Information Theoretical Limits on Cooperative Communications

Melda Yuksel and Elza Erkip

Abstract

This chapter provides an overview of the information theoretic foundations of cooperative communications. Earlier information theoretic achievements as well as the more recent developments are discussed. The analysis accounts for full/half-duplex nodes, and for multiple relays. Various channel models such as discrete memoryless, additive white Gaussian noise (AWGN) and fading channels are considered. Cooperative communication protocols are investigated using capacity, diversity and diversity-multiplexing tradeoff (DMT) as performance metrics. Overall, this chapter provides a comprehensive view on the foundations of, and the state-of-the-art reached in the theory of cooperative communications.

Index Terms

compress-and-forward, cooperation, decode-and-forward, diversity-multiplexing tradeoff, fading channels, full-duplex relay, half-duplex relay, multiple-access channel with generalized feedback, multiple-input multiple-output (MIMO), relay channel, wireless networks.

I. INTRODUCTION

In traditional communication networks data transmission directly occurs between the transmitter and the receiver. No user solicits the assistance of another one. However, in

M. Yuksel is with the Electrical and Electronics Engineering Department, TOBB University of Economics and Technology, Ankara, Turkey (e-mail: yuksel@etu.edu.tr).

E. Erkip is with the Electrical and Computer Engineering Department, Polytechnic Institute of New York University, Brooklyn, NY 11201 USA (e-mail: elza@poly.edu)

a general communication network, there are many intermediate nodes that are available to help. For example in wireless networks, when one node broadcasts its messages, all nearby nodes overhear this transmission. Processing and forwarding these messages to the intended destination, system performance, whether it be throughput, lifetime, or coverage area, can be improved. To understand how much performance improvement can ideally be possible by this "cooperative" network, we need an information theoretical study. Such a study also elucidates how cooperation should take place and helps construct the backbone for future cooperative communication applications.

In information theory the idea of cooperation was first presented in van der Meulen's 1971 paper [1], which established the foundations of the *relay channel*. The relay channel is a three-terminal network, in which Terminal 1 (source) aims to transmit to Terminal 3 (destination) with the help of Terminal 2 (relay) as in Fig. 1. The aim is to attain the highest communication rate from Terminal 1 to Terminal 3. In general the intermediate relay enhances communication rates of the direct link from Terminal 1 to 3.

To elaborate the idea of enhanced communication rates via relaying, here we introduce an example from [1]. Table I gives the channel output probabilities $P(y_2, y_3|x_1, x_2)$ conditioned on input pairs (x_1, x_2) when the destination is assumed to be silent ($x_3 = 0$).

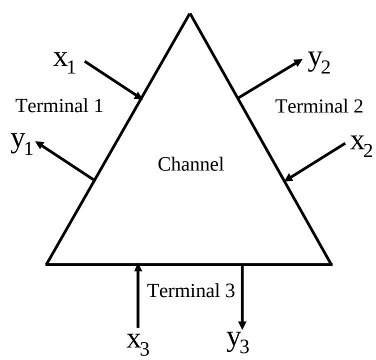

Fig. 1. The three-terminal communication channel.

Observe that if $x_2 = 1$, both y_2 and y_3 are equal to 1 no matter what x_1 is. Neither the relay, nor the destination can distinguish the correct x_1 and hence, communication rates between the source and the relay, and the source and the destination are equal to zero given $x_2 = 1$. When $x_2 = 0$, the channels between the source and the relay, and the source and the destination both become equivalent to a binary symmetric channel with crossover probability $1/2$, whose capacity is also equal to zero. Hence, no direct communication is possible from the source to the relay, or from the source to the destination. However, non-zero communication rates from the source to the destination can be achieved with the help of the relay. Observe that if $x_2 = 0$ and the destination knows the channel output y_2, then the source can send 1 bit noiselessly to the destination. On the other hand, the destination can learn about y_2, if the relay helps the source and transmits its output signal after its own observation is complete. This information can be sent at rate 0.32 bits/channel use when x_1 is set to 0. By appropriate time division between these two strategies, [1] proves that the capacity of this example channel is equal to 0.243 bits/channel use with vanishing error probability. As this example clearly shows, relaying can substantially increase achievable rates with respect to direct transmission.

The relay channel is essential to both wired and wireless networks. In wired networks many source and destination pairs are connected via intermediate relay nodes. In

x_1x_2 \ y_2y_3	00	10	01	11
00	$1/2$	0	0	$1/2$
10	0	$1/2$	$1/2$	0
01	0	0	0	1
11	0	0	0	1

TABLE I

A RELAYING EXAMPLE FROM [1]

wireless networks, due to the *wireless broadcast advantage* idle nodes overhear nearby transmissions. These nodes can relay the information and contribute to achieving higher rates.

The capacity of the relay channel was calculated in [1] only for some particular example channels. Guidelines for communication/relaying principles in a general relay channel were developed later in the landmark paper by Cover and El Gamal [2]. In [2], the memoryless relay channel is investigated and two fundamental relaying techniques; decode-and-forward (DF) and compress-and-forward (CF), are presented. The paper also includes capacity theorems for both degraded and reversely degraded [1] relay channels and the relay channel with feedback.

Despite the substantial advancement [2] provided, the capacity of the general relay channel has been unsolved for over thirty years. The recent activity on cooperative communications mostly stems for the potential wireless applications and is spurred by recent papers [3], [4] and [5]. In [3] and [4], motivated by Willems' multiple-access channel with generalized feedback (GMAC) model [6], the authors considered the benefits of mutual cooperation in fading environments and suggested how cooperation can be carried out in code-division multiple-access (CDMA) systems. In [5] the authors suggested *simple* yet high performance relaying protocols. Promising significant gains and having vast application areas, the ideas presented in these papers triggered an extensive literature on cooperative communication systems.

Later, Kramer, Gastpar and Gupta generalized the DF and CF protocols suggested in [2] to arbitrary number of relay nodes [7]. Even though the capacity of the relay channel is unknown, this paper proves that the DF and CF protocols can be capacity achieving depending on the location of the relay node(s). Recent results on multiple relay terminal

[1]In the physically degraded relay channel the destination's observation is a physically degraded version of the relay's observation. It is the opposite for the reversely degraded relay channel, where the relay's signal is physically degraded with respect to the destination's.

networks suggested strategies that are a fixed number of bits away from the capacity [8], [9].

An important constraint in the relay channel is the processing capability of the relay node. The relay can either be full or half-duplex. If the relay is full-duplex it can transmit and receive simultaneously in the same frequency band. In half-duplex systems, transmission and reception takes place in orthogonal channels. Although the full-duplex assumption is not practically feasible, it helps us understand the fundamental characteristics of the relay channel. On the other hand, the half-duplex assumption is required to study practical aspects. Half-duplex operation in the relay channel was first considered [5]. Later [10] and [11] investigated the capacity of the relay channel under the half-duplex assumption.

The above mentioned papers consider probability of error and achievable rates to measure the level of reliability and the throughput of relaying/cooperation schemes. Another performance measure is the diversity-multiplexing tradeoff (DMT) [12], which unifies the reliability and rate perspectives. The DMT is a high SNR analysis suitable for fading channels and establishes the fundamental tradeoff between diversity and multiplexing for multiple-input multiple-output (MIMO) systems. DMT is also a useful performance measure for cooperative/relay systems. While the capacity of the relay channel is not known in general, it is possible to find relaying schemes that are optimal from the DMT perspective. In the literature [5], [13], [14], [15] were the first to investigate the relay channel DMT. In these works, either the source or the destination (or both) has a single antenna. The multiple-antenna multiple-relay channel was first studied in [16]. In [16] DMT upper bounds are found, and a relaying strategy, which achieves the bound for both full and half-duplex relays, is suggested.

This chapter surveys the above mentioned models for cooperation as well as how cooperation is useful under achievable rate, diversity and DMT metrics. It is important to note that there has been a large body of information theoretical work on cooperative communications in the past few years [17], [18] and this chapter only covers some of the fundamentals. For example, we consider neither the ergodic capacity, nor the resource allocation problems. Similarly, scaling laws in large cooperative networks and cooperative channel coding are out of this chapter's scope.

In Section II we first introduce the DF and CF protocols and introduce the generalized multiple access channel model. In Section III we extend these basic models to multiple relay/ cooperating nodes. In Section IV we explain the effect of relay processing constraints on achievable rates. In Section V we examine the relay channel when there is fading. Next, in Section VI we provide the DMT analysis for the relay channel. Finally, in Section VII we conclude.

II. The Basic Models for Cooperation

In this section we introduce the full-duplex relay channel to capture unilateral cooperation and describe the two fundamental relaying protocols DF and CF. We then explain the G-MAC model to explain two-way cooperation.

A. The Relay Channel

The basic relay channel model is illustrated in Fig. 2, where $p(y_R, y_D|x_s, x_R)$ indicates the discrete memoryless channel, W denotes the message, X_S and X_R are the signals the source and relay transmit, Y_R and Y_D are the received signals at the relay and at the destination, and $\hat{W}$ is the destination's estimate of W.

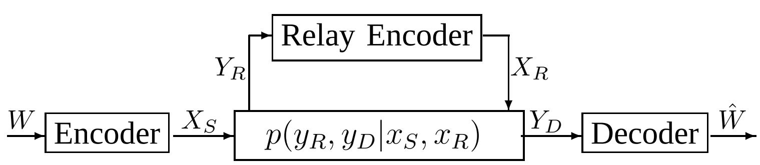

Fig. 2. The relay channel.

If the channel is real additive white Gaussian noise (AWGN), then we can write

$$Y_R = a_{SR} X_S + Z_R \tag{1}$$

$$Y_D = a_{SD} X_S + a_{RD} X_R + Z_D, \tag{2}$$

where a_{SR}, a_{SD} and a_{RD} ($\in \mathbb{R}$) are respectively the channel gains between the source and the relay, the source and the destination, and the relay and the destination. The AWGN at the relay and destination are respectively denoted by Z_R and Z_D, which are assumed to be zero mean and unit variance. Both the source and the relay have individual power constraints P_S and P_R. The objective is to find the capacity, the maximum achievable rate beyond which reliable communication is not possible. Note that the cut-set upper bound [19] suggests that if R is an the achievable rate, then

$$R \leq \max_{p(x_S, x_R)} \min\{I(X_S; Y_R, Y_D); I(X_S, X_R; Y_D)\},$$

which becomes

$$\begin{aligned} R \leq & \max_{\rho \in [0,1]} \min \left\{ \frac{1}{2} \log \left(1 + \rho a_{SR}^2 P_S + \rho a_{SD}^2 P_S \right), \right. \\ & \left. \frac{1}{2} \log \left(1 + a_{SD}^2 P_S + a_{RD}^2 P_R + 2\sqrt{(1-\rho) a_{SD}^2 P_S a_{RD}^2 P_R} \right) \right\} \end{aligned} \tag{3}$$

for the Gaussian case.

DF and CF are two of the relaying schemes proposed in [2] for the general relay channel. In DF, the relay decodes its received message, re-encodes it, and forwards it to the destination. In CF, the relay first compresses its received signal and then forwards the compressed signal through the relay-destination channel. The compression is Wyner-Ziv type [20]; i.e. the relay compresses its received signal taking into account that the destination has side information available directly from the source. Next, we introduce the ideas behind these two protocols. The formal proofs can be found in [2].

Block 1	Block 2	Block 3		Block B
$x_S^N(1, w_1)$	$x_S^N(w_1, w_2)$	$x_S^N(w_2, w_3)$	...	$x_S^N(w_{B-1}, 1)$
$x_{R_1}^N(1)$	$x_{R_1}^N(w'_1)$	$x_{R_1}^N(w'_2)$		$x_{R_1}^N(w'_{B-1})$

Fig. 3. The decode-and-forward block structure.

1) Decode-and-forward: In the DF protocol, the source, and the relay perform block Markov superposition coding. The destination can do sliding window or backward decoding [21], [7], [22], [23]. We will briefly describe the backward decoding here. The encoding structure is depicted in Fig. 3. Transmission takes place in B blocks, where each block consists of N symbols, with N and B both large. The source transmits a length-N codeword $x_S^N(w_{b-1}, w_b)$ in each block b, for which $w_0 = w_B = 1$, where w_b denotes the new message transmitted in block b.

The source rate is chosen so that the relay can reliably decode. To satisfy this constraint the source rate $R^{(DF)}$ should satisfy

$$R^{(DF)} \leq I(X_S; Y_R | X_R) \tag{4}$$

for a fixed input distribution. Under this condition, the relay finds an estimate w'_b for w_b at the end of block b and sends $x_R^N(w'_b)$ in block $b+1$. In the first block the relay sends a predetermined codeword $x_R^N(1)$. With very high probability $w'_b = w_b$ and the relay decodes the source message reliably. This way, the relay removes the effects of the channel from its received signal, and thus obtains a *clean* copy of the original source message. As the source repeats w_b in block $b+1$, the relay can then act collectively with the source. Note that the full-duplex assumption is critical to realize the block Markov encoding structure, which requires the relay to transmit and receive simultaneously.

The destination starts decoding after all B blocks are received and moves backwards [7], [22], [23]. At block B, no fresh information is sent and the destination is interested in decoding w_{B-1}. The correct message index can be identified with high probability if

$$R^{(DF)} \leq I(X_S, X_R; Y_D). \tag{5}$$

Once the destination decodes w_{B-1}, it can move backwards to decode $w_{B-2}, ..., w_1$ in a similar manner. Combining both constraints (4) and (5), we conclude that for a fixed input distribution $p(x_S, x_R)$

$$R^{(DF)} \leq \min\{I(X_S; Y_R|X_R), I(X_S, X_R; Y_D)\} \tag{6}$$

is achievable. Maximizing over all input distributions, we can write the maximum rate the DF protocol achieves is

$$R^{(DF)}_{max} = \max_{p(x_S, x_R)} \min\{I(X_S; Y_R|X_R), I(X_S, X_R; Y_D)\}.$$

For the Gaussian relay channel using Gaussian code books $R^{(DF)}_{max}$ becomes

$$\begin{aligned} R^{(DF)}_{max} &= \max_{\rho \in [0,1]} \min \left\{ \frac{1}{2} \log\left(1 + \rho a^2_{SR} P_S\right), \right. \\ &\left. \frac{1}{2} \log \left(1 + a^2_{SD} P_S + a^2_{RD} P_R + 2\sqrt{(1-\rho) a^2_{SD} P_S a^2_{RD} P_R} \right) \right\}. \end{aligned} \tag{7}$$

The first term in the minimization accounts for the rate the relay can reliably decode. The second term is the rate, which the destination can decode, when the source and the relay transmit together. The parameter ρ denotes the correlation between source and relay signals and is related to the coherent combining gain at the destination. As the source and the relay are synchronized, their signals add coherently at the destination. This leads to the additional gain in the received signal-to-noise ratio $2\sqrt{(1-\rho) a^2_{SD} P_S a^2_{RD} P_R}$. Changing ρ the source node divides its power among sending *fresh* information and repeating *old* information to coherently add with the relay's signal.

It is interesting to note that the DF protocol achieves the capacity, when the relay channel is physically degraded [2]. Physical degradedness ensures that decoding at the relay does not impose an additional constraint on the system, and DF becomes optimal.

When the relay channel is not physically degraded, then decoding at the relay may impose strict constraints; i.e. the source-relay channel can limit the achievable rates. In such cases partial decode-and-forward (PDF) can be used instead of DF. As its name suggests, in PDF the relay is required to decode part of the source message only, and the remaining part is directly sent to the destination without the relay's help. This strategy improves upon DF achievable rates [2]. We will discuss the PDF strategy more in Section II-B.

2) Compress-and-forward: In the DF protocol the relay performs a *hard decision* about the source message. Forcing the relay to make a decision can incur losses in achievable transmission rates. In some cases compression based protocols, which forward relay's *soft* signal are desirable.

A simple B-block compression protocol, which we call simple compress-and-forward (SCF), is as follows: The relay compresses its received signal Y_R in block b to form $\hat{Y}_R$, maps this to the channel codeword X_R and sends X_R to the destination in block $b+1$. As the relay is full duplex, the relay can listen and compress the message in the current block while transmitting the compressed signal of the previous block. The received signal Y_D in block b is a function of both the source and the relay signals X_S and X_R transmitted in the same block.

The destination starts decoding after all B blocks are transmitted. It first decodes the relay signal X_R, treating the source signal as noise and recovers $\hat{Y}_R$. The reliability of these steps are ensured, if the relay's compression rate is below the relay-to-destination achievable rate considering the source signal as noise. In other words, the condition

$$I(\hat{Y}_R; Y_R | X_R) \leq I(X_R; Y_D), \tag{8}$$

has to be satisfied for a fixed input distribution $p(x_S)p(x_R)p(\hat{y}_R|x_R, y_R)p(y_R, y_D|x_S, x_R)$. After the destination decodes $\hat{Y}_R$, the destination uses both $\hat{Y}_R$ and Y_D from the previous block to determine the original source signal X_S. This can be done reliably if

$$R^{(SCF)} < I(X_S; \hat{Y}_R, Y_D | X_R). \tag{9}$$

The performance of the above scheme could be improved if Wyner-Ziv type compression [20] is used instead of simple compression. The Wyner-Ziv technique allows for lower compression rates in the presence of correlated side information at the decoder. This is indeed the case in the relay channel, in which the direct signal from the source received at the destination in the previous block can be thought of as the side information. This side information lowers the compression rate from $I(\hat{Y}_R; Y_R | X_R)$ to $I(\hat{Y}_R; Y_R | X_R, Y_D)$ leading to the condition

$$I(\hat{Y}_R; Y_R | X_R, Y_D) \leq I(X_R; Y_D). \tag{10}$$

Other decoding steps remain the same and the overall CF achievable rate is

$$R^{(CF)} = I(X_S; \hat{Y}_R Y_D | X_R) \tag{11}$$

subject to (10) where the joint probability distribution is $p(x_S)p(x_R)p(\hat{y}_R|x_R, y_R)p(y_R, y_D|x_S, x_R)$. Since (10) is looser than (8), achievable rates are potentially higher when Wyner-Ziv type compression is employed.

For the Gaussian case, $R^{(CF)}$ becomes equal to

$$R^{(CF)} = \frac{1}{2} \log \left(1 + a_{SD}^2 P_S + \frac{a_{SR}^2 P_S}{1 + \hat{N}_R} \right) \tag{12}$$

subject to

$$\hat{N}_R \geq \frac{1 + a_{SD}^2 P_S + a_{SR}^2 P_S}{a_{RD}^2 P_R}. \tag{13}$$

Note that two terms, $a_{SD}^2 P_S$ and $a_{SR}^2 P_S / (1 + \hat{N}_R)$, contribute to the received SNR at the destination. The former is the received SNR due the direct source-destination link. The latter is the received SNR at the relay except the additional $\hat{N}_R$ in the denominator. This shows that the destination receives the relay's observation with an additional compression noise.

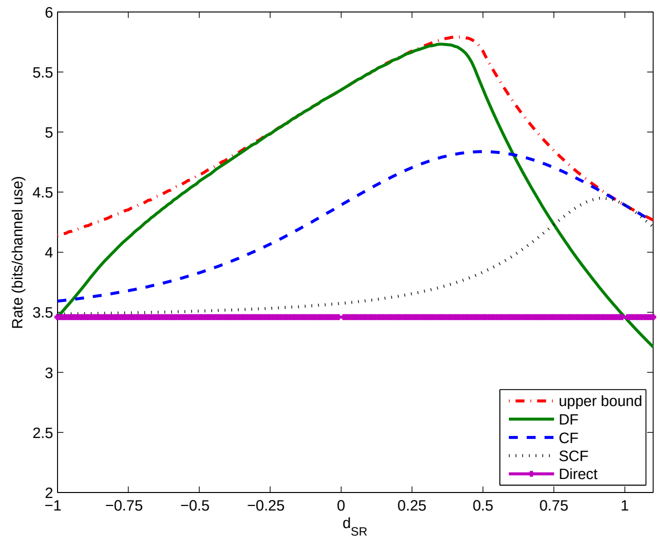

Fig. 4. DF, CF and SCF achievable rates in comparison to the upper bound and direct transmission [7].

Kramer, Gastpar and Gupta [7] provide an extensive analysis of the achievable rates of DF and CF for general channel models. One of the models considered is the path loss model, which follows the Gaussian channel in (1)-(2) with $a_{ij} = 1/\sqrt{d_{ij}^{\alpha}}$, $i, j = S, R, D$, $i \neq j$, where α indicates the path loss exponent, and d_{ij} indicates the inter-node distances.

In Fig. 4, we plot the DF and CF rates of (7) and (12) for $\alpha = 2$, and $P_S = P_R = 10$. The source and the destination are respectively located at 0 and 1. The relay's location varies on the line joining the source and the destination. The DF and CF rates are also compared to the upper bound of (3) as well as the direct transmission rates. When d_{SR} distance is small, the first terms in (3) and (7) become very large. The second term dominates both equations and DF becomes optimal. Similarly, when d_{RD} distance is small, the compression noise in (13) approaches zero and the CF rate in (7) becomes equal to the upper bound in (3). The figure clearly confirms that the DF protocol achieves the capacity when the relay is close to the source, and the CF protocol achieves the capacity when the relay is close to the destination.

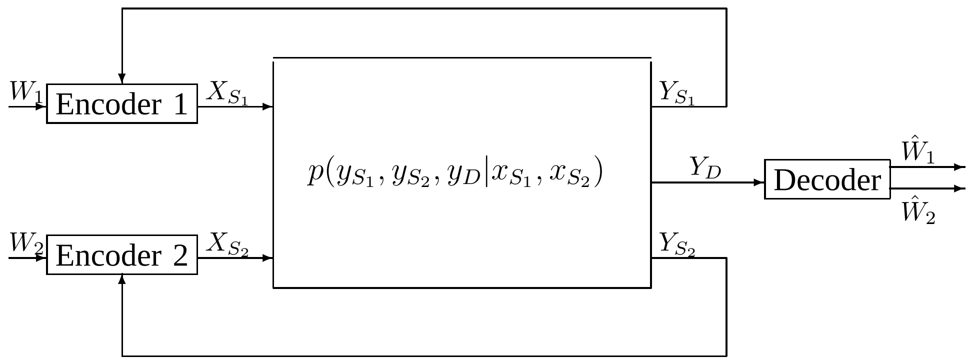

Fig. 5. The multiple access channel with generalized feedback.

B. Multiple-Access Channel with Generalized Feedback

The relay channel model describes the basic form of cooperation. In the relay channel, the relay node does not have its own information to send; it only assists the source node to enhance achievable communication rates. On the other hand, the full cooperation model captured in the multiple access channel with generalized feedback considers two or more sources, which mutually help each other to attain better performance.

The two-user discrete-memoryless GMAC is shown in Fig. 5, where W_1 and W_2 denote the messages of the first and second sources, X_{S_1} and X_{S_2} are the source signals, $p(y_{S_1}, y_{S_2}, y_D|x_{S_1}, x_{S_2})$ is the channel, Y_{S_1}, Y_{S_2} and Y_D are respectively the received signals at S_1, S_2 and the destination, and $\hat{W}_1$ and $\hat{W}_2$ are the destination's estimate of W_1 and W_2. This generalized feedback model allows for both users to *overhear* each other and thus mutual cooperation can take place. In the GMAC, the relay is another source node, which has its own information to send. The achievable rate region, which is the collection of all achievable rate pairs. This model reduces to the relay channel model, if $W_2 = \emptyset$ and $Y_{S_1} = \emptyset$.

The achievable rate region suggested in [6] for the GMAC is PDF based. In PDF some part of the message is directly sent to the destination without the help of the other. Similar to the DF in Section II-A.1 communication lasts for B blocks. In each block, each source's messages are considered to be composed of three parts. For the first part neither of the sources seek the other's assistance. This part of the message is sent directly to the destination. The second part of S_1's message in block b is aimed to be decoded only at S_2 at the end of block b. Decoding this part of the message, S_2 forms the third part of its message in block $b+1$. Similarly, at the end of block b, S_1 decodes the second part of S_2's message, and re-encoding it forms the third part of its own message in block $b+1$. This way, each source node can help each other simultaneously. We refer the readers to [6] and [24] for the achievable rate region characterization.

Although the GMAC results existed for more than 20 years, their interpretation as a means for mutual cooperation was not until [3], [4]. We will mention these results in Section V within the context of cooperation in wireless communications.

III. Extensions to Multiple Relays

In Section II, we considered the relay channel and GMAC, which consist of only two transmitters. In a general multiple-terminal network more than one relay is available to help or multiple users form coalitions. Communication strategies and achievable rates for the multiple relay channel is another important question, which was considered in [21], [7]. The paper [21] generalizes the DF protocol, and [7] generalizes the CF protocol to multiple relays, which we explain next.

When we described the DF protocol in Section II-A.1, we observed that $R^{(DF)}$ is the minimum of two rates, the source-relay communication rate, and the rate the source and relay collectively transmit to the destination. When there are multiple relays and all relays perform DF, this idea can be generalized. Suppose there are two relays. Then the DF achievable rate is the minimum of three rates: the rate source can reliably send to the first relay, the rate source and the first relay can reliably send to the second relay, and the rate the source and both relays altogether can send to the destination. Of course, this depends on the relay assignment; i.e. which relay will be the first and which relay will be the second. Therefore, the minimum of the above mentioned rates can be maximized over all relay permutations. If there are M relays, then the DF achievable rate is the minimum of $M+1$ rates maximized over all relay permutations.

Generalizing the CF protocol to multiple relays is more intricate. When there are multiple relays, the relays observe correlated signals. Forwarding by the relays becomes equivalent to sending arbitrarily correlated sources over a multiple access channel, whose complete solution is not known [25]. Moreover, the destination has side information available in addition to the signals it receives from the relays.

In the case of multiple relays, all relays do not have to perform the same relaying strategy. When some relays do DF and the rest do CF an achievable rate region is given in [7]. Here, we explain this scenario for the two-relay case, where the first relay does DF and the second does CF after listening to the first relay. Then the achievable rate becomes

$$R^{(DCF)} = \min\left\{ I(X_S; Y_{R_1}|X_{R_1}), I(X_S, X_{R_1}; \hat{Y}_{R_2}, Y_D|X_{R_2}) \right\} \tag{14}$$

subject to

$$I(\hat{Y}_{R_2}; Y_{R_2}|X_{R_2}, Y_D) \leq I(X_{R_2}; Y_D), \tag{15}$$

for the joint distribution $p(x_S, x_{R_1})p(x_{R_2})p(\hat{y}_{R_2}|x_{R_2}, y_{R_2})p(y_{R_1}, y_{R_2}, y_D|x_S, x_{R_1}, x_{R_2})$. The first mutual information expression in (14) indicates the rate the first relay can decode. If the first relay can decode, then the source and the first relay together act as a joint transmitter. Via the second relay the destination obtains a second observation $\hat{Y}_{R_2}$ in addition to its own Y_D. Therefore, the second relay and the destination mimic a joint receiver. Overall, the second term in (14) mimics the 2×2 multiple-antenna system. In a path loss model as in Section II-A, this mixed strategy is capacity achieving when the first relay is in close proximity with the source and the second relay is in close proximity with the destination [7].

The multiple relay channel problem has also been recently considered in [9]. In this work, inspired from a deterministic model [8], the authors find a lower bound on the capacity of Gaussian relay networks that is within a constant number of bits away from the cut-set upper bound. This constant depends on the topology of the network but not on the channel parameters. This result provides a good approximation of the capacity, which becomes tight as signal-to-noise ratio increases.

IV. Half-duplex

In Sections II and III, we had the idealized assumption that the relay is full-duplex. However, full-duplex operation is not possible in practical applications. Wireless transceivers cannot transmit and receive at the same time in the same band. In this section, we consider a half-duplex relay and study the impact of half-duplex operation on relaying protocols and achievable rates.

We can model half-duplex operation using the state variable Q that controls the relay operation. Q takes the value q_1 if the relay is listening, and q_2 if the relay is transmitting. A more general state configuration assumes that the relay can be in sleep, listen or talk states [11], the fundamental idea in both configurations being the same. In this chapter we consider the former for simplicity. We also consider fixed protocols, in which the relay listens for a fixed time interval fraction (t, with $0 \leq t \leq 1$) and then transmits in the remaining portion $(1 - t)$, Fig. 6. The relay does not aim to pass along additional information by breaking its transmission and reception intervals into smaller blocks and controlling its state variable [16].

The cut-set upper bound on achievable rates for the half-duplex relay channel is found in [26]. We will state this upper bound for the multiple-antenna half-duplex relay channel in Section VI-B.

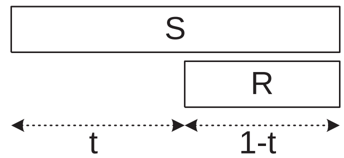

Fig. 6. Fixed relaying: the relay listens for t fraction of the time and transmits in the remaining $1 - t$ fraction.

For the half-duplex relay channel, the DF achievable rate extends from (6), if the half-duplex constraint is taken into account. When the relay is half-duplex and employs DF

$$R^{(DF)} = \min\{tI(X_S; Y_R|q_1) + (1-t)I(X_S; Y_D|X_R, q_2), \tag{16}$$

$$tI(X_S; Y_D|q_1) + (1-t)I(X_S, X_R; Y_D|q_2)\} \tag{17}$$

is achievable for a fixed input distribution and a fixed t, can be maximized over all input distributions and $t \in [0, 1]$. The first term in (17) is the sum of two mutual information expressions. The first one indicates the amount of information the relay can decode in t fraction of the time. The second is the mutual information the destination collects from the source during the time the relay is transmitting. The second term in (17) is also a sum of two terms, the first of which is the mutual information at the destination while the relay is silent, and the second is the rate the source and the relay can together send in $1-t$ fraction of the time. Similar to the full-duplex case, if the half-duplex relay channel is physically degraded, then the above rate is capacity achieving.

When the relay employs CF, the Wyner-Ziv type compression rate is such that the compressed signal at the relay can reach the destination error-free in the remaining $(1-t)$ fraction of time, in which the relay transmits. Then, for a fixed t the instantaneous mutual information at the destination is

$$R^{(CF)} = tI(X_S; \hat{Y}_R, Y_D|q_1) + (1-t)I(X_S; Y_D|X_R, q_2)$$

subject to

$$tI(\hat{Y}_R; Y_R|Y_D, q_1) \leq (1-t)I(X_R; Y_D|q_2). \tag{18}$$

Note that the above equations incorporate the half-duplex constraint into (11) and (10) and can be maximized over $t \in [0, 1]$.

The real Gaussian noise half-duplex relay channel is very similar to the full-duplex case of (1)-(2). However, in half-duplex case the relay cannot transmit and receive at the same time. Then we can write

$$\begin{aligned} Y_{R,1} &= a_{SR}X_{S,1} + Z_{R,1} \\ Y_{D,1} &= a_{SD}X_{S,1} + Z_{D,1}, \end{aligned}$$

for state q_1, and

$$Y_{D,2} = a_{SD}X_{S,2} + a_{RD}X_{R,2} + Z_{D,2},$$

for state q_2. The channel gains a_{SR}, a_{SD} and a_{RD} are defined similarly as in Section II.

For the Gaussian half-duplex relay channel the DF achievable rate becomes [26]

$$\begin{aligned} R^{(DF)}_{max} &= \max_{\rho, t \in [0,1]} \min \Bigg\{ \frac{t}{2} \log\left(1 + a_{SR}^2 P_{S,1}\right) + \frac{1-t}{2} \log\left(1 + \rho a_{SD}^2 P_{S,2}\right), \\ &\quad \frac{t}{2} \log\left(1 + a_{SD}^2 P_{S,1}\right) \\ &\quad + \frac{1-t}{2} \log\left(1 + a_{SD}^2 P_{S,2} + a_{RD}^2 P_R + 2\sqrt{(1-\rho) a_{SD}^2 P_{S,2} a_{RD}^2 P_R}\right) \Bigg\}, \end{aligned}$$

Similarly, the half-duplex CF achievable rates are [27]

$$R^{(CF)} = \max_{t \in [0,1], P_{S,1}, P_{S,2}} \frac{t}{2} \log\left(1 + a_{SD}^2 P_{S,1} + \frac{a_{SR}^2 P_{S,1}}{1 + \hat{N}_R}\right) + \frac{1-t}{2} \log\left(1 + a_{SD}^2 P_{S,2}\right),$$

where

$$\hat{N}_R = \frac{1 + (a_{SD}^2 + a_{SR}^2)P}{(1 + a_{SD}^2 P)\left(\left(1 + \frac{a_{RD}^2 P_R}{1 + a_{SD}^2 P_{S,2}}\right)^{\frac{1-t}{t}} - 1\right)}.$$

In the above expressions $P_{S,1}$ and $P_{S,2}$ are the average source power constraints in states q_1 and q_2 respectively with $P_{S,1}+P_{S,2}=P$. Both rates $R^{(DF)}$ and $R^{(CF)}$ can be optimized over $P_{S,1}$ and $P_{S,2}$ as well.

In addition to DF and CF based protocols, amplify-forward (AF) [5] and non-orthogonal amplify-and-forward (NAF) [28], [13] are two linear relaying protocols that are important for the Gaussian half-duplex relay channel. In AF the source and the relay share the time equally, and the source node remains silent while the relay transmits in the second half of the time. In AF the relay simply scales $Y_{R,1}$ according to its own power constraint and forwards $X_R = \beta Y_R$ to the destination, where $\beta \leq \sqrt{P_R/(a_{SR}^2 P_S + 1)}$. Assuming β is equal to its upper bound, the AF protocol achieves the rate

$$R^{(AF)} = \frac{1}{4}\log\left(1 + a_{SD}^2 P_S + \frac{a_{SR}^2 P_S a_{RD}^2 P_R}{1 + a_{SR}^2 P_S + a_{RD}^2 P_R}\right).$$

The NAF protocol is an extension of AF, where the source can simultaneously transmit with the relay in state q_2.

V. WIRELESS APPLICATIONS: COOPERATIVE DIVERSITY

In the previous sections we have considered cooperation in discrete memoryless and Gaussian channels. One of the most important advancements in cooperative communication has been the unearthing of its potential in wireless networks.

Wireless channels experience fading, which degrades the system performance when signal components that are received over different propagation paths add destructively [29]. Most of the error correcting codes can recover from even very deep fade levels, if the fading coherence time is on the order of a symbol time. But if the fading coherence time is much longer than the symbol duration, then a deep fade affects many symbols consecutively. Retransmission or forward error correction are too costly in this case and no recovery is possible. These slowly fading channels do not guarantee reliable communication for any transmission rate and have zero capacity. For these channels, error probability is due to two main causes: deep fade levels, and channel noise. When the wireless channel experiences a deep fade and the channel cannot support the transmission rate, it is said that the channel is in outage [30]. The probability of error can then be written as

$$P(\text{error}) = P(\text{error}|\text{outage})P(\text{outage}) + P(\text{error}|\text{no outage})P(\text{no outage}).$$

If the channel is in outage, then the error probability is almost equal to 1. Similarly, if there is no outage, good channel codes ensure arbitrarily small error probability. Thus the outage event dominates the error event and probability of error is approximately equal to probability of outage.

To mitigate the adverse effects of fading, time, frequency or spatial diversity techniques are employed. For example when there are multiple antennas at the receiver, each antenna observes an independent copy of the transmitted message. It is less likely that all observations are bad, and thus reliability is increased. Similarly, when there are multiple transmit antennas, each antenna can suitably repeat the same message to increase reliability. In addition to these traditional diversity techniques, cooperation/relaying can also be used to provide diversity [3], [4], [5].

In this section, we first assume the relay is full-duplex. Under fading, the received signals at the relay and at the destination are given by (1) and (2), but now the channel inputs, outputs and the channel noise are complex valued. Assuming Rayleigh fading, $a_{ij} = h_{ij}$, $i, j = S, R, D$. $i \neq j$, are independent, identically distributed (i.i.d.) zero mean complex Gaussian random variables with zero mean and unit variance. We consider slow fading, which stays constant for the duration of a channel block. There is no channel state information at the source. We assume there are pilot signals to measure receiver channel gains, and thus the relay and the destination know their incoming fading levels. In addition, the relay knows its outgoing gain and the source-destination channel gain, which can happen at a negligible cost with proper feedback. This information becomes necessary to implement the CF protocol and will become clear when we explain CF later in this section and in Section VI-A.

As the source node does not have channel side information, and the application is delay-limited and constant rate, the source node transmits at a fixed target data rate $R^{(T)}$. An outage occurs if the mutual information at the destination is not large enough to support this fixed target rate, and we can write the outage probability as

$$P(\text{outage}) = P(I < R^{(T)}),$$

where I is the mutual information at the destination when receiver side channel state information is available.

Under fading the DF protocol is slightly different than the one described in Section II-A.1 for non-fading channels. When there is fading, the mutual information at the destination depends on whether the relay decodes the source message or not [5]. The relay can reliably decode the source message and repeat it to the destination if the mutual information it collects $I(X_S; Y_R) = I_R$

$$I_R = \log(1 + |h_{SR}|^2 P_S)$$

is larger than the target data rate. If not, we assume the relay remains silent. Let

$$\begin{aligned} I_{SD} &= \log(1 + |h_{SD}|^2 P_S) \\ I_{SR,D} &= \log(1 + |h_{SD}|^2 P_S + |h_{RD}|^2 P_R). \end{aligned}$$

Then the mutual information at the destination is equal to

$$I^{(DF)} = \begin{cases} I_{SD} & \text{if the relay cannot decode} \\ I_{SR,D} & \text{if the relay decodes} \end{cases} .$$

[2] Note that unlike the Gaussian case, there is no coherent combining gain between the source and the relay, when they transmit together. This is because the source node does not have channel state information and cannot synchronize its phase with the relay.

[2]Note that this DF protocol is different than the one defined in [5], in which the signal received from the source is completely ignored if the relay cannot decode the message.

Then we can write the probability of outage at the destination as

$$\begin{aligned} P(\text{outage at D}) &= P(\text{outage}|\text{relay decodes})P(\text{relay decodes}) \\ &\quad + P(\text{outage}|\text{relay cannot decode})P(\text{relay cannot decode}), \end{aligned} \tag{19}$$

which becomes equal to

$$\begin{aligned} P(\text{outage at D}) &= P(I_{SR,D} < R^{(T)}|I_R > R^{(T)})P(I_R > R^{(T)}) \\ &\quad + P(I_{SD} < R^{(T)}|I_R < R^{(T)})P(I_R < R^{(T)}). \end{aligned} \tag{20}$$

When the relay employs CF then the mutual information at the destination is same as in (12) if a_{ij}^2 are replaced with $|h_{ij}|^2$, and the $1/2$ is removed. Note that the relay has to know all to channel gains in the system to determine the compression noise in (13). The probability of outage for CF is then $P(I^{(CF)} < R^{(T)})$.

Fig. 7 shows the outage probability of DF and CF protocols as a function of the target data rate $R^{(T)}$ for $P_1 = P_2 = 10$. Direct transmission and 1×2 MIMO outage probabilities are also included in the figure for comparison. We observe that cooperative protocols significantly improve upon direct transmission.

The above analysis shows the probability of outage as a function of the target data rate for a fixed P_S and P_R. It is also important to understand how probability of outage behaves as a function of P_S and P_R for a fixed target data rate. In particular the probability of error decay rate is of interest, as fast decay implies superior reliability. This decay rate is called the diversity gain and is defined as

$$d = -\lim_{\text{SNR}\to\infty} \frac{\log P_e(\text{SNR})}{\log \text{SNR}}. \tag{21}$$

Here SNR denotes the average received signal-to-noise ratio. For high SNR values (21) can be restated to write

$$P_e(\text{SNR}) \doteq \frac{1}{\text{SNR}^d}.$$

Diversity gains do not change with a constant scaling in the transmit power levels P_S and P_R. In the relay channel, we thus assume $P_S = P_R$ without any loss of generality.

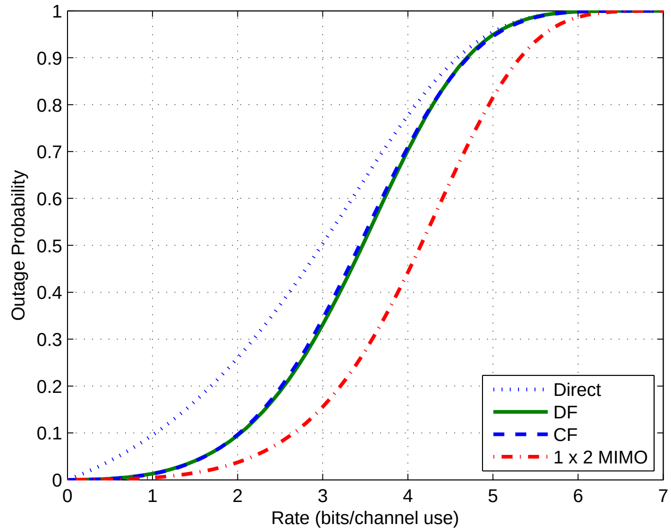

Fig. 7. Probability of outage as a function of rate. $P_1 = P_2 = 10$.

Note that the average transmit power is equal to the average SNR as we assumed the noise variance is equal to 1.

Now, we find the diversity gains for DF and CF protocols. As probability of error dominated by outage, we simply investigate the outage probability decay rate.

For the DF protocol, we first look into the relay outage probability, which is

$$\begin{aligned} P(I_R < R^{(T)}) &= P\left(|h_{SR}|^2 < \frac{2^{R^{(T)}} - 1}{P_S}\right) \\ &= 1 - \exp\left(-\frac{2^{R^{(T)}} - 1}{P_S}\right) \\ &\doteq \frac{1}{\mathrm{SNR}}. \end{aligned}$$

This also implies that $P(I_R > R^{(T)}) \doteq 1$. Note that direct transmission outage probability also results in the same limit as the channel gains are assumed to be identically distributed. Similarly, one can show that $P(\text{outage}|\text{relay cannot decode}) \doteq \mathrm{SNR}^{-1}$ and

$P(\text{outage}|\text{relay can decode}) \dot{=} \text{SNR}^{-2}$. Overall,

$$P(\text{outage at D}) \dot{=} \frac{1}{\text{SNR}^2}$$

and the DF protocol diversity gain is equal to 2.

Similar to the DF protocol, the CF protocol also achieves 2 levels of diversity gain [16]. Intuitively, from a diversity perspective, the CF protocol makes the relay channel equivalent to a single-antenna source, two-antenna destination point-to-point system, while DF results in a two-antenna source, a single-antenna destination system. In the point-to-point system each receiver antenna suffers from additive noise. In the CF protocol it is as if one of the receiver antennas experiences additional noise due to compression. However, the relay can adjust this compression noise to enable full diversity. We omit the details and refer the reader to [16] for a more complicated discussion.

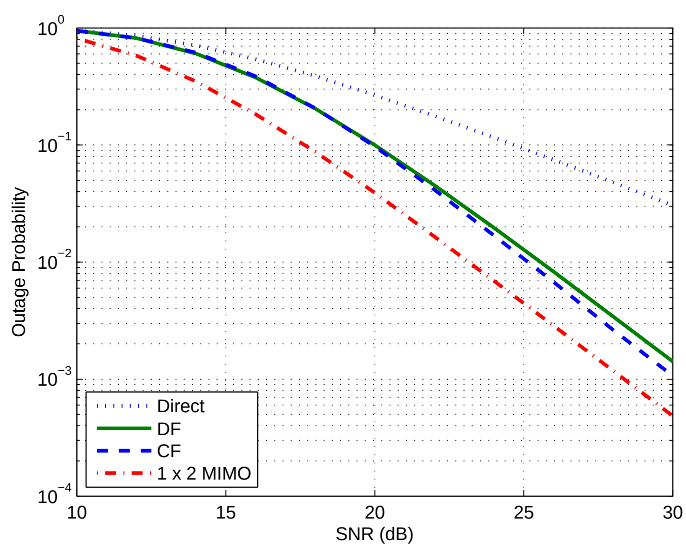

Fig. 8. Probability of outage as a function of SNR. $R^{(T)} = 5$ bits/ channel use.

Fig. 8 displays DF and CF outage probabilities as a function of per user SNR $P_S = P_R$ for $R^{(T)} = 5$ bits/channel use. The figure clearly shows that DF and CF have 2 levels of diversity as in 1×2 MIMO, whereas direct transmission has only 1 levels of diversity. This result implies that DF and CF protocols can be used to form virtual antenna arrays and to provide diversity gains similar to spacial diversity techniques.

It is important to note that half-duplex DF and CF and the AF and NAF protocols introduced in Section IV also achieve 2 levels of diversity [5], [28]. This is because the source-destination and the relay-destination links observe independent channel gains, and the diversity gain mainly depends on the number of independent observations at the destination even when half-duplex constraint is imposed.

Another protocol, which achieves 2 levels of diversity, is the dynamic decode-and-forward (DDF) protocol suggested in [13]. Unlike DF in DDF the relay listens to the source until it is able to decode reliably. When this happens, the relay re-encodes the source message and sends it in the remaining portion of the frame. In DDF the fraction of the time the relay listens, t, depends on the source-relay link quality. If this channel is of good quality, the relay can decode quickly, and transmit for a longer time interval. If this channel is bad, the relay has a short time left for transmission.

When there are multiple relays in the network, protocols other than DF, CF, AF, NAF or DDF can be used to obtain diversity gains. One method is to employ the relays to form a distributed space time code [31]. The other method is to choose the best relay and to use one relay at a time [15]. Both strategies result in full-diversity gains.

VI. Diversity-Multiplexing Tradeoff of Cooperative Communications

In the previous section we investigated the diversity gains of cooperation for single antenna terminals. We showed that cooperative protocols enable the formation of *virtual* antenna arrays to increase reliability. In a general network, nodes can have multiple antennas, increasing the system's degrees of freedom. It is important to understand how these additional degrees of freedom affect diversity gains, and how other resources, such as rate, coverage area, or network lifetime, need to be sacrificed for higher reliability. In this section we provide such an analysis for the single relay case under Rayleigh fading.

The tradeoff between reliability and rate, known as diversity-multiplexing tradeoff (DMT), is established for point-to-point multiple antenna slow-fading channels in [12]. In DMT the reliability measure is the diversity gain as defined in (21). The transmission rate $R^{(T)}$ is measured by the multiplexing gain, which is

$$r = \lim_{\mathrm{SNR}\to\infty} \frac{R^{(T)}(\mathrm{SNR})}{\log \mathrm{SNR}}.$$

It shows how fast the actual rate of the system increases with SNR. The DMT of an $m \times n$ MIMO is defined as the tradeoff between r and d and is given by $d_{mn}(r)$. $d_{mn}(r)$ is the best achievable diversity, which is a piecewise-linear function connecting the points $(k, d_{mn}(k))$, where $d_{mn}(k) = (m-k)(n-k)$, $k = 0, 1, ..., \min\{m, n\}$ [12]. Note that $d_{mn}(r) = d_{nm}(r)$.

The DMT is a powerful tool to evaluate the performance of different multiple antenna schemes at high SNR; it is also useful for cooperative/relay systems. In the rest of the chapter, we study the multiple-antenna relay channel from a DMT perspective, first for a full-duplex, then for a half-duplex relay.

A. Full-duplex relay

We assume the source, the destination and the relay have m, n and k antennas respectively. Note that our model is general enough to account for the already existing spatial diversity in the form of multiple antennas. The received signals at the relay and at the destination are

$$\mathbf{Y}_R = \mathbf{H}_{SR}\mathbf{X}_S + \mathbf{Z}_R \tag{22}$$

$$\mathbf{Y}_D = \mathbf{H}_{SD}\mathbf{X}_S + \mathbf{H}_{RD}\mathbf{X}_R + \mathbf{Z}_D, \tag{23}$$

where $\mathbf{Z}_R$ and $\mathbf{Z}_D$ are the independent complex Gaussian noise vectors at the corresponding node. $\mathbf{H}_{SR}$, $\mathbf{H}_{SD}$ and $\mathbf{H}_{RD}$ are the $k \times m$, $n \times m$ and $n \times k$ channel gain matrices between the source and the relay, the source and the destination, and the relay and the destination respectively. These channel gain matrices have i.i.d. complex Gaussian entries, with real and imaginary parts zero mean and variance $1/2$ each. Next, we first find an upper bound on DMT, then study the DMT the DF and CF protocols achieve.

An upper bound on DMT can be derived using the well-known cut-set upper bound [19], [16]. In the relay channel there are two cut-sets of interest: the cut-set around the source and the cut-set around the destination. These cut-set mutual information expressions are respectively

$$I_{\mathcal{C}_S} = I(\mathbf{X}_S; \mathbf{Y}_R, \mathbf{Y}_D | \mathbf{X}_R) \tag{24}$$

$$I_{\mathcal{C}_D} = I(\mathbf{X}_S, \mathbf{X}_R; \mathbf{Y}_D). \tag{25}$$

To find the best DMT upper bound, we need to find the maximum of these mutual information expressions. We can further upper bound the mutual information expressions as

$$I_{\mathcal{C}_S} \leq I'_{\mathcal{C}_S} = \log K_{S,RD} \tag{26}$$

$$I_{\mathcal{C}_D} \leq I'_{\mathcal{C}_D} = \log K_{SR,D}, \tag{27}$$

where

$$K_{S,RD} \triangleq \left| \mathbf{I}_{k+n} + \mathbf{H}_{S,RD}\mathbf{H}^{\dagger}_{S,RD} P_S \right|, \tag{28}$$

$$K_{SR,D} \triangleq \left| \mathbf{I}_{n} + \mathbf{H}_{SR,D}\mathbf{H}^{\dagger}_{SR,D} (P_S + P_R) \right|. \tag{29}$$

and

$$\mathbf{H}_{S,RD} = \begin{bmatrix} \mathbf{H}_{SR} \\ \mathbf{H}_{SD} \end{bmatrix}, \quad \mathbf{H}_{SR,D} = \begin{bmatrix} \mathbf{H}_{SD} & \mathbf{H}_{RD} \end{bmatrix}. \tag{30}$$

Note that $P(I'_{\mathcal{C}_i} < R^{(T)}) \doteq \mathrm{SNR}^{-d'_{\mathcal{C}_i}(r)}$, $i = S, D$, with the target data rate $R^{(T)} = r \log \mathrm{SNR}$, $d'_{\mathcal{C}_S}(r) = d_{m(n+k)}(r)$ and $d'_{\mathcal{C}_D}(r) = d_{(m+k)n}(r)$. Then one can easily upper bound the system DMT by

$$d(r) \leq \min\{d_{m(n+k)}(r), d_{(m+k)n}(r)\},$$

for $R^{(T)}$.

This DMT upper bound is quite intuitive. The cut-set around the source node idealizes the relay-destination link and assumes it to be perfect. The system becomes equivalent to an $m \times (n+k)$ MIMO, whose DMT is $d_{m(n+k)}(r)$. Similarly, the cut-set around the destination assumes that the source-relay link is perfect leading to the DMT $d_{(m+k)n}(r)$.

We first study the CF protocol and argue that it is DMT optimal. In the multiple antenna CF protocol the relay performs vector Wyner-Ziv compression with side information taken as the destination's received signal. The CF protocol achieves the rate in (11) subject to (10) if the random variables X_S, X_R, Y_R, Y_D and $\hat{Y}_R$ are respectively replaced with their vector counterparts $\mathbf{X}_S$, $\mathbf{X}_R$, $\mathbf{Y}_R$, $\mathbf{Y}_D$ and $\hat{\mathbf{Y}}_R$. Note that the relay needs to know all the channel gains in the system to ensure that the compression rate constraint in (10) is satisfied. Then CF achieves the rate [16]

$$R^{(CF)} = \log \frac{L_{S,RD}}{\left(\sqrt[k]{\frac{L_{S,RD}}{L_{SR,D}}}+1\right)^k}, \tag{31}$$

subject to

$$\log \frac{L_{S,RD}}{L_{S,D}\hat{N}_R^k} \leq \log \frac{L_{SR,D}}{L_{S,D}},$$

where

$$\begin{aligned}
L_{S,D} &\triangleq \left|\mathbf{H}_{SD}\mathbf{H}_{SD}^{\dagger}\frac{P_S}{m} + \mathbf{I}_n\right| \\
L_{SR,D} &\triangleq \left|\mathbf{H}_{SD}\mathbf{H}_{SD}^{\dagger}\frac{P_S}{m} + \mathbf{H}_{RD}\mathbf{H}_{RD}^{\dagger}\frac{P_R}{k} + \mathbf{I}_n\right|, \\
L_{S,RD} &\triangleq \left|\mathbf{H}_{S,RD}\mathbf{H}_{S,RD}^{\dagger}\frac{P_S}{m} + \begin{bmatrix} (\hat{N}_R+1)\mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix}\right|, \\
L'_{S,RD} &\triangleq \left|\mathbf{H}_{S,RD}\mathbf{H}_{S,RD}^{\dagger}\frac{P_S}{m} + \mathbf{I}_{k+n}\right|.
\end{aligned}$$

Then it can be shown that CF achieves the DMT

$$d_{CF}(r) = \min\{d_{m(n+k)}(r), d_{(m+k)n}(r)\}.$$

We avoid the mathematical details of the proof in this chapter and refer the interested readers to [16].

In the CF protocol as the relay knows all the channel gains the network, it adjusts its compression rate such that the destination can always reliably decode the relay signal. The destination then obtains a perfect copy of the compressed version ($\hat{\mathbf{Y}}_R$) of the relay's observation ($\mathbf{Y}_R$). The compression noise at the relay is never too large and does not hinder diversity gains. The CF protocol adapts well to the current channel conditions, the relay and destination work in accordance and imitate $m \times (n+k)$ MIMO behavior. It is worth mentioning that the SCF protocol in Section II-A.2 has a suboptimal performance and the CF adaptivity is required for DMT optimal behavior.

As an alternative to CF, the relay can use the DF protocol. When the source, the destination and the relay all have a single antenna each, following the analysis in Section II-A.1 we can show that the DF protocol also achieves the DMT upper bound, which is equal to $d_{12}(r)$. But this optimality does not generalize to arbitrary m, n and k. The DF protocol achieves the DMT

$$d_{DF}(r) = \begin{cases} \min\{d_{(m+k)n}(r), d_{mn}(r) + d_{mk}(r)\} & \text{if } 0 \leq r \leq \min\{m,n,k\} \\ d_{mn}(r) & \text{if } \min\{m,n,k\} < r \leq \min\{m,n\} \end{cases} .$$

If m or n (or both) is equal to 1, we find that DF meets the DMT upper bound. Similarly we can show that for cases such as $(m,n,k) = (3,2,2)$ or $(m,n,k) = (4,2,3)$, as $d_{(m+k)n}(r) < d_{mn}(r) + d_{mk}(r)$ for all r, DF is optimal. A general necessary condition for DF to be optimal for all multiplexing gains is $m \geq n$. If $m < n$, then $d_{mn}(r) + d_{mk}(r) \leq d_{m(n+k)}(r) < d_{(m+k)n}$, and DF is suboptimal. Whenever $\min\{m,n,k\} = k$, the degrees of freedom in the direct link is larger than the degrees of freedom in the source to relay link, that is $\min\{m,n\} \geq \min\{m,k\}$. For multiplexing gains in the range $\min\{m,n,k\} < r \leq \min\{m,n\}$, the relay can never help and the system has the direct link DMT $d_{mn}(r)$. Therefore, DF loses its optimality.

Intuitively the suboptimality of DF is because of the hard decision at the relay. The

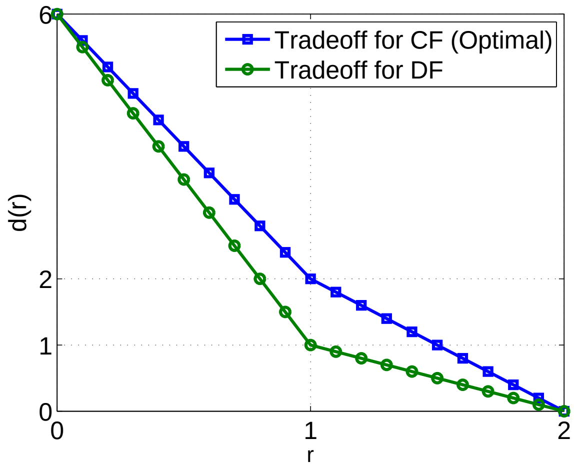

Fig. 9. The source has 2, the destination has 2, and the relay has 1 antenna, $(m, n, k) = (2, 2, 1)$.

point-to-point MIMO DMT, which is the piecewise linear function connecting the points $(k, (m-k)(n-k))$, $k = 1, ..., \min\{m, n\}$, suggests that each additional unit of multiplexing gain uses an antenna both from the transmitter and the receiver, while the remaining antennas provide the diversity gain. Similarly, in the DF protocol each additional multiplexing gain effectively uses an antenna at every node (source, relay and destination) due to hard decision. So the cost of each additional multiplexing gain costs is, in terms of antennas, is one more than the ideal MIMO behavior.

In Fig. 9 we compare CF and DF DMT for $(m, n, k) = (2, 2, 1)$. We see that the CF protocol is always DMT optimal, but the DF protocol is significantly worse. Note that any minor difference in the DMT curve is quite important because diversity gain is the exponent of $1/\mathrm{SNR}$ at high SNR. The suboptimal behavior of DF arises because the outage event when the relay cannot decode can dominate for general m, n and k. In addition to this, for multiplexing gains larger than $\min\{m, n, k\}$, the relay never participates in the communication because it is degrees of freedom limited and cannot decode large multiplexing gain signals. For this region, we observe the direct link behavior. For DF, the available relay CSI does not improve the DMT performance compared to the case where the relay only knows the source-relay channel gains. With this CSI the relay can at best perform beamforming, which brings in power gains but no improvement on DMT. We conclude that soft information transmission, as in the CF protocol, is necessary at the relay not to lose diversity or multiplexing gains.

In Section II we showed that there is a symmetry between DF and CF. DF achieves capacity when the relay is close to the source, and CF achieves capacity if the relay is close to the destination. However, the above analysis reveals that CF and DF protocols do not always behave similar, unlike the single antenna relay system. The degrees of freedom available also has an effect on relaying strategies.

B. Half-duplex relay

In the previous section, we studied the relay channel when the relay is full-duplex. Although this is an ideal assumption about the relay's physical capabilities, it helps us understand the fundamental differences between the DF and CF protocols. In this section we assume a half-duplex relay to study how this affects the DMT behavior of the relay channel. In this section we investigate *static* half-duplex channels, in which the relay state is not controlled based on channel realizations. The relay can also perform *dynamic* behavior as in the DDF protocol. In dynamic protocols, the relay determines the fraction of time it listens and transmits depending on the channel gains.

In the previous subsection we found an upper bound on DMT, when the relay is full-duplex. As this bound may not be tight when the relay is half-duplex, next we consider a half-duplex DMT upper bound. The cut-set mutual information expressions for a half-duplex relay are respectively [10]

$$I_{\mathcal{C}_S}(t) = tI(\mathbf{X}_S;\mathbf{Y}_R,\mathbf{Y}_D|q_1) + (1-t)I(\mathbf{X}_S;\mathbf{Y}_D|\mathbf{X}_R,q_2) \quad (32)$$

$$I_{\mathcal{C}_D}(t) = tI(\mathbf{X}_S;\mathbf{Y}_D|q_1) + (1-t)I(\mathbf{X}_S,\mathbf{X}_R;\mathbf{Y}_D|q_2). \quad (33)$$

Note that these expressions depend on t, the amount of time the relay listens. Then an upper bound on these expressions are

$$I_{\mathcal{C}_S}(t) \leq I'_{\mathcal{C}_S}(t) = t\log K_{S,RD} + (1-t)\log K_{S,D} \tag{34}$$

$$I_{\mathcal{C}_D}(t) \leq I'_{\mathcal{C}_D}(t) = t\log K_{S,D} + (1-t)\log K_{SR,D} \tag{35}$$

where $K_{S,RD}$ and $K_{SR,D}$ are respectively defined in (28) and (29), and

$$K_{S,D} \triangleq \left|\mathbf{H}_{SD}\mathbf{H}_{SD}^{\dagger}P_S + \mathbf{I}_n\right|. \tag{36}$$

For a target data rate $R^{(T)} = r\log \mathrm{SNR}$, and for a fixed t, if $P(I'_{\mathcal{C}_i}(t) < R^{(T)}) \dot{=} \mathrm{SNR}^{-d'_{\mathcal{C}_i}(r,t)}$, $i = S, D$, then $d(r,t)$, the best achievable diversity for the half-duplex relay channel for fixed t, satisfies

$$d(r,t) \leq \min\{d'_{\mathcal{C}_S}(r,t), d'_{\mathcal{C}_D}(r,t)\}. \tag{37}$$

Optimizing over t we find an upper bound on the static multiple antenna half-duplex relay channel DMT as

$$d(r) \leq \max_{t\in[0,1]} \min\{d'_{\mathcal{C}_S}(r,t), d'_{\mathcal{C}_D}(r,t)\}. \tag{38}$$

In general it is hard to compute the exact DMT of (38). In particular for static protocols, to find $d'_{\mathcal{C}_S}(r,t)$ and $d'_{\mathcal{C}_D}(r,t)$ for general m, n and k we need to calculate the joint eigenvalue distribution of two correlated Hermitian matrices, $\mathbf{H}_{SD}\mathbf{H}_{SD}^{\dagger}$ and $\mathbf{H}_{S,RD}\mathbf{H}_{S,RD}^{\dagger}$ or $\mathbf{H}_{SD}\mathbf{H}_{SD}^{\dagger}$ and $\mathbf{H}_{SR,D}\mathbf{H}_{SR,D}^{\dagger}$. However, when $m = 1$, both $\mathbf{H}_{SD}\mathbf{H}_{SD}^{\dagger}$ and $\mathbf{H}_{S,RD}\mathbf{H}_{S,RD}^{\dagger}$ reduce to vectors and $d'_{\mathcal{C}_S}(r,t)$ is easier to compute. Similarly, when $n = 1$, $\mathbf{H}_{SD}\mathbf{H}_{SD}^{\dagger}$ and $\mathbf{H}_{SR,D}\mathbf{H}_{SR,D}^{\dagger}$ are vectors, and $d'_{\mathcal{C}_D}(r,t)$ can be found. An explicit form for $d'_{\mathcal{C}_S}(r,t)$ for $m = 1$ is given as

$$d'_{\mathcal{C}_S}(r,t) = \begin{cases} n + k - k\frac{r}{t} & \text{if} \quad r \leq t, \text{ and } t \leq \frac{k}{n+k} \\ n\left(\frac{1-r}{1-t}\right) & \text{if} \quad r \geq t, \text{ and } t \leq \frac{k}{n+k} \\ (n+k)(1-r) & \text{if} \quad t \geq \frac{k}{n+k} \end{cases}.$$

For $n = 1$ and for arbitrary m and k, $d'_{\mathcal{C}_D}(r,t)$ has the same expression as $d'_{\mathcal{C}_S}(r,t)$ if n and t are replaced with m and $(1-t)$ in the above expressions [16].

Although we do not have an explicit expression for $d'_{\mathcal{C}_S}(r,t)$ or $d'_{\mathcal{C}_D}(r,t)$ for general (m,n,k), we can comment on some special cases and get insights about multiple antenna, half-duplex behavior. First we observe that $d'_{\mathcal{C}_S}(r,t)$ and $d'_{\mathcal{C}_D}(r,t)$ depend on the choice of t, and the upper bound of (38) is not always equal to the full-duplex bound. As an example consider $(m,n,k) = (1,1,2)$, for which $d'_{\mathcal{C}_S}(r,t)$ is shown in Fig. 10. To achieve the full-duplex bound for all r, $d'_{\mathcal{C}_S}(r,t)$ needs to have $t \geq 2/3$, whereas $d'_{\mathcal{C}_D}(r,t)$ needs $t \leq 1/3$. As both cannot be satisfied simultaneously, $d(r,t)$ is less than the full-duplex bound for all t.

On the other hand, to maximize the half-duplex DMT it is optimal to choose $t = 1/2$ whenever $m = n$. To see this, we compare (34) with (35), and note that both $K'_{S,RD} \geq K_{S,D}$ and $K_{SR,D} \geq K_{S,D}$ for $m = n$. Furthermore, for $m = n$ $d'_{\mathcal{C}_S}(r,t) = d'_{\mathcal{C}_D}(r,1-t)$, and $d'_{\mathcal{C}_S}(r,t)$ is a non-decreasing function in t. Therefore $\min\{d'_{\mathcal{C}_S}(r,t), d'_{\mathcal{C}_D}(r,t)\}$ must reach its maximum at $t = 1/2$.

Similar to the full-duplex case, in the multiple-antenna half-duplex relay channel [16] the CF protocol is DMT optimal and achieves the half-duplex cut-set bound DMT

$$d_{CF}(r) = \max_{t\in[0,1]} \min\{d'_{\mathcal{C}_S}(r,t), d'_{\mathcal{C}_D}(r,t)\}.$$

Note that if the relay is dynamic, CF can also behave dynamically, choosing t as a function of CSI available at the relay, and still achieve the DMT upper bound for dynamic protocols at high SNR.

Fig. 11 shows the CF DMT in comparison to DDF and NAF DMT for $(m,n,k) = (1,1,1)$. For this single antenna case the half-duplex DMT bound is equal to the full-duplex case. We observe that DDF does not meet the upper bound for $\frac{1}{2} \leq r \leq 1$, as in this range, the relay does not have enough time to transmit the high rate information it received [13]. NAF protocol is also suboptimal. However, CF is DMT optimal for $(m,n,k) = (1,1,1)$ as well as arbitrary antenna configurations.

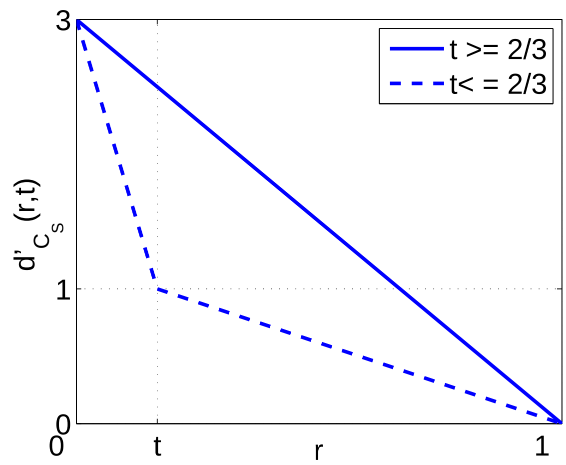

Fig. 10. DMT upper bound for the cut-set around the source, $\mathcal{C}_S$. The source has 1, the destination has 1, and the relay has 2 antennas, $(m, n, k) = (1, 1, 2)$. Note that as $m = n$, $d'_{\mathcal{C}_S}(r, t) = d'_{\mathcal{C}_D}(r, 1 - t)$. The upper bound in (38) reaches its maximum for $t = 1/2$. The solid line in the figure is also equal to the full-duplex bound.

We showed that for a multiple antenna full-duplex relay system, the probability that the relay cannot decode is dominant and the DF protocol becomes suboptimal. Therefore, we do not expect any relay decoding based protocol to achieve the DMT upper bound in the multiple antenna half-duplex system either.

VII. Conclusion

In this chapter we outline some of the major information theoretical advances in the cooperative communications literature. For discrete memoryless and Gaussian channels, we show that the relay has a large potential to improve achievable rates with respect to direct transmission. We explain two fundamental strategies DF and CF for the relay channel. We also provide an overview of mutual cooperation among two users under the generalized multiple access channel model. For multiple relay systems we give an example on how multiple relays can be coordinated.

In addition to non-fading channels, cooperation offers substantial gains under fading.

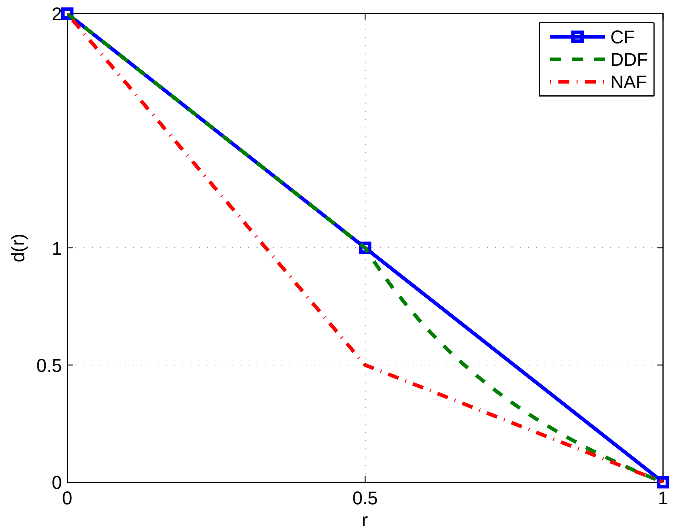

Fig. 11. The DMT for CF, DDF, and NAF protocols for single antenna nodes.

We show that via cooperative protocols *virtual* antenna arrays can be formed and spatial diversity gains similar to physical antenna arrays can be achieved. To provide a complete characterization, we also study the diversity gains in the DMT context. We investigate the effects of increased degrees of freedom and relay processing capability on cooperative DMT. We show that although full-duplex DF and CF have the same DMT when each node has a single antenna, this is not the case for multiple-antenna relay systems. DF can be suboptimal, yet CF is DMT optimal for arbitrary antenna configurations. For the half-duplex relay CF remains to be DMT optimal. Although it is hard to find the DMT upper bound explicitly for arbitrary number of antennas at the nodes, we have solutions for special cases. We also show that the static half-duplex DMT bound is tighter than the full-duplex bound in general.

Despite the recent significant advances we outlined in this chapter, there are many open problems related to the relay channel in particular and to cooperative networks in general. For example there are no practical channel codes designed for CF, which hinders the implementation of this robust protocol. On the other hand, CF protocol performance strongly depends on the channel state information available at the relay. Suggesting new relaying protocols, which process the soft information, but do not require this information are of utmost interest. In general, utilization of feedback in the relay channel is not fully understood. In addition to suggesting new protocols, low to moderate SNR studies need to be carried for the relay channel to complement the high SNR results of DMT. To understand larger cooperative networks better and to extend the cooperation benefits to large networks we also need to understand the role of relaying/cooperation in interference, broadcast and multiple-access channels, and to explore network gains of cooperation, relay assignment, synchronization and coordination issues, and incentives to cooperate.

References

[1] E. C. van der Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, p. 121, 1971.

[2] T. M. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, p. 572, September 1979.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part I: System description," *IEEE Transactions on Communications*, vol. 51, no. 11, p. 1927, November 2003.

[4] ——, "User cooperation diversity-Part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, p. 1939, November 2003.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, p. 3062, December 2004.

[6] F. M. J. Willems, E. C. van der Meulen, and J. P. M. Schalkwijk, "Achievable rate region for the multiple access channel with generalized feedback," in *Proc. Annu. Allerton Conf. Communication, Control, Computing*, 1983, p. 284.

[7] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, p. 3037, 2005.

[8] S. Avestimehr, S. Diggavi, and D. Tse, "A deterministic approach to wireless relay networks," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, 2007.

[9] ——, "Approximate capacity of Gaussian relay networks," in *Proceedings of IEEE International Symposium on Information Theory*, 2008.

[10] M. Khojastepour, A. Sabharwal, and B. Aazhang, “On the capacity of ‘cheap’ relay networks,” in *Proceedings of 37th Conference on Information Sciences and Systems*, 2003.

[11] G. Kramer, “Models and theory for relay channels with receive constraints,” in *Proceedings of 42nd Allerton Conference on Communication, Control and Computing*, 2004.

[12] L. Zheng and D. N. C. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels,” *IEEE Transactions on Information Theory*, vol. 49, p. 1073, May 2003.

[13] K. Azarian, H. El-Gamal, and P. Schniter, “On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, p. 4152, December 2005.

[14] N. Prasad and M. K. Varanasi, “Diversity and multiplexing tradeoff bounds for cooperative diversity protocols,” in *Proceedings of IEEE International Symposium on Information Theory*, 2004, p. 268.

[15] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, “A simple cooperative diversity method based on network path selection,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, p. 659, March 2006.

[16] M. Yuksel and E. Erkip, “Multiple-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective,” *IEEE Transactions on Information Theory, Special Issue on Models, Theory, and Codes for Relaying and Cooperation in Communication Networks*, vol. 53, no. 10, p. 3371, October 2007.

[17] IEEE Transactions on Information Theory, Special Issue on Models, Theory, and Codes for Relaying and Cooperation in Communication Networks, October 2007.

[18] G. Kramer, I. Maric, and R. D. Yates, “Cooperative Communications,” *Foundations and Trends in Networking, NOW Publishers Inc.*, vol. 1, no. 3-4, 2006.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John-Wiley & Sons, Inc., 1991.

[20] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, p. 1, January 1976.

[21] L. Xie and P. R. Kumar, “A network information theory for wireless communication: Scaling laws and optimal operation,” *IEEE Transactions on Information Theory*, vol. 50, no. 5, p. 748, May 2004.

[22] C. Zeng, F. Kuhlmann, and A. Buzo, “Achievability proof of some multiuser channel coding theorems using backward decoding,” *IEEE Transactions on Information Theory*, vol. 35, no. 6, p. 1160, November 1989.

[23] F. M. J. Willems and E. C. van der Meulen, “The discrete memoryless multiple-access channel with cribbing encoders,” *IEEE Transactions on Information Theory*, vol. 31, p. 313, May 1985.

[24] F. M. J. Willems, “Informationtheoretical results for the discrete memoryless multiple access channel,” PhD thesis, Katholieke Universiteit, Leuven, Leuven, Belgium, October 1982. [Online]. Available: http://www.sps.ele.tue.nl/members/F.M.J.Willems/RESEARCH_files/PhDThesis/PhDthesis.htm

[25] T. M. Cover, A. E. Gamal, and M. Salehi, “Multiple access channels with arbitrarily correlated sources,” *IEEE Transactions on Information Theory*, vol. 26, p. 648, November 1980.

[26] M. Khojastepour, A. Sabharwal, and B. Aazhang, “On capacity of Gaussian ’cheap’ relay channel,” in *Proceedings of IEEE Global Telecommunications Conference*, p. 1776.

[27] L. Lai, K. Liu, and H. E. Gamal, "The three node wireless network: Achievable rates and cooperation strategies," *IEEE Transactions on Information Theory*, vol. 52, no. 3, p. 805, March 2006.

[28] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, p. 1099, August 2004.

[29] J. G. Proakis, *Digital Communications*, Fourth ed. McGraw-Hill, Inc., New York, 2000.

[30] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[31] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, p. 2415, October 2003.