut-set and decode-and-forward rates for $P_1 = P_2 = 10$ and the same range of d as in Fig. 18. We have also plotted the decode-and-forward rates from Fig. 18, and the compress-and-forward rates when the relay again uses $\hat{Y}_2 = Y_2 + \hat{Z}_2$. The latter rates are now

$$R_{\mathrm{CF}} = \log\left(1 + \frac{P_1}{d_{12}^\alpha(1 + \hat{N}_2)} + \frac{2P_1}{d_{13}^\alpha}\right) \quad (87)$$

where the choice

$$\hat{N}_2 = \frac{P_1(1/d_{12}^\alpha + 2/d_{13}^\alpha) + 1}{(a + \sqrt{a^2 - b^2})/2 - (2P_1/d_{13}^\alpha + 1)} \quad (88)$$

satisfies (30) with equality. Compress-and-forward again performs well for all d and achieves capacity for $d = 0$ and $d = 1$.

Fig. 21 plots the region of relay positions that satisfy (86) for a two-dimensional geometry (cf. Fig. 19). 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 to the source. Observe that the regions

www.manaraa.com

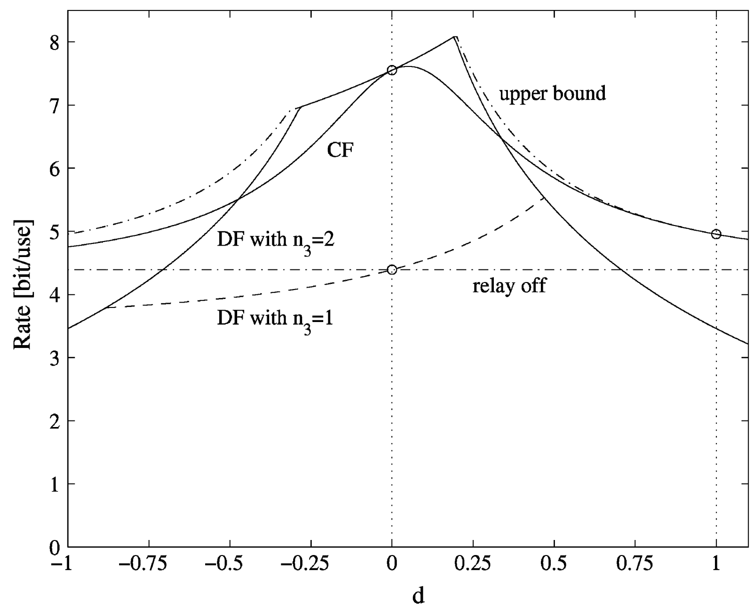

Fig. 20. Rates for a single-relay network with phase fading, $n_1 = n_2 = 1, n_3 = 2, P_1 = P_2 = 10$, and $\alpha = 2$.

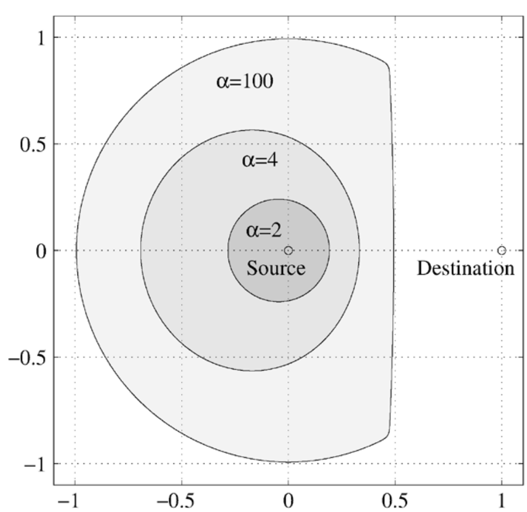

Fig. 21. Positions of the relay where decode-and-forward achieves capacity with phase fading, $n_1 = n_2 = 1, n_3 = 2$, and $P_1 = P_2 = 10$.

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 $n_3 = 1$.

Remark 42: An important limitation of our models is that the network operates *synchronously*. The transmitting nodes might therefore need to be symbol synchronized, and this is likely difficult to implement in wireless networks. However, we point out that as long as the signals are *band limited*, the decode-and-forward strategies with *independent* X_t do not require symbol synchronization between nodes. This statement can be justified as follows [75]. Suppose the $Y_t^N, t = 2, 3, \ldots, T$, are samples of $T-1$ band-limited waveforms. The Y_t^N are sufficient statistics about the transmitted signals if the sampling rates are at or above the Nyquist rate [96, p. 53]. Suppose further that one uses the regular encoding/sliding-window decoding version of the decode-and-forward strategy. Node t then decodes after block b by interpolating its $Y_{tb'}^n, b' = b, b-1, \ldots, b-t+2$, by using different delays for every b'. This will permit communication at the rates (21) because, ignoring boundary effects, the interference energy in every block b' is unaffected by the interpolator time shifts (the energy of a time-shifted and Nyquist-rate sampled version of a band-limited signal is the same for every time shift [77, Ch. 8.1]).

G. Fading With Directions

Single-bounce fading and fading with directions can be dealt with as above, in the sense that the best input distribution has independent $\underline{X}_t, t = 1, 2, \ldots, T-1$. However, now the best $Q_{\underline{X}_t}$ are not necessarily given by (82) and they might not be diagonal. For example, suppose we have Rayleigh fading with directions, i.e., G_{st} has Gaussian entries. It is unclear what the best choice of $Q_{\underline{X}_s}$ should be because each $\underline{X}_s$ goes through multiple B_{st} and C_{st}. Nevertheless, capacity is again achieved if all relays are near the source node, because the best choice of $Q_{\underline{X}_s}$ will be the same for both (5) and (21).

There are naturally some simple cases where we can say more. For example, suppose there is single-bounce fading, i.e., we have $H_{st} = B_{st}D_{st}C_{st}$ where B_{st}, D_{st}, and C_{st} are $n_t \times n_{st}, n_{st} \times n_{st}$ and $n_{st} \times n_s$ matrices, respectively.

Proposition 3: Consider single-bounce fading with $n_t = n_{st} = n_s = n_1, B_{st} = I, |D_{st}^{(i,i)}| = 1$, and $C_{st} = I$ for all s, t, i. Decode-and-forward achieves capacity if

$$\sum_{t=1}^{T-1} \frac{P_t}{d_{tT}^\alpha} \leq \max_{\pi(\cdot)} \min_{1 \leq s \leq T-2} \sum_{t \in \pi(1:s)} \frac{P_t}{d_{t\pi(s+1)}^\alpha} \tag{89}$$

and the resulting capacity is

$$C = n_1 \log\left(1 + \frac{1}{n_1} \sum_{t=1}^{T-1} \frac{P_t}{d_{tT}^\alpha}\right). \tag{90}$$

The condition (89) is identical to (68), but the capacity (90) increases with n_1.

Proof: We effectively have n_1 parallel phase fading channels between the nodes. Using the same steps as in the proof of Theorem 8, we find that the best input covariance matrices are

المنارة للاستشارات
www.manaraa.com

given by (82). Using (51), the decode-and-forward rates (21) are thus

$$I\left(X_{\pi(1:s)};Y_{\pi(s+1)}\left|X_{\pi(s+1:T-1)}\right.\right) = \log\left(\left|\left\{1+\sum_{t\in\pi(1:s)}\frac{P_t/n_1}{d_{t\pi(s+1)}^{\alpha}}\right\}I\right|\right). \quad (91)$$

for $s=1,2,\ldots,T-1$. The remaining steps are the same as in the proof of Theorem 7 or Theorem 8. □

H. Multisource Networks and Phase Fading

The capacity theorems derived above generalize to several multisource and multidestination networks. For instance, consider MARCs with phase fading. We find that the best input distribution for (24) and (25) is Gaussian (see Section VII-A). We further find that $U_1=U_2=0$ is best, i.e., the $\underline{X}_t$ are independent. One again achieves capacity if the source and relay nodes are near each other, and we summarize this with the following theorem.

Theorem 9: The decode-and-forward strategy of Section IV-D achieves all points inside the capacity region of MARCs with phase fading if

$$\begin{aligned}&P_1/d_{14}^{\alpha}+P_3/d_{34}^{\alpha}\le P_1/d_{13}^{\alpha}\\&P_2/d_{24}^{\alpha}+P_3/d_{34}^{\alpha}\le P_2/d_{23}^{\alpha}\\&P_1/d_{14}^{\alpha}+P_2/d_{24}^{\alpha}+P_3/d_{34}^{\alpha}\le P_1/d_{13}^{\alpha}+P_2/d_{23}^{\alpha}\end{aligned} \quad (92)$$

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

$$\begin{aligned}&0\le R_1\le\log(1+P_1/d_{14}^{\alpha}+P_3/d_{34}^{\alpha})\\&0\le R_2\le\log(1+P_2/d_{14}^{\alpha}+P_3/d_{34}^{\alpha})\\&R_1+R_2\le\log(1+P_1/d_{14}^{\alpha}+P_2/d_{24}^{\alpha}+P_3/d_{34}^{\alpha}).\end{aligned} \quad (93)$$

Proof: The proof is omitted. The idea is to use (24), (25), and the same steps as the proof of Theorem 6. □

Generalizations of Theorem 9 to include more sources and relays, as well as multiple antennas and Rayleigh fading, are clearly 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 bound [25, p. 445] and use (28) with independent X_1 and X_2 to prove the following theorem.

Theorem 10: The decode-and-forward strategy of Section IV-E achieves the capacity of BRCs with phase fading and with a common message if

$$\min\left[\frac{P_1}{d_{13}^{\alpha}}+\frac{P_2}{d_{23}^{\alpha}},\frac{P_1}{d_{14}^{\alpha}}+\frac{P_2}{d_{24}^{\alpha}}\right]\le P_1/d_{12}^{\alpha} \quad (94)$$

and the resulting capacity is

$$C=\min\left[\log\left(1+\frac{P_1}{d_{13}^{\alpha}}+\frac{P_2}{d_{23}^{\alpha}}\right),\log\left(1+\frac{P_1}{d_{14}^{\alpha}}+\frac{P_2}{d_{24}^{\alpha}}\right)\right]. \quad (95)$$

Proof: The proof uses the same steps as the proof of Theorem 6, and is omitted. □

Theorem 10 generalizes to other fading models. However, some care is needed because the BRCs might not be degraded.

I. Quasi-Static Fading

Quasi-static fading has the H_{st} chosen randomly at the beginning of time and held fixed for all channel uses [89, Sec. 5]. The information rate for a specified $Q_{\underline{X}_\mathcal{T}}$ is therefore a random variable that is a function of the fading random variables $H_{st}^{(i,j)}$ [98, p. 2631]. The resulting analysis is rather more complicated than for ergodic fading. We illustrate the differences by studying one relay, single-antenna nodes, and phase fading.

Suppose we use decode-and-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 sliding-window decoding. The destination will likely make errors if either $I(X_1X_2;Y_3\,|\,H_{13}H_{23})$ or $I(X_1;Y_2\,|\,X_2H_{12})$ 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 information expressions is too small.

The best $P(x_1,x_2)$ for all our bounds and for any realization of H_{12},H_{13}, and H_{23} is again zero-mean Gaussian (see Section VII-A), but one must now carefully adjust $\rho=\mathrm{E}[X_1X_2^*]/\sqrt{P_1P_2}$. Recall that $H_{st}=e^{j\theta_{st}}$, and let $\theta=\theta_{13}-\theta_{23}+\theta_\rho$ where θ_ρ is the phase of ρ. The minimum in (12) is thus the minimum in (54) for a fixed ρ, i.e., it is the random variable

$$\Psi(\rho,\theta)=\min\left[\log\left(1+\frac{P_1}{d_{12}^{\alpha}}(1-|\rho|^2)\right),\right.\left.\log\left(1+\frac{P_1}{d_{13}^{\alpha}}+\frac{P_2}{d_{23}^{\alpha}}+\frac{2\sqrt{P_1P_2}|\rho|\cos\theta}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}}\right)\right] \quad (96)$$

that is a function of the random variable θ. Since θ is uniform over $[0,2\pi)$, we can restrict attention to real and nonnegative ρ. The decode-and-forward outage probability is thus

$$P_{\text{out}}^{\text{DF}}(R)=\min_{0\le\rho\le1}\Pr(\Psi(\rho,\theta)\le R). \quad (97)$$

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

Continuing with (97), observe that if

$$R>\log\left(1+\frac{P_1}{d_{12}^{\alpha}}\right) \quad (98)$$

then $P_{\text{out}}^{\text{DF}}(R)=1$. For smaller R, we infer from (96) that one should choose ρ as large as possible, i.e., choose the positive ρ satisfying

$$R=\log\left(1+\frac{P_1}{d_{12}^{\alpha}}(1-\rho^2)\right). \quad (99)$$

The random component of (96) is $\xi=\cos\theta$ that has the cumulative distribution

$$\Pr(\xi\le x)=\begin{cases}0, & x<-1\\ \frac{1}{2}+\frac{1}{\pi}\arcsin(x), & -1\le x\le1\\ 1, & x>1.\end{cases} \quad (100)$$

Using (99) and (100), we compute

$$P_{\text{out}}^{\text{DF}}(R)=\begin{cases}0, & f(R)<0\\ \frac{1}{2}+\frac{1}{\pi}\arcsin(f(R)), & 0\le f(R)\le1\\ 1, & f(R)>1\end{cases} \quad (101)$$

www.manaraa.com

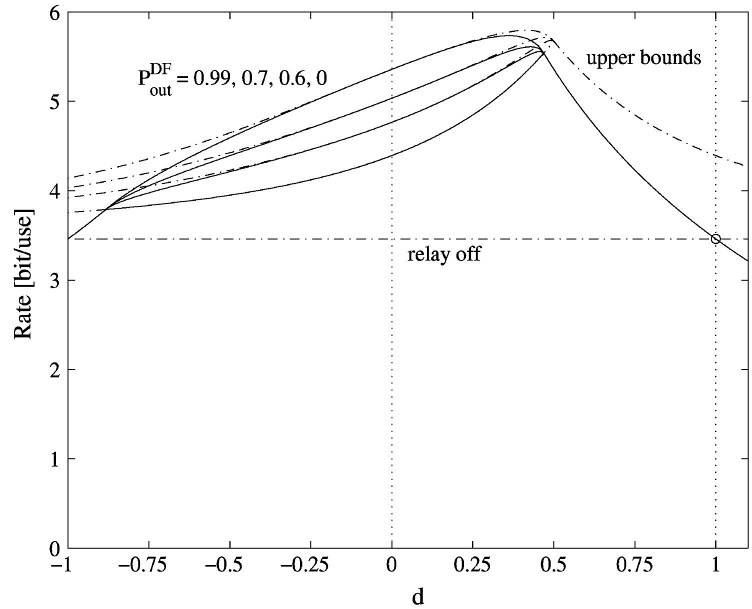

Fig. 22. Outage rates for decode-and-forward and a single-relay network with phase fading, $P_1 = P_2 = 10$, and $\alpha = 2$.

where

$$f(R) = \frac{\left(e^R - 1 - \frac{P_1}{d_{13}^\alpha} - \frac{P_2}{d_{23}^\alpha}\right) d_{13}^{\alpha/2} d_{23}^{\alpha/2}}{2\sqrt{P_2}\sqrt{P_1 - d_{12}^\alpha(e^R - 1)}}. \quad (102)$$

We remark that $P_{\rm out}^{\rm DF}(R) = 0$ clearly implies $P_{\rm out}(R) = 0$. Also, if $P_{\rm out}^{\rm DF}(R) \neq 0$ then $P_{\rm out}^{\rm DF}(R) \geq 1/2$. However, decode-and-forward is not necessarily optimal when $P_{\rm out}(R) \neq 0$.

A lower bound on $P_{\rm out}(R)$ can be computed using (53) and (100). We have

$$P_{\rm out}(R) = 1, \quad \text{if } R \geq \log\left(1 + P_1\left(\frac{1}{d_{12}^\alpha} + \frac{1}{d_{13}^\alpha}\right)\right). \quad (103)$$

If R is less than the far right-hand side of (103), we have

$$P_{\rm out}(R) \geq \begin{cases} 0, & g(R) < 0 \\ \frac{1}{2} + \frac{1}{\pi}\arcsin(g(R)), & 0 \leq g(R) \leq 1 \\ 1, & g(R) > 1 \end{cases} \quad (104)$$

where

$$g(R) = \frac{\left(e^R - 1 - \frac{P_1}{d_{13}^\alpha} - \frac{P_2}{d_{23}^\alpha}\right) d_{13}^{\alpha/2} d_{23}^{\alpha/2}}{2\sqrt{P_2}\sqrt{P_1 - d_{12}^\alpha d_{13}^\alpha (e^R - 1)/(d_{12}^\alpha + d_{13}^\alpha)}}. \quad (105)$$

To illustrate these results, consider again the geometry of Fig. 15. We plot the decode-and-forward rates for $P_{\rm out}^{\rm DF} = 0, 0.6, 0.7, 0.99$ as the solid lines in Fig. 22 (the rates for $P_{\rm out}^{\rm DF}$ satisfying $0 \leq P_{\rm out}^{\rm DF} \leq 0.5$ are all the same). Observe that $R_{\rm DF}$ with $P_{\rm out}^{\rm DF} = 0$ is the same as $R_{\rm DF}$ *with* phase fading (see Fig. 18). Similarly, $R_{\rm DF}$ with $P_{\rm out}^{\rm DF} \to 1$ is the same as $R_{\rm DF}$ *without* phase fading (see Fig. 16). The dash-dotted curves are upper bounds on the best possible rates for $P_{\rm out} = 0, 0.6, 0.7, 0.99$. These rates were computed with (103)–(105), and they approach the upper bound in Fig. 16 as $P_{\rm out} \to 1$.

Remark 43: The analysis for Rayleigh fading is similar to the above. The information rate is now the random variable

$$\Psi\left(\rho, H_{\{1,2\}\{2,3\}}\right) = \min\left[\log\left(1 + \frac{P_1|H_{12}|^2}{d_{12}^\alpha}(1 - |\rho|^2)\right), \log\left(1 + \frac{P_1|H_{13}|^2}{d_{13}^\alpha} + \frac{P_2|H_{23}|^2}{d_{23}^\alpha} + \frac{2\sqrt{P_1P_2}\Re(\rho H_{13}H_{23}^*)}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}}\right)\right] \quad (106)$$

that is a function of the Gaussian fading random variables H_{12}, H_{13}, and H_{23}. Note that the two expressions inside the minimization are independent random variables, which helps to simplify the analysis somewhat. The outage statistics for the first random variable can be computed by using the incomplete gamma function as in [89, Sec. 5.1].

Remark 44: Suppose that, instead of quasi-static fading, we have *block fading* where the H_{st} are chosen independently from block to block. The relay outage probability with decode-and-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 (97).

Remark 45: Suppose that only the channel *amplitudes* exhibit quasi-static fading, while the channel *phases* behave in an ergodic fashion. This problem was studied in [60], and it is easier to treat than the above problem because the best correlation coefficient ρ is always zero. One thus gets more extensive "quasi-static" capacity results than those described here.

www.manaraa.com

VIII. Conclusion

We developed several coding strategies for relay networks. The decode-and-forward strategies are useful for relays that are close to the source, and the compress-and-forward strategies are useful for relays that are close to the destination (and sometimes even close to the source). A strategy that mixes decode-and-forward and compress-and-forward achieves capacity if the network nodes form two closely spaced clusters. We further showed that decode-and-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. The capacity results extend to multisource 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 [99].

Appendix A
Decode-and-Forward for MARCs

Consider MARCs with $T-2$ source nodes, a relay node $T-1$, and a destination node T. In what follows, we give only a sketch of the proof because the analysis is basically the same as for MAC decoding (see [25, p. 403]) and MAC backward decoding (see [21, Ch. 7]).

Let $\mathcal{T}'=\{1,2,\dots,T-2\}$ 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 node $t\in\mathcal{T}'$ sends B message blocks $w_{t1},w_{t2},\dots,w_{tB}$ in B+1 transmission blocks.

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

$$p(\underline{x}_{T-1}\,|\,\underline{u}_{\mathcal{T}'}(r_{\mathcal{T}'}))=\prod_i p_{X_t\,|\,U_{\mathcal{T}'}}(x_{ti}\,|\,u_{\mathcal{T}'i}(r_{\mathcal{T}'})).$$

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 by using $\underline{y}_{(T-1)b}$ and by assuming its past message estimates are correct. The decoding problem is the same as for a MAC with side information $U_{\mathcal{T}'}$ and X_{T-1}, and one can decode reliably if (see [25, p. 403])

$$R_{\mathcal{S}}<I\left(X_{\mathcal{S}};Y_{T-1}\,|U_{\mathcal{T}'}X_{\mathcal{S}^C}X_{T-1}\right)\qquad(107)$$

for all $\mathcal{S}\subseteq\mathcal{T}'$, where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{T}'$.

2) *(At the destination)* The destination waits until all transmission is completed. It then decodes $w_{\mathcal{T}'b}$ for $b=B,B-1,\dots,1$ by using its output block $\underline{y}_{T(b+1)}$ and by assuming its message estimates $\hat{w}^{(T)}_{\mathcal{T}'(b+1)}$ are correct. The techniques of [21, Ch. 7] ensure that one can decode reliably if

$$R_{\mathcal{S}}<I\left(X_{\mathcal{S}}X_{T-1};Y_T\,|U_{\mathcal{S}^C}X_{\mathcal{S}^C}\right)\qquad(108)$$

for all $\mathcal{S}\subseteq\mathcal{T}'$, where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{T}'$.

Remark 46: For $T=4$ and $\mathcal{T}'=\{1,2\}$, we recover (24) and (25).

Remark 47: The above rates might be improved by introducing a time-sharing random variable as described in [25, p. 397] (see also [100]).

Appendix B
Decode-and-Forward for BRCs

Consider a BRC with four nodes as in Sections III-D and IV-E. We again give only a sketch of the proof, albeit a detailed one, because the analysis is based on standard arguments (see [25, Sec. 14.2] and [82]–[85]).

The source node divides its messages W_0,W_1,W_2 into B blocks w_{0b},w_{1b},w_{2b} for $b=1,2,\dots,B$, 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 (see [84, Lemma 1]). 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. 23 (the bits w_{0b} in w'_{1b} and w'_{2b} must be the same on the right in Fig. 23). The corresponding restrictions on the rates R_0,R_1,R_2 are

$$\begin{aligned}R_0+R_1&\le R'_0+R'_1\\R_0+R_2&\le R'_0+R'_2\\R_0+R_1+R_2&\le R'_0+R'_1+R'_2.\end{aligned}\qquad(109)$$

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

$$R''_1\ge R'_1,\quad R''_2\ge R'_2.\qquad(110)$$

Random Code Construction: We construct the following codebooks independently for *all* blocks $b, b=1,2,\dots,B+1$. However, for economy of notation, we will not label the codewords with their block. The reason we generate new codebooks

المنارة للاستشارات
www.manaraa.com

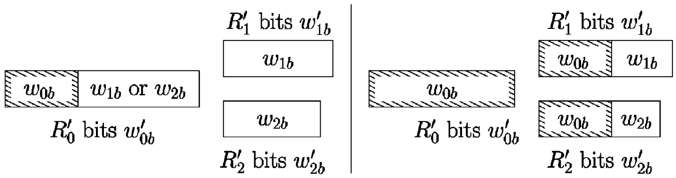

Fig. 23. Reorganization of the message bits for BRC coding.

for each block is to guarantee statistical independence for certain random coding arguments.

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

$$p(\underline{u}_0 \,|\, \underline{x}_2(q)) = \prod_i p_{U_0|X_2}(u_{0i} \,|\, x_{2i}(q)).$$

Label these $\underline{u}_0(q, r_0), r_0 \in [1, 2^{nR_0'}]$.

3) For each $\underline{u}_0(q, r_0)$ choose $2^{nR_1''}$ i.i.d. $\underline{u}_1$ with

$$p(\underline{u}_1 \,|\, \underline{x}_2(q), \underline{u}_0(q, r_0)) = \prod_i p_{U_1|X_2U_0}(u_{1i} \,|\, x_{2i}(q), u_{0i}(q, r_0)).$$

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

$$p(\underline{u}_2 \,|\, \underline{x}_2(q), \underline{u}_0(q, r_0)) = \prod_i p_{U_2|X_2U_0}(u_{2i} \,|\, x_{2i}(q), u_{0i}(q, r_0)).$$

Label these $\underline{u}_2(q, r_0, s_2), s_2 \in [1, 2^{nR_2''}]$.

5) Randomly partition the set $\{1, \ldots, 2^{nR_1''}\}$ into $2^{nR_1'}$ cells S_{r_1} with $r_1 \in [1, 2^{nR_1'}]$.

6) Randomly partition the set $\{1, \ldots, 2^{nR_2''}\}$ 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 cells in this step and the previous step. 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

$$p(\underline{x}_1 \,|\, \underline{x}_2(q), \underline{u}_0(q, r_0), \underline{u}_1(q, r_0, s_1), \underline{u}_2(q, r_0, s_2)) = \prod_i p_{X_1\,|\,X_2U_0U_1U_2}(x_{1i} \,|\, x_{2i}(q), u_{0i}(q, r_0), u_{1i}(q, r_0, s_1), u_{2i}(q, r_0, s_2)).$$

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 node finds a pair

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

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 node 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 $\underline{x}_2(w_{0(b-1)}')$.

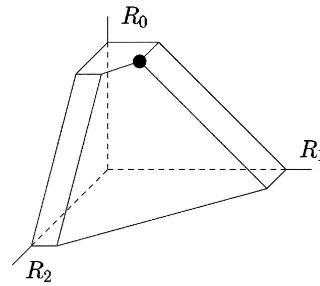

Fig. 24. Achievable rate region for the BRC.

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 [85]) guarantee that such a pair exists with high probability if n is large, ϵ is small, and

$$(R_1'' - R_1') + (R_2'' - R_2') > I(U_1; U_2 \,|\, U_0 X_2). \quad (111)$$

Decoding: After block b decoding proceeds as follows.

1) *(At the relay)* The relay decodes w_{0b}' by using $\underline{y}_{2b}$ and by assuming its past message estimates are correct. Decoding can be done reliably if

$$R_0' < I(U_0; Y_2 \,|\, X_2). \quad (112)$$

2) *(At the destinations)* Node 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. 11), and by assuming its past message estimates are correct. Similarly, node 4 decodes $w_{0(b-1)}'$ and $s_{2(b-1)}$ by using $\underline{y}_{4(b-1)}$ and $\underline{y}_{4b}$, and by assuming its past message estimates are correct. The techniques of [20], [32], [33] ensure that both nodes decode reliably if

$$\begin{aligned} R_1'' &< I(U_1; Y_3 \,|\, U_0 X_2) \\ R_0' + R_1'' &< I(U_0 U_1 X_2; Y_3) \\ R_2'' &< I(U_2; Y_4 \,|\, U_0 X_2) \\ R_0' + R_2'' &< I(U_0 U_2 X_2; Y_4). \end{aligned} \quad (113)$$

Nodes 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}').

The rate region of (26) has the form depicted in Fig. 24 (cf. [83, Fig. 1]). We proceed as in [82, Sec. III] and show that one can approach the corner point

$$\begin{aligned} R_0 &= \min(I_2, I_3, I_4) \\ R_1 &= \min(I_2, I_3) + I(U_1; Y_3 \,|\, U_0 X_2) - \min(I_2, I_3, I_4) \\ R_2 &= \min(I_2, I_3, I_4) + I(U_2; Y_4 \,|\, U_0 X_2) \\ &\quad - I(U_1; U_2 \,|\, U_0 X_2) - \min(I_2, I_3) \end{aligned} \quad (114)$$

where we recall that

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

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

$$\begin{aligned} R_1'' &= R_1' \\ R_2'' &= R_2' + I(U_1; U_2 \,|\, U_0 X_2) + \epsilon \end{aligned} \quad (115)$$

for a small positive ϵ. This choice satisfies (110) and (111). We next consider two cases separately.

www.manaraa.com

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

$$\begin{aligned} R_0 &= R_0' = \min(I_2, I_3) - \epsilon \\ R_1 &= R_1' = I(U_1; Y_3 \,|\, U_0 X_2) - \epsilon \\ R_2 &= R_2' = I(U_2; Y_4 \,|\, U_0 X_2) - I(U_1; U_2 \,|\, U_0 X_2) - 2\epsilon \end{aligned} \tag{116}$$

to satisfy (109), (112), and (113). We can thus approach the corner point (114) by letting $\epsilon \to 0$.

2) Suppose we have $\min(I_2, I_3, I_4) = I_4$. We choose

$$\begin{aligned} R_0 &= I_4 - \epsilon \\ R_0' &= \min(I_2, I_3) - \epsilon \\ R_1 &= \min(I_2, I_3) + I(U_1; Y_3 \,|\, U_0 X_2) - I_4 - \epsilon \\ R_1' &= I(U_1; Y_3 \,|\, U_0 X_2) - \epsilon \\ R_2 &= R_2' = I(U_2; Y_4 \,|\, U_0 X_2) - I(U_1; U_2 \,|\, U_0 X_2) \\ &\quad + I_4 - \min(I_2, I_3) - 2\epsilon \end{aligned} \tag{117}$$

which satisfies (109), (112), and (113) (note that we are using the mapping on the left in Fig. 23 because we have $R_0 \le R_0'$). We can again approach the corner point (114) by letting $\epsilon \to 0$.

One can thus approach both $R_0 = \min(I_2, I_3, I_4)$ corner points in Fig. 24. 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. 24 are achieved by time sharing.

Appendix C
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, Definition 1])

$$T_\epsilon^{(n)}(X) = \left\{ \underline{x} : \left| \frac{n_{\underline{x}}(a)}{n} - p_X(a) \right| < \frac{\epsilon}{|\mathcal{X}|} \text{ for all } a \in \mathcal{X} \right\} \tag{118}$$

where $\mathcal{X}$ 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 [25, p. 387].

Lemma 1: Consider the distribution $p(s_1, s_2, \ldots, s_M)$, and let $(\underline{s}_1', \underline{s}_2', \ldots, \underline{s}_M')$ be a vector of M-tuples

$$(s_{1i}', s_{2i}', \ldots, s_{Mi}'), \qquad i = 1, 2, \ldots, n$$

that are chosen i.i.d. with distribution $p(s_1) \prod_{m=2}^{M} p(s_m \,|\, s_1)$. The probability P that $(\underline{s}_1', \underline{s}_2', \ldots, \underline{s}_M')$ is in $T_\epsilon^{(n)}(S_1, S_2, \ldots, S_M)$ is bounded by

$$(1-\epsilon) 2^{-n[2M\epsilon + \gamma]} \le P \le 2^{-n[-2M\epsilon + \gamma]} \tag{119}$$

where

$$\gamma = -H(S_2 \ldots S_M \,|\, S_1) + \sum_{m=2}^{M} H(S_m \,|\, S_1). \tag{120}$$

Proof: We omit the proof because of its similarity to [25, p. 387]. Lemma 1 includes unconditioned bounds by making S_1 a constant. □

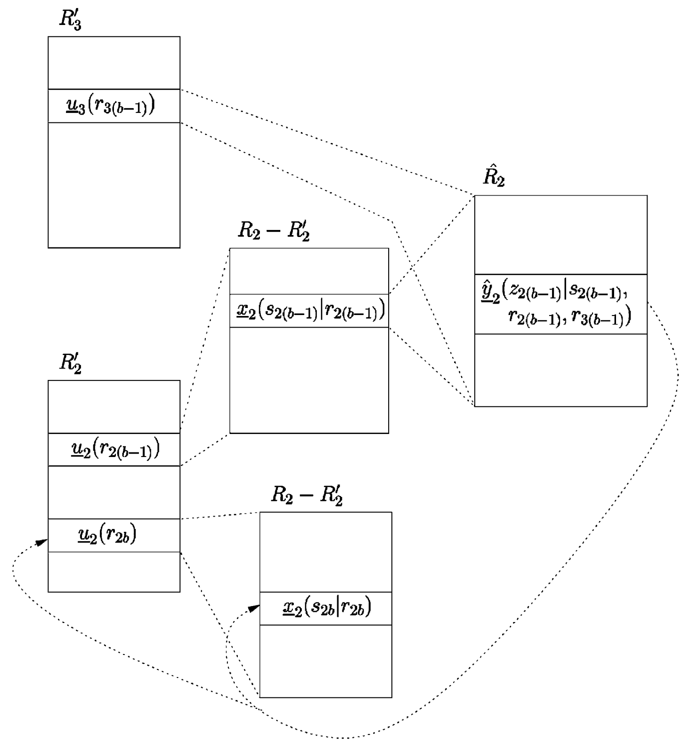

Fig. 25. The code used by node 2 for compress-and-forward.

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 code construction is illustrated for $T = 4$ in Fig. 25, and it uses rates $\hat{R}_t, t \in \mathcal{T}$, that will be chosen during the course of the proof.

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 $\underline{\hat{y}}_t(z_t \,|\, s_t, r_{\mathcal{T}}), z_t \in [1, 2^{n\hat{R}_t}]$.

www.manaraa.com

5) For all $t \in \mathcal{T}$, randomly partition the set $\{1,\ldots,2^{n\hat{R}_t}\}$ 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 node *and* the relays, while the $\underline{x}_t$ are decoded by the destination node 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 node sends $\underline{x}_1(w_b)$.
2) For all $t \in \mathcal{T}$, node t knows $z_{t(b-1)}$ from decoding step 4) described below. Node t chooses (r_{tb}, s_{tb}) so that $z_{t(b-1)} \in S_{r_{tb},s_{tb}}$, and it transmits $\underline{x}_t(s_{tb} \mid r_{tb})$.

Decoding: After block b decoding proceeds as follows.

1) *(At the destination)* The destination first decodes the $T-2$ indices $r_{\mathcal{T}b}$ using $\underline{y}_{Tb}$. This can be done reliably if (see [25, p. 403])

$$R'_{\mathcal{S}} < I\left(U_{\mathcal{S}}; Y_T \,|U_{\mathcal{S}^C}\right) \tag{121}$$

for all $\mathcal{S} \subseteq \mathcal{T}$, where $\mathcal{S}^C$ is the complement of $\mathcal{S}$ in $\mathcal{T}$. The destination next decodes $s_{\mathcal{T}b}$. This can be done reliably if

$$R_{\mathcal{S}} - R'_{\mathcal{S}} < I\left(X_{\mathcal{S}}; Y_T \,|U_{\mathcal{T}} X_{\mathcal{S}^C}\right) \tag{122}$$

for all $\mathcal{S} \subseteq \mathcal{T}$.

2) *(At the destination)* The destination determines the set $\mathcal{L}(\underline{y}_{T(b-1)})$ of $z_{\mathcal{T}}$ such that

$$\Big(\Big\{\underline{u}_t\left(r_{t(b-1)}\right), \underline{x}_t\left(s_{t(b-1)}\,\middle|r_{t(b-1)}\right), \\ \underline{\hat{y}}_t\left(z_t\,\middle|s_{t(b-1)}, r_{\mathcal{T}(b-1)}\right) : t \in \mathcal{T}\Big\}, \underline{y}_{T(b-1)}\Big) \tag{123}$$

is jointly ϵ-typical, where the $r_{t(b-1)}$ and $s_{t(b-1)}$ were obtained from the first decoding step. The destination node declares that $z_{\mathcal{T}(b-1)}$ was sent in block $b-1$ if

$$z_{\mathcal{T}} \in \left(S_{s_{2b},r_{2b}} \times \cdots \times S_{s_{(T-1)b},r_{(T-1)b}}\right) \cap \mathcal{L}\left(\underline{y}_{T(b-1)}\right) \tag{124}$$

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 chosen incorrectly. We first partition $\mathcal{L}(\underline{y}_{T(b-1)})$ into 2^{T-2} sets

$$\mathcal{L}_{\mathcal{S}} = \Big\{z_{\mathcal{T}} : z_{\mathcal{T}} \in \mathcal{L}\left(\underline{y}_{T(b-1)}\right), z_t \neq z_{t(b-1)} \text{ for } t \in \mathcal{S}, \\ z_t = z_{t(b-1)} \text{ for } t \notin \mathcal{S}\Big\} \tag{125}$$

where $\mathcal{S} \subseteq \mathcal{T}$. These sets could be empty. We proceed to determine their average size. We follow [4] and write F_{b-1}^c for the event that all decisions in block $b-1$ were correct. We have

$$\mathrm{E}\left[\|\mathcal{L}_{\mathcal{S}}\| \,\middle|F_{b-1}^c\right] = \sum_{z_{\mathcal{T}} \in \mathcal{L}_{\mathcal{S}}} \mathrm{E}\left[\psi\left(Z_{\mathcal{T}}\,\middle|\underline{y}_{T(b-1)}\right) \,|F_{b-1}^c\right] \tag{126}$$

where

$$\psi\left(z_{\mathcal{T}}\,\middle|\underline{y}_{T(b-1)}\right) = \begin{cases} 1, & \text{if (123) is jointly typical} \\ 0, & \text{otherwise.} \end{cases} \tag{127}$$

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 ensures that

$$\begin{aligned} &\mathrm{E}\left[\psi\left(Z_{\mathcal{T}}\,\middle|\underline{y}_{T(b-1)}\right) |F_{b-1}^c\right] \\ &\le 2^{-n[-2(|\mathcal{S}|+2)\epsilon - H(\hat{Y}_{\mathcal{S}}\hat{Y}_{\mathcal{S}^C}Y_T \,|\, U_{\mathcal{T}}X_{\mathcal{T}}) + H(\hat{Y}_{\mathcal{S}^C}Y_T \,|\, U_{\mathcal{T}}X_{\mathcal{T}})} \\ &\quad ^{+\sum_{t\in\mathcal{S}} H(\hat{Y}_t \,|\, U_{\mathcal{T}}X_{\mathcal{T}})]} \\ &= 2^{-n[-2(|\mathcal{S}|+2)\epsilon - H(\hat{Y}_{\mathcal{S}} \,|\, U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{S}^C}Y_T) + \sum_{t\in\mathcal{S}} H(\hat{Y}_t \,|\, U_{\mathcal{T}}X_{\mathcal{T}})]}. \end{aligned} \tag{128}$$

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 $\mathrm{E}[\|\mathcal{L}_{\mathcal{S}}\| \,|\, F_{b-1}^c]$ by

$$2^{n\hat{R}_{\mathcal{S}}} 2^{-n[-2(|\mathcal{S}|+2)\epsilon - H(\hat{Y}_{\mathcal{S}} \,|\, U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{S}^C}Y_T) + \sum_{t\in\mathcal{S}} H(\hat{Y}_t \,|\, U_{\mathcal{T}}X_{\mathcal{T}})]}. \tag{129}$$

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

$$\begin{aligned} &\Pr(\text{error in step 2}) \\ &\le \Pr\left(\bigcup_{\mathcal{S}\subseteq\mathcal{T},\mathcal{S}\neq\emptyset} \bigcup_{z_{\mathcal{T}}\in\mathcal{L}_{\mathcal{S}}} \text{event (124) occurs} \,\middle|\, F_{b-1}^c\right) \\ &\le \sum_{\mathcal{S}\subseteq\mathcal{T},\mathcal{S}\neq\emptyset} \Pr\left(\bigcup_{z_{\mathcal{T}}\in\mathcal{L}_{\mathcal{S}}} \text{event (124) occurs} \,\middle|\, F_{b-1}^c\right) \\ &\le \sum_{\mathcal{S}\subseteq\mathcal{T},\mathcal{S}\neq\emptyset} \Pr\left(\bigcup_{z_{\mathcal{T}}\in\mathcal{L}_{\mathcal{S}}} \{z_t \in S_{s_{tb},r_{tb}} \text{ for all } t \in \mathcal{S}\} \,\middle|\, F_{b-1}^c\right) \\ &= \sum_{\mathcal{S}\subseteq\mathcal{T},\mathcal{S}\neq\emptyset} \mathrm{E}\left[\|\mathcal{L}_{\mathcal{S}}\| \,\middle|F_{b-1}^c\right] 2^{-nR_{\mathcal{S}}}. \end{aligned} \tag{130}$$

Inserting (129) into (130), we see that as long as

$$R_{\mathcal{S}} > \hat{R}_{\mathcal{S}} + H\left(\hat{Y}_{\mathcal{S}}\,\middle|U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{S}^C}Y_T\right) - \sum_{t\in\mathcal{S}} H(\hat{Y}_t \,|\, U_{\mathcal{T}}X_{\mathcal{T}}) \tag{131}$$

for all $\mathcal{S} \subseteq \mathcal{T}$ then the destination can determine $z_{\mathcal{T}(b-1)}$ reliably.

Decoding (continued):

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

$$\Big(\underline{x}_1(w_{b-1}), \Big\{\underline{u}_t\left(r_{t(b-1)}\right), \underline{x}_t\left(s_{t(b-1)}\,\middle|r_{t(b-1)}\right), \\ \underline{\hat{y}}_t\left(z_{t(b-1)}\,\middle|s_{t(b-1)}, r_{\mathcal{T}(b-1)}\right) : t \in \mathcal{T}\Big\}, \underline{y}_{T(b-1)}\Big) \tag{132}$$

is ϵ-typical. The Markov lemma (see [93, Lemma 4.1]) ensures that the probability that (132) is ϵ-typical approaches 1 as $n \to \infty$ for the transmitted w_{b-1}. Using Lemma 1, the probability that there exists a $w \neq w_{b-1}$ such that $\underline{x}_1(w)$ satisfies (132) is upper-bounded by

$$2^{-n(I(X_1; U_{\mathcal{T}}X_{\mathcal{T}}\hat{Y}_{\mathcal{T}}Y_T) + 6\epsilon)}. \tag{133}$$

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

$$R_1 < I\left(X_1; \hat{Y}_{\mathcal{T}}Y_T \,|U_{\mathcal{T}}X_{\mathcal{T}}\right) \tag{134}$$

www.manaraa.com

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

$$R'_{\mathcal{S}} < I(U_{\mathcal{S}}; Y_t | U_{\mathcal{S}^C} X_t) \tag{135}$$

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 (135). Suppose the $T-3$ estimates of the r_{sb} are correct. Relay t chooses any of the z_t so that

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

is jointly ϵ-typical. The probability P that there is *no* such z_t is bounded by

$$P \leq \left(1 - (1-\epsilon)2^{-n[4\epsilon + I(\hat{Y}_t; Y_t \,|\, U_{\mathcal{T}} X_t)]}\right)^{2^{n\hat{R}_t}} \tag{136}$$

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

$$\log P \leq 2^{n\hat{R}_t} \log\left(1 - (1-\epsilon)2^{-n[4\epsilon + I(\hat{Y}_t; Y_t \,|\, U_{\mathcal{T}} X_t)]}\right)$$
$$\leq -(1-\epsilon)2^{n\hat{R}_t} 2^{-n[4\epsilon + I(\hat{Y}_t; Y_t \,|\, U_{\mathcal{T}} X_t)]}. \tag{137}$$

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

$$\hat{R}_t = I(\hat{Y}_t; Y_t \,|\, U_{\mathcal{T}} X_t) + \delta. \tag{138}$$

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

Finally, we combine (131) and (138), and use (see (34))

$$H(\hat{Y}_t \,|\, U_{\mathcal{T}} X_t Y_t) = H\left(\hat{Y}_t \middle| U_{\mathcal{T}} X_{\mathcal{T}} \hat{Y}_{\mathcal{S}^C} Y_{\mathcal{T}} Y_T\right)$$

to obtain the left-hand side of (33). We combine (121), (122), and (135) to obtain the right-hand side of (33).

Appendix E
Proof of Theorem 5

One could prove Theorem 5 by using backward or sliding-window decoding. Instead, we use the proof technique of [4]. The code construction is illustrated in Fig. 26.

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

$$p(\underline{x}_1 \,|\, \underline{u}_2(r_2), \underline{x}_2(s_2 \,|\, r_2))$$
$$= \prod_i p_{X_1 \,|\, U_2 X_2}(x_{1i} \,|\, u_{2i}(r_2), x_{2i}(s_2 \,|\, r_2)).$$

Label these $\underline{x}_1(w \,|\, s_2, r_2), w \in [1, 2^{nR_1}]$.

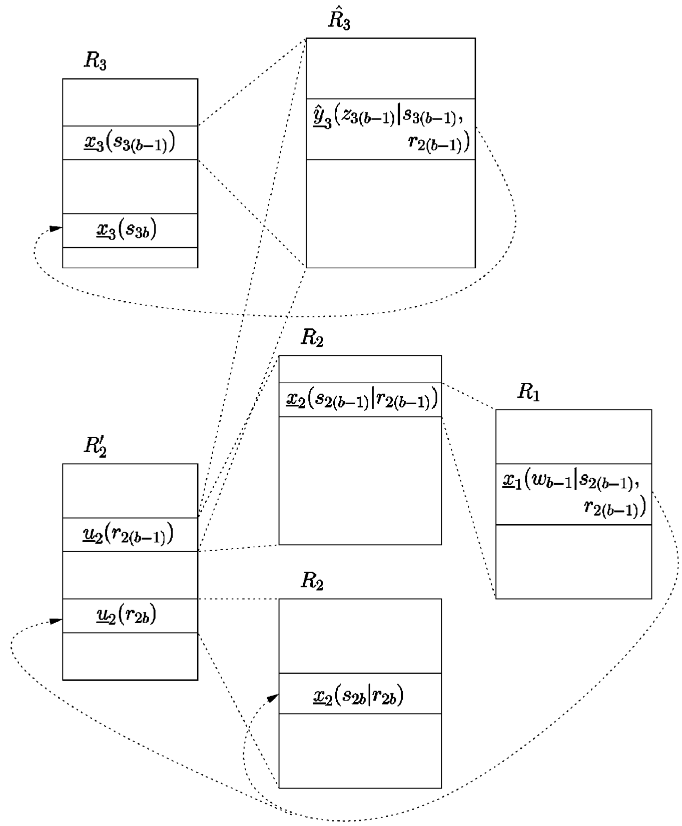

Fig. 26. The code construction for Theorem 5.

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

$$p(\underline{\hat{y}}_3 \,|\, \underline{x}_3(s_3), \underline{u}_2(r_2))$$
$$= \prod_i p_{\hat{Y}_3 \,|\, U_2 X_3}(\hat{y}_{3i} \,|\, u_{2i}(r_2), x_{3i}(s_3)).$$

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 $\{1, \ldots, 2^{nR_1}\}$ into 2^{nR_2} cells $S_{r_2, s_2}, r_2 \in [1, 2^{nR'_2}], s_2 \in [1, 2^{n(R_2 - R'_2)}]$.

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

Encoding:

1) The source transmits $\underline{x}_1(w_{b-1} \,|\, s_{2(b-1)}, r_{2(b-1)})$ in block $b-1$. It chooses (r_{2b}, s_{2b}) so that $w_{b-1} \in S_{r_{2b}, s_{2b}}$.

2) Node 2 knows w_{b-1} from decoding step 1) described below, and it chooses (r_{2b}, s_{2b}) so that $w_{b-1} \in S_{r_{2b}, s_{2b}}$. Node 2 transmits $\underline{x}_2(s_{2b} \,|\, r_{2b})$ in block b.

3) Node 3 knows $z_{3(b-1)}$ from decoding step 4) described below, and it chooses s_{3b} so that $z_{3(b-1)} \in S_{s_{3b}}$. Node 3 transmits $\underline{x}_3(s_{3b})$ in block b.

المنارة للاستشارات
www.manaraa.com

Decoding: After block b decoding proceeds as follows.

1) *(At node 2)* Node 2 chooses (one of) the message(s) w_b so that $(\underline{u}_2(r_{2b}), \underline{x}_1(w_b \mid s_{2b}), \underline{x}_2(s_{2b} \mid r_{2b}), \underline{y}_{2b})$ is jointly typical. This step can be made reliable if

$$R_1 < I(X_1; Y_2 \mid U_2 X_2). \tag{139}$$

2) *(At the destination)* The destination decodes r_{2b} and s_{3b}. This step can be made reliable if

$$R_2' < I(U_2; Y_4 \mid X_3) \tag{140}$$
$$R_3 < I(X_3; Y_4 \mid U_2) \tag{141}$$
$$R_2' + R_3 < I(U_2 X_3; Y_4). \tag{142}$$

3) *(At the destination)* The destination determines the set $\mathcal{L}(\underline{y}_{4(b-1)})$ of z_3 such that

$$\left(\underline{u}_2\left(r_{2(b-1)}\right), \underline{x}_3\left(s_{3(b-1)}\right), \hat{\underline{y}}_3\left(z_3 \left| s_{3(b-1)}, r_{2(b-1)}\right.\right), \underline{y}_{4(b-1)}\right)$$

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 (see (131))

$$R_3 > \hat{R}_3 + H(\hat{Y}_3 \mid U_2 X_3 Y_4) - H(\hat{Y}_3 \mid U_2 X_3)$$
$$= \hat{R}_3 - I(\hat{Y}_3; Y_4 \mid U_2 X_3). \tag{143}$$

This is identical to step (ii) of the proof of [4, Theorem 6].

4) *(At the destination)* The destination chooses $s_{2(b-1)}$ so that

$$\left(\underline{u}_2\left(r_{2(b-1)}\right), \underline{x}_2\left(s_{2(b-1)} \left| r_{2(b-1)}\right.\right), \underline{x}_3\left(s_{3(b-1)}\right), \hat{\underline{y}}_3\left(z_{3(b-1)} \left| s_{3(b-1)}, r_{2(b-1)}\right.\right), \underline{y}_{4(b-1)}\right)$$

is jointly typical. Using Lemma 1, this step can be made reliable if

$$R_2 - R_2' < I(X_2; \hat{Y}_3 Y_4 \mid U_2 X_3). \tag{144}$$

5) *(At the destination)* The destination determines the set $\mathcal{L}_1(\underline{y}_{4(b-1)})$ of w such that

$$\left(\underline{u}_2\left(r_{2(b-1)}\right), \underline{x}_1\left(w \left| s_{2(b-1)}, r_{2(b-1)}\right.\right), \underline{x}_2\left(s_{2(b-1)} \left| r_{2(b-1)}\right.\right), \underline{x}_3\left(s_{3(b-1)}\right), \hat{\underline{y}}_3\left(z_{3(b-1)} \left| s_{3(b-1)}, r_{2(b-1)}\right.\right), \underline{y}_{4(b-1)}\right)$$

is jointly typical. The destination knows (s_{2b}, r_{2b}) and generates the intersection of $\mathcal{L}_1(\underline{y}_{4(b-1)})$ with those w in $S_{r_{2b}, s_{2b}}$. The correct w_b can be found reliably if

$$R_1 < I(X_1; \hat{Y}_3 Y_4 \mid U_2 X_2 X_3) + R_2. \tag{145}$$

This is the analog of step (iii) of the proof of [4, Theorem 6].

6) *(At node 3)* Node 3 decodes r_{2b} using $\underline{y}_{3b}$. This can be done reliably if

$$R_2' < I(U_2; Y_3 \mid X_3). \tag{146}$$

Finally, node 3 tries to find a z_3 such that

$$(\hat{\underline{y}}_3(z_3 \mid s_{3b}, r_{2b}, \underline{y}_{3b}, \underline{x}_3(s_{3b}), \underline{u}_2(r_{2b}))$$

is jointly typical. Such a z_3 exists with high probability for large n if

$$\hat{R}_3 > I(\hat{Y}_3; Y_3 \mid U_2 X_3). \tag{147}$$

This is the analog of step (iv) of [4, Theorem 6].

In summary, R_1 is bounded by (139) and (145). Inserting (144) into (145), we have

$$R_1 < \min\{I(X_1; Y_2 \mid U_2 X_2), I(X_1 X_2; \hat{Y}_3 Y_4 \mid U_2 X_3) + R_2'\}. \tag{148}$$

Combining (147) with (143) yields

$$R_3 > I(\hat{Y}_3; Y_3 \mid U_2 X_3 Y_4) \tag{149}$$

where we have used (47). Furthermore, R_2' and R_3 must satisfy (140)–(142) and (146), i.e., we have

$$R_2' < \min\{I(U_2; Y_3 \mid X_3), I(U_2; Y_4 \mid X_3)\} \tag{150}$$
$$R_3 < I(X_3; Y_4 \mid U_2) \tag{151}$$
$$R_2' + R_3 < I(U_2 X_3; Y_4). \tag{152}$$

The inequalities (148)–(152) correspond to those in (45)–(46).

ACKNOWLEDGMENT

The authors would like to thank the reviewers and the Associate Editor for their useful suggestions. M. Gastpar wishes to thank Bell Laboratories, Lucent Technologies, for their support during a summer internship in 2001. G. Kramer would like to thank L. Sankaranarayanan and A. J. de Lind van Wijngaarden for contributions regarding the MARC, and A. J. Grant and J. N. Laneman for discussions on the MAC-GF.

REFERENCES

[1] E. C. van der Meulen, "Transmission of information in a T-terminal discrete memoryless channel," Ph.D. dissertation, Univ. California, Berkeley, CA, Jun. 1968.
[2] ——, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, pp. 120–154, 1971.
[3] ——, "A survey of multiway channels in information theory: 1961–1976," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 1–37, Jan. 1977.
[4] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.
[5] R. C. King, "Multiple access channels with generalized feedback," Ph.D. dissertation, Stanford Univ., Stanford, CA, Mar. 1978.
[6] A. A. El Gamal, "Results in multiple user channel capacity," Ph.D. dissertation, Stanford Univ., Stanford, CA, May 1978.
[7] M. R. Aref, "Information flow in relay networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, Oct. 1980.
[8] A. A. El Gamal, "On information flow in relay networks," in *Proc. IEEE National Telecommunications Conf.*, vol. 2, Miami, FL, Nov. 1981, pp. D4.1.1–D4.1.4.
[9] A. A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 3, p. 536, May 1982.
[10] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1106–1110, Sep. 1988.

المنارة للاستشارات
www.manaraa.com

[11] P. Vanroose and E. C. van der Meulen, "Uniquely decodable codes for deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1203–1212, Jul. 1992.

[12] R. Ahlswede and A. H. Kaspi, "Optimal coding strategies for certain permuting channels," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 3, pp. 310–314, May 1987.

[13] K. Kobayashi, "Combinatorial structure and capacity of the permuting relay channel," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 813–826, Nov. 1987.

[14] J. N. Laneman, "Cooperative diversity in wireless networks: Algorithms and architectures," Ph.D. dissertation, MIT, Cambridge, MA, Aug. 2002.

[15] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjeee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80–89, Sep. 2004.

[16] A. Nosratinia and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

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

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

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

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

[21] F. M. J. Willems, "Informationtheoretical Results for the Discrete Memoryless Multiple Access Channel," Doctor in de Wetenschappen Proefschrift dissertation, Katholieke Univ. Leuven, Leuven, Belgium, Oct. 1982.

[22] C.-M. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1160–1165, Nov. 1989.

[23] J. N. Laneman and G. Kramer, "Window decoding for the multiaccess channel with generalized feedback," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 281.

[24] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Capacity theorems for the multiple-access relay channel," in *Proc. 42nd Annu. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Sep./Oct. 2004, pp. 1782–1791.

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

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

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

[28] B. E. Schein, "Distributed coordination in network information theory," Ph.D. dissertation, MIT, Cambridge, MA, Oct. 2001.

[29] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 136.

[30] 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.

[31] P. Gupta and P. R. Kumar, "Toward an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

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

[33] ——, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.

[34] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc. 41st Annu. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 1074–1083.

[35] A. Reznik, S. R. Kulkarni, and S. Verdú, "Degraded Gaussian multirelay channel: Capacity and optimal power allocation," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3037–3046, Dec. 2004.

[36] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Lower bounds on the capacity of Gaussian relay channel," in *Proc. 38th Annu. Conf. Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2004, pp. 597–602.

[37] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–11, Jan. 1976.

[38] M. Gastpar, "On Wyner-Ziv networks," in *Proc. 37th Asilomar Conf. Signals, Systems, and Computers*, Asilomar, CA, Nov. 2003, pp. 855–859.

[39] ——, "The Wyner–Ziv problem with multiple sources," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2762–2768, Nov. 2004.

[40] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity–Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[41] ——, "User cooperation diversity—Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

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

[43] T. E. Hunter and A. Nosratinia, "Cooperation diversity through coding," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 220.

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

[45] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1470–1476, Sep. 2004.

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

[47] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "Multihop diversity in wireless relaying channels," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1820–1830, Oct. 2004.

[48] P. Mitran, H. Ochiai, and V. Tarokh, "Space–time diversity enhancements using collaborative communications," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2041–2057, Jun. 2005.

[49] H. Ochiai, P. Mitran, and V. Tarokh, "Design and analysis of collaborative communication protocols for wireless sensor networks," *IEEE Trans. Inform. Theory*, submitted for publication.

[50] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proc. IEEE INFOCOM 2002*, New York, June 2002, pp. 1577–1586.

[51] ——, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.

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

[53] A. Høst-Madsen, "On the capacity of wireless relaying," in *Proc. IEEE Vehicular Technology Conf., VTC 2002 Fall*, vol. 3, Vancouver, BC, Canada, Sep. 2002, pp. 1333–1337.

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

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

[56] M. O. Hasna and M.-S. Alouini, "Optimal power allocation for relayed transmissions over Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conf., VTC 2003 Spring*, vol. 4, Apr. 2003, pp. 2461–2465.

[57] ——, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 216–218, May 2003.

[58] A. El Gamal and S. Zahedi, "Minimum energy communication over a relay channel," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 344.

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

[60] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for the wireless relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.

[61] R. U. Nabar, Ö. Oyman, H. Bölcskei, and A. J. Paulraj, "Capacity scaling laws in MIMO wireless networks," in *Proc. 41st Annu. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 378–389.

[62] U. Mitra and A. Sabharwal, "On achievable rates of complexity constrained relay channels," in *Proc. 41st Annu. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 551–560.

المنارة للاستشارات
www.manaraa.com

[63] B. Wang and J. Zhang, "MIMO relay channel and its application for cooperative communication in ad hoc networks," in *Proc. 41st Annu. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 1556–1565.

[64] 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.

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

[66] Z. Dawy and H. Kamoun, "The general Gaussian relay channel: Analysis and insights," in *Proc. 5th Int. ITG Conf. Source and Channel Coding*, Erlangen-Nuremberg, Germany, Jan. 2004, pp. 469–476.

[67] G. Kramer, M. Gastpar, and P. Gupta, "Information-theoretic multihopping for relay networks," in *Proc. 2004 Int. Zurich Seminar*, Zurich, Switzerland, Feb. 2004, pp. 192–195.

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

[69] M. Katz and S. Shamai (Shitz), "Communicating to co-located ad-hoc receiving nodes in a fading environment," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun.Jul. 2004, p. 115.

[70] S. Zahedi, M. Mohseni, and A. El Gamal, "On the capacity of AWGN relay channels with linear relaying functions," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 399.

[71] Y. Liang and V. V. Veeravalli, "The impact of relaying on the capacity of broadcast channels," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 403.

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

[73] Y. Liang and V. V. Veeravalli, "Gaussian orthogonal relay channels: Optimal resource allocation and capacity," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3284–3289, Sep. 2005.

[74] A. El Gamal, M. Mohseni, and S. Zahedi, "On reliable communication over additive white Gaussian noise relay channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[75] G. Kramer, "Models and theory for relay channels with receive constraints," in *Proc. 42nd Annu. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Sep./Oct. 2004, pp. 1312–1321.

[76] B. Wang, J. Zhang, and A. Høst-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

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

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

[79] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[80] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[81] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.

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

[83] S. I. Gel'fand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Pered. Inform.*, vol. 16, no. 1, pp. 24–34, Jan./Mar. 1980.

[84] T. S. Han, "The capacity region for the deterministic broadcast channel with a common message," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 122–125, Jan. 1981.

[85] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.

[86] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.

[87] M. Gastpar, "To code or not to code," Doctoral dissertation, Swiss Federal Institute of Technology, Lausanne (EPFL), Lausanne, Switzerland, Dec. 2002.

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

[89] İ. E. Telatar, "Capacity of multiantenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[90] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[91] M. Debbah and R. R. Müller, "MIMO channel modeling and the principle of maximum entropy," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1667–1690, May 2005.

[92] J. A. Thomas, "Feedback can at most double Gaussian multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 5, pp. 711–716, Sep. 1987.

[93] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1977.

[94] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inform. Contr.*, vol. 38, pp. 60–80, Jul. 1978.

[95] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1912–1923, Nov. 1997.

[96] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[97] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[98] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

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

[100] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Probab.*, vol. 2, pp. 805–814, Oct. 1974.

المنارة للاستشارات
www.manaraa.com