Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior

J. Nicholas Laneman, *Member, IEEE*, David N. C. Tse, *Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

***Abstract*—We develop and analyze low-complexity cooperative diversity protocols that combat fading induced by multipath propagation in wireless networks. The underlying techniques exploit space diversity available through cooperating terminals' relaying signals for one another. We outline several strategies employed by the cooperating radios, including fixed relaying schemes such as amplify-and-forward and decode-and-forward, selection relaying schemes that adapt based upon channel measurements between the cooperating terminals, and incremental relaying schemes that adapt based upon limited feedback from the destination terminal. We develop performance characterizations in terms of outage events and associated outage probabilities, which measure robustness of the transmissions to fading, focusing on the high signal-to-noise ratio (SNR) regime. Except for fixed decode-and-forward, all of our cooperative diversity protocols are efficient in the sense that they achieve full diversity (i.e., second-order diversity in the case of two terminals), and, moreover, are close to optimum (within 1.5 dB) in certain regimes. Thus, using distributed antennas, we can provide the powerful benefits of space diversity without need for physical arrays, though at a loss of spectral efficiency due to half-duplex operation and possibly at the cost of additional receive hardware. Applicable to any wireless setting, including cellular or *ad hoc* networks—wherever space constraints preclude the use of physical arrays—the performance characterizations reveal that large power or energy savings result from the use of these protocols.**

***Index Terms*—Diversity techniques, fading channels, outage probability, relay channel, user cooperation, wireless networks.**

I. Introduction

IN wireless networks, signal fading arising from multipath propagation is a particularly severe channel impairment that can be mitigated through the use of diversity [1]. Space, or multiple-antenna, diversity techniques are particularly attractive as they can be readily combined with other forms of diversity, e.g., time and frequency diversity, and still offer dramatic performance gains when other forms of diversity are unavailable. In contrast to the more conventional forms of space diversity with physical arrays [2]–[4], this work builds upon the classical relay channel model [5] and examines the problem of creating and exploiting space diversity using a collection of distributed antennas belonging to multiple terminals, each with its own information to transmit. We refer to this form of space diversity as *cooperative diversity* (cf., user cooperation diversity of [6]) because the terminals share their antennas and other resources to create a "virtual array" through distributed transmission and signal processing.

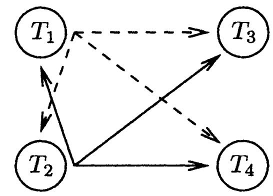

Fig. 1. Illustration of radio signal paths in an example wireless network with terminals T_1 and T_2 transmitting information to terminals T_3 and T_4, respectively.

A. Motivating Example

To illustrate the main concepts, consider the example wireless network in Fig. 1, in which terminals T_1 and T_2 transmit to terminals T_3 and T_4, respectively. This example might correspond to a snapshot of a wireless network in which a higher level network protocol has allocated bandwidth to two terminals for transmission to their intended destinations or next hops. For example, in the context of a cellular network, T_1 and T_2 might correspond to handsets and $T_3 = T_4$ might correspond to the base station [7]. As another example, in the context of a wireless local-area network (LAN), the case $T_3 \neq T_4$ might correspond to an *ad hoc* configuration among the terminals, while the case $T_3 = T_4$ might correspond to an infrastructure configuration, with T_3 serving as an access point [8]. The broadcast nature of the wireless medium is the key property that allows for cooperative diversity among the transmitting terminals: transmitted signals can, in principle, be received and processed by any of a number of terminals. Thus, instead of transmitting independently to their intended destinations, T_1 and T_2 can listen to each other's transmissions and jointly communicate their information. Although these extra observations of the transmitted signals are available for free (except, possibly, for the cost of

Manuscript received January 22, 2002; revised June 10, 2004. The work of J. N. Laneman and G. W. Wornell was supported in part by ARL Federated Labs under Cooperative Agreement DAAD19-01-2-0011, and by the National Science Foundation under Grant CCR-9979363. The work of J. N. Laneman was also supported in part by the State of Indiana through the 21st Century Research and Technology Fund, and by the National Science Foundation under Grant ECS03-29766. The work of D. N. C. Tse was supported in part by the National Science Foundation under Grant ANI-9872764. The material in this paper was presented in part at the 38th Annual Allerton Conference on Communications, Control and Computing, Monticello, IL, October 2000, and at the IEEE International Symposium on Information Theory, Washington, DC, June 2001.

J. N. Laneman was with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge. He is now with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: jnl@nd.edu).

D. N. C. Tse is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA (e-mail: dtse@eecs.berkeley.edu).

G. W. Wornell is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: gww@mit.edu).

Communicated by L. Tassiulas, Associate Editor for Communication Networks.

Digital Object Identifier 10.1109/TIT.2004.838089

additional receive hardware) wireless network protocols often ignore or discard them.

In the most general case, T_1 and T_2 can pool their resources, such as power and bandwidth, to cooperatively transmit their information to their respective destinations, corresponding to a wireless multiple-access channel with relaying for $T_3 = T_4$, and to a wireless interference channel with relaying for $T_3 \neq T_4$. At one extreme, corresponding to a wireless relay channel, the transmitting terminals can focus all their resources on transmitting the information of T_1; in this case, T_1 acts as the "source" of the information, and T_2 serves as a "relay." Such an approach might provide diversity in a wireless setting because, even if the fading is severe between T_1 and T_3, the information might be successfully transmitted through T_2. Similarly, T_1 and T_2 can focus their resources on transmitting the information of T_2, corresponding to another wireless relay channel.

B. Related Work

Relay channels and their extensions form the basis for our study of cooperative diversity. This section summarizes some of the relevant literature in this area. Because relaying and cooperative diversity essentially create a virtual antenna array, work on multiple-antenna systems, or multiple-input, multiple-output (MIMO) systems, is of course relevant, as are different ways of characterizing fundamental performance limits in wireless channels, in particular outage probability for nonergodic settings. Throughout the rest of the paper, we assume that the reader is familiar with these latter areas, and refer the interested reader to [2]–[4], [9], [10], and references therein, for an introduction to the relevant concepts from multiple-antenna systems and to [11] for an introduction to outage capacity for fading channels.

1) Relay Channels: The classical relay channel models a class of three-terminal communication channels originally examined by van der Meulen [12], [13]. Cover and El Gamal [5] treat certain discrete memoryless and additive white Gaussian noise relay channels, and they determine channel capacity for the class of *physically degraded*[1] relay channels. More generally, they develop lower bounds on capacity, i.e., achievable rates, via three structurally different random coding schemes:

- *facilitation* [5, Theorem 2], in which the relay does not actively help the source, but rather, facilitates the source transmission by inducing as little interference as possible;
- *cooperation* [5, Theorem 1], in which the relay fully decodes the source message and retransmits, jointly with the source, a bin index (in the sense of Slepian–Wolf coding [14], [15]) of the previous source message;
- *observation*[2] [5, Theorem 6], in which the relay encodes a quantized version of its received signal, using ideas from source coding with side information [14], [16], [17].

[1]At a high level, degradedness means that the destination receives a corrupted version of what the relay receives, all conditioned on the relay transmit signal. While this class is mathematically convenient, none of the wireless channels found in practice are well modeled by this class.

[2]The names *facilitation* and *cooperation* were introduced in [5], but Cover and El Gamal did not give a name to their third approach. We use the name *observation* throughout the paper for convenience.

Loosely speaking, cooperation yields highest achievable rates when the source-relay channel quality is very high, and observation yields highest achievable rates when the relay-destination channel quality is very high. Various extensions to the case of multiple relays have appeared in the work of Schein and Gallager [18], [19], Gupta and Kumar [20], [21], Gastpar *et al.* [22]–[24], and Reznik *et al.* [25]. For channels with multiple information sources, Kramer and Wijngaarden [26] consider a multiple-access channel in which the sources communicate to a single destination and share a single relay.

2) Multiple-Access Channels With Generalized Feedback: Work by King [27], Carleial [28], and Willems *et al.* [29]–[32] examines multiple-access channels with generalized feedback. Here, the generalized feedback allows the sources to essentially act as relays for one another. This model relates most closely to the wireless channels we have in mind. The constructions in [28]–[30] can be viewed as two-terminal generalizations of the cooperation scheme in [5]; the construction [27] may be viewed as a two-terminal generalization of the observation scheme in [5]. Sendonaris *et al.* introduce multipath fading into the model of [28], [30], calling their approaches for this system model *user cooperation diversity* [6], [33], [34]. For ergodic fading, they illustrate that the adapted coding scheme of [30] enlarges the achievable rate region.

C. Summary of Results

We now highlight the results of the present paper, many of which were initially reported in [35], [36], and recently extended in [37]. This paper develops low-complexity cooperative diversity protocols that explicitly take into account certain implementation constraints in the cooperating radios. Specifically, while previous work on relay and cooperative channels allows the terminals to transmit and receive simultaneously, i.e., *full-duplex*, we constrain them to employ *half-duplex* transmission. Furthermore, although previous work employs channel state information (CSI) at the transmitters in order to exploit coherent transmission, we utilize CSI at the receivers only. Finally, although previous work focuses primarily on ergodic settings and characterizes performance via Shannon capacity or capacity regions, we focus on nonergodic or delay-constrained scenarios and characterize performance by outage probability [11].

We outline several cooperative protocols and demonstrate their robustness to fairly general channel conditions. In addition to direct transmission, we examine fixed relaying protocols in which the relay either amplifies what it receives, or fully decodes, re-encodes, and retransmits the source message. We call these options *amplify-and-forward* and *decode-and-forward*, respectively. Obviously, these approaches are inspired by the observation [5], [18], [28] and cooperation [5], [6], [30] schemes, respectively, but we intentionally limit the complexity of our protocols for ease of potential implementation. Furthermore, our analysis suggests that cooperating radios may also employ threshold tests on the measured channel quality between them, to obtain adaptive protocols, called *selection relaying*, that choose the strategy with best performance. In addition, adaptive protocols based upon limited feedback from the destination terminal, called *incremental relaying*, are also

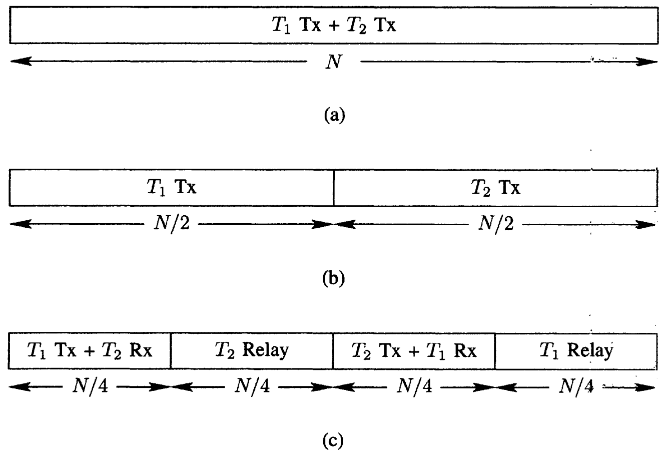

Fig. 2. Example time-division channel allocations for (a) direct transmission with interference, (b) orthogonal direct transmission, and (c) orthogonal cooperative diversity. We focus on orthogonal transmissions of the form (b) and (c) throughout the paper.

developed. Selection and incremental relaying protocols represent new directions for relay and cooperative transmission, building upon existing ideas.

For scenarios in which CSI is unavailable to the transmitters, even full-duplex cooperation cannot improve the sum capacity for ergodic fading [38]. Consequently, we focus on delay-limited or nonergodic scenarios, and evaluate performance of our protocols in terms of outage probability [11]. We show analytically that, except for fixed decode-and-forward, each of our cooperative protocols achieves *full diversity*, i.e., outage probability decays proportional to $1/\mathtt{SNR}^2$, where $\mathtt{SNR}$ is signal-to-noise ratio (SNR) of the channel, whereas it decays proportional to $1/\mathtt{SNR}$ without cooperation. At fixed low rates, amplify-and-forward and selection decode-and-forward are at most 1.5 dB from optimal and offer large power or energy savings over direct transmission. For sufficiently high rates, direct transmission becomes preferable to fixed and selection relaying, because these protocols repeat information all the time. Incremental relaying exploits limited feedback to overcome this bandwidth inefficiency by repeating only rarely. More broadly, the relative attractiveness of the various schemes can depend upon the network architecture and implementation considerations.

D. Outline

An outline of the remainder of the paper is as follows. Section II describes our system model for the wireless networks under consideration. Section III outlines fixed, selection, and incremental relaying protocols at a high level. Section IV characterizes the outage behavior of the various protocols in terms of outage events and outage probabilities, using several results for exponential random variables developed in Appendix I. Section V compares the results from a number of perspectives, and Section VI offers some concluding remarks.

II. System Model

In our model for the wireless channel in Fig. 1, narrow-band transmissions suffer the effects of frequency nonselective fading and additive noise. Our analysis in Section IV focuses on the case of slow fading, to capture scenarios in which delay constraints are on the order of the channel coherence time, and measures performance by outage probability, to isolate the benefits of space diversity. While our cooperative protocols can be naturally extended to the kinds of wide-band and highly mobile scenarios in which frequency- and time-selective fading, respectively, are encountered, the potential impact of our protocols becomes less substantial when other forms of diversity can be exploited in the system.

A. Medium Access

As in many current wireless networks, such as cellular and wireless LANs, we divide the available bandwidth into orthogonal channels and allocate these channels to the transmitting terminals, allowing our protocols to be readily integrated into existing networks. As a convenient by-product of this choice, we are able to treat the multiple-access (single receiver) and interference (multiple receivers) cases described in Section I-A simultaneously, as a pair of relay channels with signaling between the transmitters. Furthermore, removing the interference between the terminals at the destination radio(s) substantially simplifies the receiver algorithms and the outage analysis for purposes of exposition.

For all of our cooperative protocols, transmitting terminals must also process their received signals; however, current limitations in radio implementation preclude the terminals from full-duplex operation, i.e., transmitting and receiving at the same time in the same frequency band. Because of severe attenuation over the wireless channel, and insufficient electrical isolation between the transmit and receive circuitry, a terminal's transmitted signal drowns out the signals of other terminals at its receiver input.[3] Thus, to ensure half-duplex operation, we further divide each channel into orthogonal subchannels. Fig. 2 illustrates our channel allocation for an example time-division approach with two terminals.

We expect that some level of synchronization between the terminals is required for cooperative diversity to be effective. As suggested by Fig. 2 and the modeling discussion to follow, we

[3]Typically, a terminal's transmit signal is 100–150 dB above its received signal.

consider the scenario in which the terminals are block, carrier, and symbol synchronous. Given some form of network block synchronization, carrier and symbol synchronization for the network can build upon the same between the individual transmitters and receivers. Exactly how this synchronization is achieved, and the effects of small synchronization errors on performance, is beyond the scope of this paper.

B. Equivalent Channel Models

Under the above orthogonality constraints, we can now conveniently, and without loss of generality, characterize our channel models using a time-division notation; frequency-division counterparts to this model are straightforward. Due to the symmetry of the channel allocations, we focus on the message of the "source" terminal T_s, which potentially employs terminal T_r as a "relay," in transmitting to the "destination" terminal T_d, where $s, r \in \{1, 2\}$ and $d \in \{3, 4\}$. We utilize a baseband-equivalent, discrete-time channel model for the continuous-time channel, and we consider N consecutive uses of the channel, where N is large.

For direct transmission, our baseline for comparison, we model the channel as

$$\mathsf{y}_d[n] = \mathsf{a}_{s,d}\,\mathsf{x}_s[n] + \mathsf{z}_d[n] \tag{1}$$

for, say, $n = 1, \ldots, N/2$, where $\mathsf{x}_s[n]$ is the source transmitted signal, and $\mathsf{y}_d[n]$ is the destination received signal. The other terminal transmits for $n = N/2+1, \ldots, N$ as depicted in Fig. 2(b). Thus, in the baseline system, each terminal utilizes only half of the available degrees of freedom of the channel.

For cooperative diversity, we model the channel during the first half of the block as

$$\mathsf{y}_r[n] = \mathsf{a}_{s,r}\,\mathsf{x}_s[n] + \mathsf{z}_r[n] \tag{2}$$

$$\mathsf{y}_d[n] = \mathsf{a}_{s,d}\,\mathsf{x}_s[n] + \mathsf{z}_d[n] \tag{3}$$

for, say, $n = 1, \ldots, N/4$, where $\mathsf{x}_s[n]$ is the source transmitted signal and $\mathsf{y}_r[n]$ and $\mathsf{y}_d[n]$ are the relay and destination received signals, respectively. For the second half of the block, we model the received signal as

$$\mathsf{y}_d[n] = \mathsf{a}_{r,d}\,\mathsf{x}_r[n] + \mathsf{z}_d[n] \tag{4}$$

for $n = N/4+1, \ldots, N/2$, where $\mathsf{x}_r[n]$ is the relay transmitted signal and $\mathsf{y}_d[n]$ is the destination received signal. A similar setup is employed in the second half of the block, with the roles of the source and relay reversed, as depicted in Fig. 2(c). Note that, while again half the degrees of freedom are allocated to each source terminal for transmission to its destination, only a quarter of the degrees of freedom are available for communication to its relay.

In (1)–(4), $\mathsf{a}_{i,j}$ captures the effects of path-loss, shadowing, and frequency nonselective fading, and $\mathsf{z}_j[n]$ captures the effects of receiver noise and other forms of interference in the system, where $i \in \{s, r\}$ and $j \in \{r, d\}$. We consider the scenario in which the fading coefficients are known to, i.e., accurately measured by, the appropriate receivers, but not fully known to, or not exploited by, the transmitters. Statistically, we model $\mathsf{a}_{i,j}$ as zero-mean, independent, circularly symmetric complex Gaussian random variables with variances $\sigma^2_{i,j}$. Furthermore, we model $\mathsf{z}_j[n]$ as zero-mean mutually independent, circularly symmetric, complex Gaussian random sequences with variance N_0.

C. Parameterizations

Two important parameters of the system are the SNR without fading and the spectral efficiency. We now define these parameters in terms of standard parameters in the continuous-time channel. For a continuous-time channel with bandwidth W hertz available for transmission, the discrete-time model contains W two-dimensional symbols per second (2D/s).

If the transmitting terminals have an average power constraint in the continuous-time channel model of P_c joules per second, we see that this translates into a discrete-time power constraint of $P = 2P_c/W$ J/2D since each terminal transmits in half of the available degrees of freedom, under both direct transmission and cooperative diversity. Thus, the channel model is parameterized by the SNR random variables $\mathtt{SNR}\,|\mathsf{a}_{i,j}|^2$, where

$$\mathtt{SNR} := \frac{2P_c}{N_0 W} = \frac{P}{N_0} \tag{5}$$

is the common SNR without fading. Throughout our analysis, we vary $\mathtt{SNR}$, and allow for different (relative) received SNRs through appropriate choice of the fading variances. As we will see, increasing the source-relay SNR proportionally to increases in the source-destination SNR leads to the full diversity benefits of the cooperative protocols.

In addition to SNR, transmission schemes are further parameterized by the rate r bits per second, or spectral efficiency

$$\mathtt{R} := 2r/W \text{ b/s/Hz} \tag{6}$$

attempted by the transmitting terminals. Note that (6) is the rate normalized by the number of degrees of freedom utilized by each terminal, not by the total number of degrees of freedom in the channel.

Nominally, one could parameterize the system by the pair $(\mathtt{SNR}, \mathtt{R})$; however, our results lend more insight, and are substantially more compact, when we parameterize the system by either of the pairs $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$ or $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$, where[4]

$$\mathtt{SNR}_{\text{norm}} := \frac{\mathtt{SNR}}{2^{\mathtt{R}} - 1}, \qquad \mathtt{R}_{\text{norm}} := \frac{\mathtt{R}}{\log\left(1 + \mathtt{SNR}\,\sigma^2_{s,d}\right)}. \tag{7}$$

For a complex additive white Gaussian noise (AWGN) channel with bandwidth $(W/2)$ and SNR given by $\mathtt{SNR}\,\sigma^2_{s,d}$, $\mathtt{SNR}_{\text{norm}} > 1$ is the SNR normalized by the minimum SNR required to achieve spectral efficiency $\mathtt{R}$ [39]. Similarly, $\mathtt{R}_{\text{norm}} < 1$ is the spectral efficiency normalized by the maximum achievable spectral efficiency, i.e., channel capacity [9], [10]. In this sense, parameterizations given by $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$ and $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$ are duals of one another. For our setting with fading, as we will see, the two parameterizations yield tradeoffs between different aspects of system performance: results under $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$ exhibit a tradeoff between the normalized SNR gain and spectral efficiency of a protocol, while results under

[4]Unless otherwise indicated, logarithms in this paper are taken to base 2.

$(\mathtt{SNR}, \mathtt{R}_{\mathrm{norm}})$ exhibit a tradeoff between the diversity order and normalized spectral efficiency of a protocol. The latter tradeoff has also been called the *diversity-multiplexing* tradeoff in [9], [10].

Note that, although we have parameterized the transmit powers and noise levels to be symmetric throughout the network for purposes of exposition, asymmetries in average SNR and path loss can be lumped into the fading variances $\sigma_{i,j}^2$. Furthermore, while the tools are powerful enough to consider general rate pairs $(\mathtt{R}_1, \mathtt{R}_2)$, we consider the equal rate point, i.e., $\mathtt{R}_1 = \mathtt{R}_2 = \mathtt{R}$, for purposes of exposition.

III. Cooperative Diversity Protocols

In this section, we describe a variety of low-complexity cooperative diversity protocols that can be utilized in the network of Fig. 1, including fixed, selection, and incremental relaying. These protocols employ different types of processing by the relay terminals, as well as different types of combining at the destination terminals. For fixed relaying, we allow the relays to either amplify their received signals subject to their power constraint, or to decode, re-encode, and retransmit the messages. Among many possible adaptive strategies, selection relaying builds upon fixed relaying by allowing transmitting terminals to select a suitable cooperative (or noncooperative) action based upon the measured SNR between them. Incremental relaying improves upon the spectral efficiency of both fixed and selection relaying by exploiting limited feedback from the destination and relaying only when necessary.

In any of these cases, the radios may employ repetition or more powerful codes. We focus on repetition coding throughout the sequel, for its low implementation complexity and ease of exposition. Destination radios can appropriately combine their received signals by exploiting control information in the protocol headers.

A. Fixed Relaying

1) Amplify-and-Forward: For amplify-and-forward transmission, the appropriate channel model is (2)–(4). The source terminal transmits its information as $\mathsf{x}_s[n]$, say, for $n = 1, \ldots, N/4$. During this interval, the relay processes $\mathsf{y}_r[n]$, and relays the information by transmitting

$$\mathsf{x}_r[n] = \beta\, \mathsf{y}_r[n - N/4] \tag{8}$$

for $n = N/4 + 1, \ldots, N/2$. To remain within its power constraint (with high probability), an amplifying relay must use gain

$$\beta \leq \sqrt{\frac{P}{|\mathsf{a}_{s,r}|^2 P + N_0}} \tag{9}$$

where we allow the amplifier gain to depend upon the fading coefficient $\mathsf{a}_{s,r}$ between the source and relay, which the relay estimates to high accuracy. This scheme can be viewed as repetition coding from two separate transmitters, except that the relay transmitter amplifies its own receiver noise. The destination can decode its received signal $\mathsf{y}_d[n]$ for $n = 1, \ldots, N/2$ by first appropriately combining the signals from the two subblocks using one of a variety of combining techniques; in the sequel, we focus on a suitably designed matched filter, or maximum-ratio combiner.

2) Decode-and-Forward: For decode-and-forward transmission, the appropriate channel model is again (2)–(4). The source terminal transmits its information as $\mathsf{x}_s[n]$, say, for $n = 0, \ldots, N/4$. During this interval, the relay processes $\mathsf{y}_r[n]$ by decoding an estimate $\hat{\mathsf{x}}_s[n]$ of the source transmitted signal.

Under a repetition-coded scheme, the relay transmits the signal

$$\mathsf{x}_r[n] = \hat{\mathsf{x}}_s[n - N/4]$$

for $n = N/4 + 1, \ldots, N/2$.

Decoding at the relay can take on a variety of forms. For example, the relay might fully decode, i.e., estimate without error, the entire source codeword, or it might employ symbol-by-symbol decoding and allow the destination to perform full decoding. These options allow for trading off performance and complexity at the relay terminal. Note that we focus on full decoding in the sequel; symbol-by-symbol decoding of binary transmissions has been treated from an uncoded perspective in [40]. Again, the destination can employ a variety of combining techniques; we focus in the sequel on a suitably modified matched filter.

B. Selection Relaying

As we might expect, and the analysis in Section IV confirms, fixed decode-and-forward is limited by direct transmission between the source and relay. However, since the fading coefficients are known to the appropriate receivers, $\mathsf{a}_{s,r}$ can be measured to high accuracy by the cooperating terminals; thus, they can adapt their transmission format according to the realized value of $\mathsf{a}_{s,r}$.

This observation suggests the following class of selection relaying algorithms. If the measured $|\mathsf{a}_{s,r}|^2$ falls below a certain threshold, the source simply continues its transmission to the destination, in the form of repetition or more powerful codes. If the measured $|\mathsf{a}_{s,r}|^2$ lies above the threshold, the relay forwards what it received from the source, using either amplify-and-forward or decode-and-forward, in an attempt to achieve diversity gain.

Informally speaking, selection relaying of this form should offer diversity because, in either case, two of the fading coefficients must be small in order for the information to be lost. Specifically, if $|\mathsf{a}_{s,r}|^2$ is small, then $|\mathsf{a}_{s,d}|^2$ must also be small for the information to be lost when the source continues its transmission. Similarly, if $|\mathsf{a}_{s,r}|^2$ is large, then both $|\mathsf{a}_{s,d}|^2$ and $|\mathsf{a}_{r,d}|^2$ must be small for the information to be lost when the relay employs amplify-and-forward or decode-and-forward. We formalize this notion when we consider outage performance of selection relaying in Section IV.

C. Incremental Relaying

As we will see, fixed and selection relaying can make inefficient use of the degrees of freedom of the channel, especially for high rates, because the relays repeat all the time. In this section, we describe incremental relaying protocols that exploit limited feedback from the destination terminal, e.g., a single bit indicating the success or failure of the direct transmission, that we will see can dramatically improve spectral efficiency over fixed and selection relaying. These incremental relaying protocols can be viewed as extensions of incremental redundancy, or hybrid automatic-repeat-request (ARQ), to the relay context. In ARQ, the source retransmits if the destination provides a negative acknowledgment via feedback; in incremental relaying, the relay retransmits in an attempt to exploit spatial diversity.

As one example, consider the following protocol utilizing feedback and amplify-and-forward transmission. We nominally allocate the channels according to Fig. 2(b). First, the source transmits its information to the destination at spectral efficiency $\mathtt{R}$. The destination indicates success or failure by broadcasting a single bit of feedback to the source and relay, which we assume is detected reliably by at least the relay.[5] If the source-destination SNR is sufficiently high, the feedback indicates success of the direct transmission, and the relay does nothing. If the source-destination SNR is not sufficiently high for successful direct transmission, the feedback requests that the relay amplify-and-forward what it received from the source. In the latter case, the destination tries to combine the two transmissions. As we will see, protocols of this form make more efficient use of the degrees of freedom of the channel, because they repeat only rarely. Incremental decode-and-forward is also possible; however, its analysis is more involved, and its performance is slightly worse than the above protocol.

IV. Outage Behavior

In this section, we characterize performance of the protocols of Section III in terms of outage events and outage probabilities [11]. To facilitate their comparison in the sequel, we also derive high-SNR approximations of the outage probabilities using results from Appendix I. For fixed fading realizations, the effective channel models induced by the protocols are variants of well-known channels with AWGN. As a function of the fading coefficients viewed as random variables, the mutual information for a protocol is a random variable denoted by I; in turn, for a target rate $\mathtt{R}$, $\mathsf{I} < \mathtt{R}$ denotes the outage event, and $\Pr[\mathsf{I} < \mathtt{R}]$ denotes the outage probability.

A. Direct Transmission

To establish baseline performance under direct transmission, the source terminal transmits over the channel (1). The maximum average mutual information between input and output in this case, achieved by independent and identically distributed (i.i.d.) zero-mean, circularly symmetric complex Gaussian inputs, is given by

$$\mathsf{I}_D = \log\left(1 + \mathtt{SNR}\,|\mathsf{a}_{s,d}|^2\right) \tag{10}$$

as a function of the fading coefficient $\mathsf{a}_{s,d}$. The outage event for spectral efficiency $\mathtt{R}$ is given by $\mathsf{I}_D < \mathtt{R}$ and is equivalent to the event

$$|\mathsf{a}_{s,d}|^2 < \frac{2^{\mathtt{R}} - 1}{\mathtt{SNR}}. \tag{11}$$

For Rayleigh fading, i.e., $|\mathsf{a}_{s,d}|^2$ exponentially distributed with parameter $\sigma_{s,d}^{-2}$, the outage probability satisfies[6]

$$\begin{aligned} p_D^{\text{out}}(\mathtt{SNR}, \mathtt{R}) &:= \Pr[\mathsf{I}_D < \mathtt{R}] = \Pr\left[|\mathsf{a}_{s,d}|^2 < \frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right] \\ &= 1 - \exp\left(-\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}\,\sigma_{s,d}^2}\right) \\ &\sim \frac{1}{\sigma_{s,d}^2}\cdot\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}} \end{aligned}$$

where we have utilized the result of Fact 1 in Appendix I with $\lambda = 1/\sigma_{s,d}^2$, $t = \mathtt{SNR}$, and $g(t) = (2^{\mathtt{R}} - 1)/t$.

B. Fixed Relaying

1) Amplify-and-Forward: The amplify-and-forward protocol produces an equivalent one-input, two-output complex Gaussian noise channel with different noise levels in the outputs. As explained in detail in Appendix II, the maximum average mutual information between the input and the two outputs, achieved by i.i.d. complex Gaussian inputs, is given by

$$\mathsf{I}_{AF} = \frac{1}{2}\log\left(1 + \mathtt{SNR}\,|\mathsf{a}_{s,d}|^2 + f\left(\mathtt{SNR}\,|\mathsf{a}_{s,r}|^2,\ \mathtt{SNR}\,|\mathsf{a}_{r,d}|^2\right)\right) \tag{12}$$

as a function of the fading coefficients, where

$$f(x,y) := \frac{xy}{x+y+1}. \tag{13}$$

We note that the amplifier gain β does not appear in (12), because the constraint (9) is met with equality.

The outage event for spectral efficiency $\mathtt{R}$ is given by $\mathsf{I}_{AF} < \mathtt{R}$ and is equivalent to the event

$$|\mathsf{a}_{s,d}|^2 + \frac{1}{\mathtt{SNR}} f\left(\mathtt{SNR}\,|\mathsf{a}_{s,r}|^2,\ \mathtt{SNR}\,|\mathsf{a}_{r,d}|^2\right) < \frac{2^{2\mathtt{R}} - 1}{\mathtt{SNR}}. \tag{14}$$

For Rayleigh fading, i.e., $|\mathsf{a}_{i,j}|^2$ independent and exponentially distributed with parameters $\sigma_{i,j}^{-2}$, analytic calculation of the outage probability becomes involved, but we can approximate its high-SNR behavior as

$$\begin{aligned} p_{AF}^{\text{out}}(\mathtt{SNR}, \mathtt{R}) &:= \Pr[\mathsf{I}_{AF} < \mathtt{R}] \\ &\sim \left(\frac{1}{2\sigma_{s,d}^2}\,\frac{\sigma_{s,r}^2 + \sigma_{r,d}^2}{\sigma_{s,r}^2\sigma_{r,d}^2}\right)\left(\frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}}\right)^2 \end{aligned}$$

where we have utilized the result of Claim 1 in Appendix I, with

$$\begin{aligned} \mathsf{u} = |\mathsf{a}_{s,d}|^2, \quad \mathsf{v} = |\mathsf{a}_{s,r}|^2, \quad \mathsf{w} = |\mathsf{a}_{r,d}|^2 \\ \lambda_{\mathsf{u}} = \sigma_{s,d}^{-2}, \quad \lambda_{\mathsf{v}} = \sigma_{s,r}^{-2}, \quad \lambda_{\mathsf{w}} = \sigma_{r,d}^{-2} \\ g(\epsilon) = (2^{2\mathtt{R}} - 1)\epsilon, \quad t = \mathtt{SNR}, \quad h(t) = 1/t\,. \end{aligned}$$

[5] Such an assumption is reasonable if the destination encodes the feedback bit with a very-low-rate code. Even if the relay cannot reliably decode, useful protocols can be developed and analyzed. For example, a conservative protocol might have the relay amplify-and-forward what it receives from the source in all cases except when the destination reliably receives the direct transmission *and* the relay reliably decodes the feedback bit.

[6] As we develop more formally in Appendix I, the approximation $f(\mathtt{SNR}) \sim g(\mathtt{SNR})$ for $\mathtt{SNR}$ large is in the sense of $f(\mathtt{SNR})/g(\mathtt{SNR}) \to 1$ as $\mathtt{SNR} \to \infty$.

2) Decode-and-Forward: To analyze decode-and-forward transmission, we examine a particular decoding structure at the relay. Specifically, we require the relay to fully decode the source message; examination of symbol-by-symbol decoding at the relay becomes involved because it depends upon the particular coding and modulation choices. The maximum average mutual information for repetition-coded decode-and-forward can be readily shown to be

$$\mathsf{I}_{DF} = \frac{1}{2}\min\{\log\left(1+\mathtt{SNR}\,|\mathsf{a}_{s,r}|^2\right), \log\left(1+\mathtt{SNR}\,|\mathsf{a}_{s,d}|^2+\mathtt{SNR}\,|\mathsf{a}_{r,d}|^2\right)\} \quad (15)$$

as a function of the fading random variables. The first term in (15) represents the maximum rate at which the relay can reliably decode the source message, while the second term in (15) represents the maximum rate at which the destination can reliably decode the source message given repeated transmissions from the source and destination. Requiring both the relay and destination to decode the entire codeword without error results in the minimum of the two mutual informations in (15). We note that such forms are typical of relay channels with full decoding at the relay [5].

The outage event for spectral efficiency $\mathtt{R}$ is given by $\mathsf{I}_{DF} < \mathtt{R}$ and is equivalent to the event

$$\min\left\{|\mathsf{a}_{s,r}|^2, |\mathsf{a}_{s,d}|^2+|\mathsf{a}_{r,d}|^2\right\} < \frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}}. \quad (16)$$

For Rayleigh fading, the outage probability for repetition-coded decode-and-forward can be computed according to

$$\begin{aligned} p_{DF}^{\text{out}}(\mathtt{SNR},\mathtt{R}) &:= \Pr\left[\mathsf{I}_{DF} < \mathtt{R}\right] \\ &= \Pr\left[|\mathsf{a}_{s,r}|^2 < g(\mathtt{SNR})\right] \\ &\quad + \Pr\left[|\mathsf{a}_{s,r}|^2 \geq g(\mathtt{SNR})\right]\Pr\left[|\mathsf{a}_{s,d}|^2+|\mathsf{a}_{r,d}|^2 < g(\mathtt{SNR})\right] \end{aligned} \quad (17)$$

where $g(\mathtt{SNR}) = [2^{2\mathtt{R}}-1]/\mathtt{SNR}$. Although we may readily compute a closed-form expression for (17), for compactness we examine the large $\mathtt{SNR}$ behavior of (17) by computing the limit

$$\begin{aligned} &\frac{1}{g(\mathtt{SNR})}p_{DF}^{\text{out}}(\mathtt{SNR},\mathtt{R}) \\ &= \underbrace{\frac{1}{g(\mathtt{SNR})}\Pr\left[|\mathsf{a}_{s,r}|^2 < g(\mathtt{SNR})\right]}_{\to 1/\sigma_{s,r}^2} \\ &+ \underbrace{\Pr\left[|\mathsf{a}_{s,r}|^2 \geq g(\mathtt{SNR})\right]}_{\to 1}\underbrace{\frac{1}{g(\mathtt{SNR})}\Pr\left[|\mathsf{a}_{s,d}|^2+|\mathsf{a}_{r,d}|^2 < g(\mathtt{SNR})\right]}_{\to 0} \\ &\to 1/\sigma_{s,r}^2 \end{aligned}$$

as $\mathtt{SNR}\to\infty$, using the results of Facts 1 and 2 in Appendix I. Thus, we conclude that

$$p_{DF}^{\text{out}}(\mathtt{SNR},\mathtt{R}) \sim \frac{1}{\sigma_{s,r}^2}\cdot\frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}}. \quad (18)$$

The $1/\mathtt{SNR}$ behavior in (18) indicates that fixed decode-and-forward does not offer diversity gains for large $\mathtt{SNR}$, because requiring the relay to fully decode the source information limits the performance of decode-and-forward to that of direct transmission between the source and relay.

C. Selection Relaying

To overcome the shortcomings of decode-and-forward transmission, we described selection relaying corresponding to adaptive versions of amplify-and-forward and decode-and-forward, both of which fall back to direct transmission if the relay cannot decode. We cannot conclude whether or not these protocols are optimal, because the capacities of general relay and related channels are long-standing open problems; however, as we will see, selection decode-and-forward enables the cooperating terminals to exploit full spatial diversity and overcome the limitations of fixed decode-and-forward.

As an example analysis, we determine the performance of selection decode-and-forward. Its mutual information is somewhat involved to write down in general; however, in the case of repetition coding at the relay, using (10) and (15), it can be readily shown to be

$$\mathsf{I}_{SDF} = \begin{cases} \frac{1}{2}\log\left(1+2\mathtt{SNR}|\mathsf{a}_{s,d}|^2\right), & |\mathsf{a}_{s,r}|^2 < g(\mathtt{SNR}) \\ \frac{1}{2}\log\left(1+\mathtt{SNR}[|\mathsf{a}_{s,d}|^2+|\mathsf{a}_{r,d}|^2]\right), & |\mathsf{a}_{s,r}|^2 \geq g(\mathtt{SNR}) \end{cases} \quad (19)$$

where $g(\mathtt{SNR}) = [2^{2\mathtt{R}}-1]/\mathtt{SNR}$. This threshold is motivated by our discussion of direct transmission, and is analogous to (11). The first case in (19) corresponds to the relay's not being able to decode and the source's repeating its transmission; here, the maximum average mutual information is that of repetition coding from the source to the destination, hence the extra factor of 2 in the SNR. The second case in (19) corresponds to the relay's ability to decode and repeat the source transmission; here, the maximum average mutual information is that of repetition coding from the source and relay to the destination.

The outage event for spectral efficiency $\mathtt{R}$ is given by $\mathsf{I}_{SDF} < \mathtt{R}$ and is equivalent to the event

$$\begin{aligned} &\left(\{|\mathsf{a}_{s,r}|^2 < g(\mathtt{SNR})\}\bigcap\{2|\mathsf{a}_{s,d}|^2 < g(\mathtt{SNR})\}\right) \\ &\bigcup\left(\{|\mathsf{a}_{s,r}|^2 \geq g(\mathtt{SNR})\}\bigcap\{|\mathsf{a}_{s,d}|^2+|\mathsf{a}_{r,d}|^2 < g(\mathtt{SNR})\}\right). \end{aligned} \quad (20)$$

The first (resp., second) event of the union in (20) corresponds to the first (resp., second) case in (19). We observe that adapting to the realized fading coefficient ensures that the protocol performs no worse than direct transmission, except for the fact that it potentially suffers the bandwidth inefficiency of repetition coding.

Because the events in the union of (20) are mutually exclusive, the outage probability becomes a sum

$$\begin{aligned} p_{SDF}^{\text{out}}(\mathtt{SNR},\mathtt{R}) &:= \Pr\left[\mathsf{I}_{SDF} < \mathtt{R}\right] \\ &= \Pr\left[|\mathsf{a}_{s,r}|^2 < g(\mathtt{SNR})\right]\Pr\left[2|\mathsf{a}_{s,d}|^2 < g(\mathtt{SNR})\right] \\ &\quad + \Pr\left[|\mathsf{a}_{s,r}|^2 \geq g(\mathtt{SNR})\right]\Pr\left[|\mathsf{a}_{s,d}|^2+|\mathsf{a}_{r,d}|^2 < g(\mathtt{SNR})\right] \end{aligned} \quad (21)$$

and we may readily compute a closed-form expression for (21). For comparison to our other protocols, we examine the large $\mathtt{SNR}$ behavior of (21) by computing the limit

$$\begin{aligned}
&\frac{1}{g^2(\mathtt{SNR})}\, p_{SDF}^{\mathrm{out}}(\mathtt{SNR},\mathtt{R}) \\
&= \underbrace{\frac{1}{g(\mathtt{SNR})}\Pr\left[|\mathsf{a}_{s,r}|^2 < g(\mathtt{SNR})\right]}_{\to 1/\sigma_{s,r}^2}\ \underbrace{\frac{1}{g(\mathtt{SNR})}\Pr\left[2|\mathsf{a}_{s,d}|^2 < g(\mathtt{SNR})\right]}_{\to 1/(2\sigma_{s,d}^2)} \\
&+ \underbrace{\Pr\left[|\mathsf{a}_{s,r}|^2 \geq g(\mathtt{SNR})\right]}_{\to 1}\ \underbrace{\frac{1}{g^2(\mathtt{SNR})}\Pr\left[|\mathsf{a}_{s,d}|^2 + |\mathsf{a}_{r,d}|^2 < g(\mathtt{SNR})\right]}_{\to 1/(2\sigma_{s,d}^2\sigma_{r,d}^2)} \\
&\to \left(\frac{1}{2\sigma_{s,d}^2}\frac{\sigma_{s,r}^2+\sigma_{r,d}^2}{\sigma_{s,r}^2\sigma_{r,d}^2}\right) \qquad (22)
\end{aligned}$$

as $\mathtt{SNR} \to \infty$, using the results of Facts 1 and 2 of Appendix I. Thus, we conclude that the large $\mathtt{SNR}$ performance of selection decode-and-forward is identical to that of fixed amplify-and-forward.

Analysis of more general selection relaying becomes involved because there are additional degrees of freedom in choosing the thresholds for switching between the various options such as direct, amplify-and-forward, and decode-and-forward. These issues represent a potentially useful direction for future research, but a detailed analysis of such protocols is beyond the scope of this paper.

D. Bounds for Cooperative Diversity

We now develop performance limits for fixed and selection relaying. If we suppose that the source and relay know each other's messages *a priori*, then instead of direct transmission, each would benefit from using a space–time code for two transmit antennas. In this sense, the outage probability of conventional transmit diversity [2]–[4] represents an optimistic lower bound on the outage probability of cooperative diversity. The following sections develop two such bounds: an unconstrained transmit diversity bound and an orthogonal transmit diversity bound that takes into account the half-duplex constraint.

1) Transmit Diversity Bound: To utilize a space–time code for each terminal, we allocate the channel as in Fig. 2(b). Both terminals transmit in all the degrees of freedom of the channel, so their transmitted power is $P/2$ J/2D, half that of direct transmission. The spectral efficiency for each terminal remains $\mathtt{R}$.

For transmit diversity, we model the channel as

$$\mathsf{y}_d[n] = \begin{bmatrix}\mathsf{a}_{s,d} & \mathsf{a}_{r,d}\end{bmatrix}\begin{bmatrix}\mathsf{x}_s[n]\\ \mathsf{x}_r[n]\end{bmatrix} + \mathsf{z}_d[n] \qquad (23)$$

for, say, $n = 0,\ldots,N/2$. As developed in Appendix III, an optimal signaling strategy, in terms of minimizing outage probability in the large $\mathtt{SNR}$ regime, is to encode information using $[\mathsf{x}_s \quad \mathsf{x}_r]^{\mathrm{T}}$ i.i.d. complex Gaussian, each with power $P/2$. Using this result, the maximum average mutual information as a function of the fading coefficients is given by

$$\mathsf{I}_T = \log\left(1 + \frac{\mathtt{SNR}}{2}\left[|\mathsf{a}_{s,d}|^2 + |\mathsf{a}_{r,d}|^2\right]\right). \qquad (24)$$

The outage event $\mathsf{I}_T < \mathtt{R}$ is equivalent to the event

$$|\mathsf{a}_{s,d}|^2 + |\mathsf{a}_{r,d}|^2 < \frac{2^{\mathtt{R}}-1}{(\mathtt{SNR}/2)}. \qquad (25)$$

For $|\mathsf{a}_{i,j}|^2$ exponentially distributed with parameters $\sigma_{i,j}^{-2}$, the outage probability satisfies

$$\begin{aligned}
p_T^{\mathrm{out}}(\mathtt{SNR},\mathtt{R}) &:= \Pr\left[\mathsf{I}_T < \mathtt{R}\right] \\
&\sim \frac{2}{\sigma_{s,d}^2\sigma_{r,d}^2}\cdot\left(\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right)^2 \qquad (26)
\end{aligned}$$

where we have applied the result of Fact 2 in Appendix I.

2) Orthogonal Transmit Diversity Bound: The transmit diversity bound (26) does not take into account the half-duplex constraint. To capture this effect, we constrain the transmit diversity scheme to be orthogonal.

When the source and relay can cooperate perfectly, an equivalent model to (23), incorporating the relay orthogonality constraint, consists of parallel channels

$$\mathsf{y}_d[n] = \mathsf{a}_{s,d}\,\mathsf{x}_s[n] + \mathsf{z}_d[n], \quad n = 0,\ldots,N/4 \qquad (27)$$

$$\mathsf{y}_d[n] = \mathsf{a}_{r,d}\,\mathsf{x}_r[n] + \mathsf{z}_d[n], \quad n = N/4+1,\ldots,N/2. \qquad (28)$$

This pair of parallel channels is utilized half as many times as the corresponding direct transmission channel, so the source must transmit at twice the spectral efficiency in order to achieve the same spectral efficiency as direct transmission.

For each fading realization, the maximum average mutual information can be obtained using independent complex Gaussian inputs. Allocating a fraction α of the power to x_s, and the remaining fraction $\overline{\alpha}:=(1-\alpha)$ of the power to x_r, the average mutual information is given by

$$\mathsf{I}_P = \frac{1}{2}\log\left[\left(1+2\alpha\mathtt{SNR}|\mathsf{a}_{s,d}|^2\right)\left(1+2\overline{\alpha}\mathtt{SNR}|\mathsf{a}_{r,d}|^2\right)\right]. \qquad (29)$$

The outage event $\mathsf{I}_P < \mathtt{R}$ is equivalent to the outage region

$$\alpha|\mathsf{a}_{s,d}|^2 + \overline{\alpha}|\mathsf{a}_{r,d}|^2 + 2\alpha\overline{\alpha}\mathtt{SNR}|\mathsf{a}_{s,d}|^2|\mathsf{a}_{r,d}|^2 < \frac{2^{2\mathtt{R}}-1}{2\,\mathtt{SNR}}. \qquad (30)$$

As in the case of amplify-and-forward, analytical calculation of the outage probability becomes involved; however, we can approximate its high-SNR behavior for Rayleigh fading as

$$\begin{aligned}
p_P^{\mathrm{out}}(\mathtt{SNR},\mathtt{R}) &:= \Pr\left[\mathsf{I}_P < \mathtt{R}\right] \\
&\sim \frac{1}{4\alpha\overline{\alpha}\sigma_{s,d}^2\sigma_{r,d}^2}\cdot\frac{2^{2\mathtt{R}}[2\mathtt{R}\ln(2)-1]+1}{\mathtt{SNR}^2} \qquad (31)
\end{aligned}$$

using the result of Claim 2 in Appendix I, with

$$\begin{aligned}
\mathsf{u} &= \alpha|\mathsf{a}_{s,d}|^2, \quad \mathsf{v} = \overline{\alpha}|\mathsf{a}_{r,d}|^2 \\
\lambda_{\mathsf{u}} &= 1/(\alpha\sigma_{s,d}^2), \quad \lambda_{\mathsf{v}} = 1/(\overline{\alpha}\sigma_{r,d}^2) \\
\epsilon &= [2^{2\mathtt{R}}-1]/(2\,\mathtt{SNR}), \quad t = 2^{2\mathtt{R}}-1.
\end{aligned}$$

Clearly, (31) is minimized for $\alpha = 1/2$, yielding

$$p_P^{\mathrm{out}}(\mathtt{SNR},\mathtt{R}) \sim \frac{1}{\sigma_{s,d}^2\sigma_{r,d}^2}\cdot\frac{2^{2\mathtt{R}}[2\mathtt{R}\ln(2)-1]+1}{\mathtt{SNR}^2} \qquad (32)$$

so that i.i.d. complex Gaussian inputs again minimize outage probability for large $\mathtt{SNR}$. Note that for $\mathtt{R} \to 0$, (32) converges to (26), the transmit diversity bound without orthogonality constraints. Thus, the orthogonality constraint has little effect for small $\mathtt{R}$, but induces a loss in $\mathtt{SNR}$ proportional to

$$\sqrt{\mathtt{R}\ln(2)}$$

with respect to the unconstrained transmit diversity bound for large $\mathtt{R}$.

E. Incremental Relaying

Outage analysis of incremental relaying is complicated by its variable-rate nature. Specifically, the protocols operate at spectral efficiency $\mathtt{R}$ when the source-destination transmission is successful, and spectral efficiency $\mathtt{R}/2$ when the relay repeats the source transmission. Thus, we examine outage probability as a function of $\mathtt{SNR}$ and the *expected* spectral efficiency $\overline{\mathtt{R}}$.

For incremental amplify-and-forward, the outage probability as a function of $\mathtt{SNR}$ and $\mathtt{R}$ is given by

$$\begin{aligned} p_{IAF}^{\text{out}}&(\mathtt{SNR},\mathtt{R}) \\ &= \Pr[\mathsf{I}_D < \mathtt{R}] \Pr[\mathsf{I}_{AF} < \mathtt{R}/2 | \mathsf{I}_D < \mathtt{R}] \\ &= \Pr[\mathsf{I}_{AF} < \mathtt{R}/2] \\ &= \Pr\left[|\mathsf{a}_{s,d}|^2 + \frac{1}{\mathtt{SNR}} f(\mathtt{SNR}|\mathsf{a}_{s,r}|^2, \mathtt{SNR}|\mathsf{a}_{r,d}|^2) < g(\mathtt{SNR})\right] \end{aligned} \tag{33}$$

where I_D and I_{AF} are given by (10) and (12), respectively,

$$g(\mathtt{SNR}) = [2^{\mathtt{R}} - 1]/\mathtt{SNR}$$

and $f(\cdot,\cdot)$ is given in (13). The second equality follows from the fact that the intersection of the direct and amplify-and-forward outage events is exactly the amplify-and-forward outage event at half the rate. Furthermore, the expected spectral efficiency can be computed as

$$\begin{aligned} \overline{\mathtt{R}} &= \mathtt{R}\Pr\left[|\mathsf{a}_{s,d}|^2 \geq \frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right] + \frac{\mathtt{R}}{2}\Pr\left[|\mathsf{a}_{s,d}|^2 < \frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right] \\ &= \mathtt{R}\exp\left(-\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right) + \frac{\mathtt{R}}{2}\left[1 - \exp\left(-\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right)\right] \\ &= \frac{\mathtt{R}}{2}\left[1 + \exp\left(-\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right)\right] := h_{\mathtt{SNR}}(\mathtt{R}) \end{aligned} \tag{34}$$

where the second equality follows from substituting standard exponential results for $|\mathsf{a}_{s,d}|^2$.

An important question is the value of $\mathtt{R}$ to employ in (33) for a given expected spectral efficiency $\overline{\mathtt{R}}$. A fixed value of $\overline{\mathtt{R}}$ can arise from several possible $\mathtt{R}$, depending upon the value of $\mathtt{SNR}$; thus, we see that the pre-image $h_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}})$ can contain several points. We define a function $\tilde{h}_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}}) := \min h_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}})$ to capture a useful mapping from $\overline{\mathtt{R}}$ to $\mathtt{R}$; for a given value of $\overline{\mathtt{R}}$, it seems clear from the outage expression (33) that we want the smallest $\mathtt{R}$ possible.

For fair comparison to protocols without feedback, we characterize a modified outage expression in the large-SNR regime. Specifically, we compare outage of fixed and selection relaying protocols to the modified outage $p_{IAF}^{\text{out}}(\mathtt{SNR}, \tilde{h}_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}}))$. For large $\mathtt{SNR}$, we have

$$p_{IAF}^{\text{out}}\left(\mathtt{SNR}, \tilde{h}_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}})\right) \sim \left(\frac{1}{2\sigma_{s,d}^2}\frac{\sigma_{s,r}^2 + \sigma_{r,d}^2}{\sigma_{s,r}^2\sigma_{r,d}^2}\right)\cdot\left(\frac{2^{\overline{\mathtt{R}}}-1}{\mathtt{SNR}}\right)^2 \tag{35}$$

where we have combined the results of Claims 1 and 3 in Appendix I.

Bounds for incremental relaying can be obtained by suitably normalizing the results developed in Section IV-D; however, we stress that treating protocols that exploit more general feedback, along with their associated performance limits, is beyond the scope of this paper.

V. Discussion

In this section, we compare the outage results developed in Section IV in various regimes. To partition the discussion, and make it clear in which context certain observations hold, we divide the exposition into two sections. Section V-A considers general asymmetric networks in which the fading variances $\sigma_{i,j}^2$ may be distinct. The primary observations of this section are comparing the performance of the cooperative protocols to the transmit diversity bound and examining how spectral efficiency and network asymmetry affect that comparison. Section V-B focuses on the special case of symmetric networks in which the fading variances are identical, e.g., ${\sigma_{i,j}}^2 = 1$, without loss of generality. Focusing on this special case allows us to substantially simplify the exposition for comparison of performance under different parameterizations, e.g., $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$ and $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$.

A. Asymmetric Networks

As indicated by the results in Section IV, for fixed rates, simple protocols such as fixed amplify-and-forward, selection decode-and-forward, and incremental amplify-and-forward each achieve full (i.e., second-order) diversity: their outage probability decays proportional to $1/\mathtt{SNR}^2$ (cf. (15), (22), and (35)). We now compare these protocols to the transmit diversity bound, discuss the impacts of spectral efficiency and network geometry on performance, and examine their outage events.

1) Comparison to Transmit Diversity Bound: Equating performance of amplify-and-forward (15) or selection decode-and-forward (22) to the transmit diversity bound (26), we can determine the relative SNR losses of cooperative diversity. In the low-spectral-efficiency regime, the protocols without feedback are within a factor of

$$\left[\frac{2^{2\mathtt{R}}-1}{2(2^{\mathtt{R}}-1)}\right]\sqrt{1+\left(\frac{\sigma_{r,d}^2}{\sigma_{s,r}^2}\right)} \approx \sqrt{1+\left(\frac{\sigma_{r,d}^2}{\sigma_{s,r}^2}\right)}$$

in SNR from the transmit diversity bound, suggesting that the powerful benefits of multiple-antenna systems can indeed be obtained without the need for physical arrays. For statistically symmetric networks, e.g., $\sigma_{i,j}^2 = 1$, the loss is only $\sqrt{2}$ or 1.5 dB; more generally, the loss decreases as the source-relay path improves relative to the relay-destination path.

For larger spectral efficiencies, fixed and selection relaying lose an additional 3 dB per transmitted bit per second per hertz (bit/s/Hz) with respect to the transmit diversity bound. This additional loss is due to two factors: the half-duplex constraint and the repetition-coded nature of the protocols. As suggested by Fig. 3, of the two, repetition coding appears to be the more significant source of inefficiency in our protocols. In Fig. 3, the SNR loss of orthogonal transmit diversity with respect to unconstrained transmit diversity is intended to indicate the cost of the half-duplex constraint, and the loss of our cooperative diversity protocols with respect to the transmit diversity bound indicates the cost of both imposing the half-duplex constraint and employing repetition-like codes. The figure suggests that, although

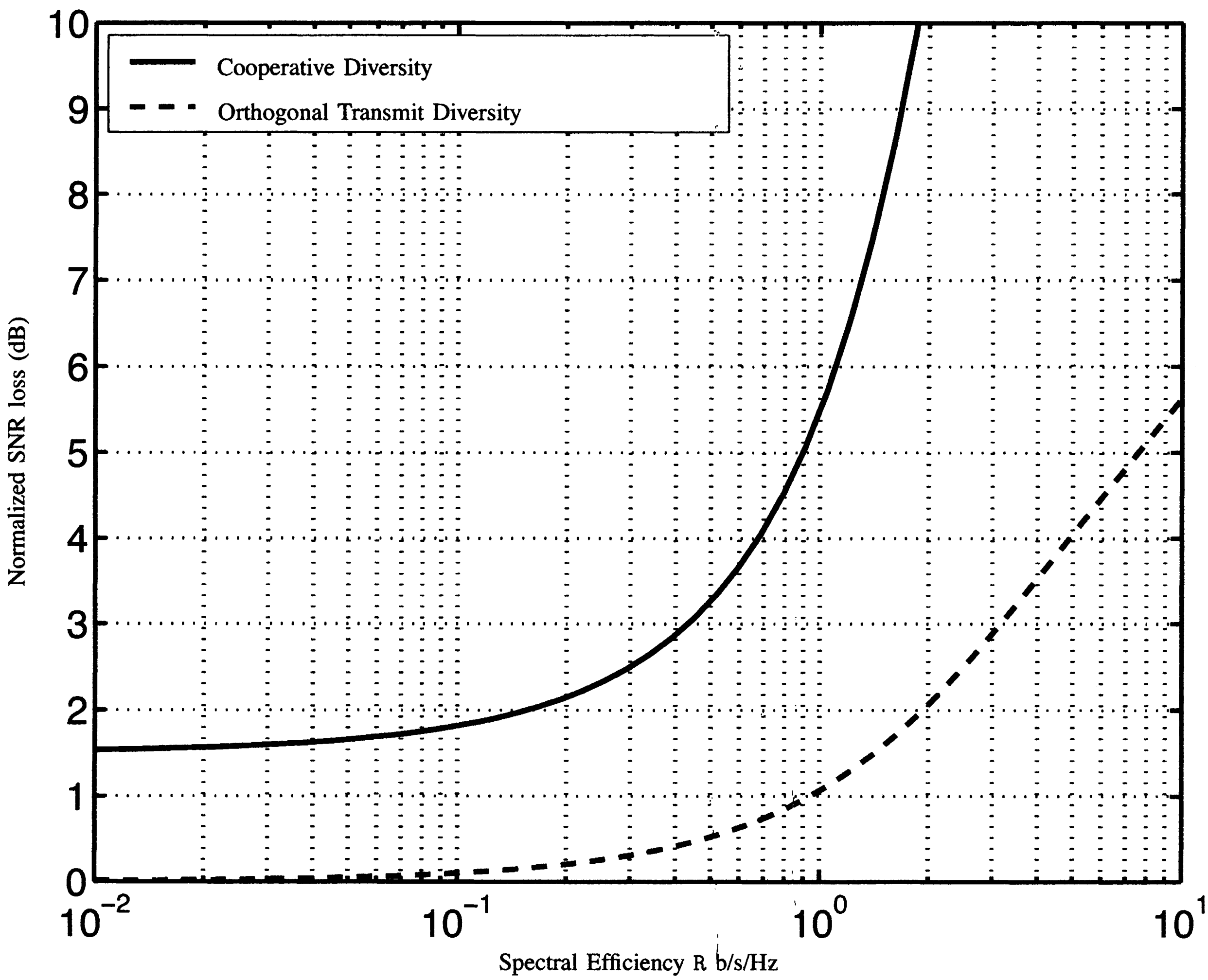

Fig. 3. SNR loss for cooperative diversity protocols (solid) and orthogonal transmit diversity bound (dashed) relative to the (unconstrained) transmit diversity bound.

the half-duplex constraint contributes, "repetition" in the form of amplification or repetition coding is the major cause of SNR loss for high rates. By contrast, incremental amplify-and-forward overcomes these additional losses by repeating only when necessary.

2) Outage Events: It is interesting that amplify-and-forward and selection decode-and-forward have the same high-SNR performance, especially considering the different shapes of their outage events (cf. (14), (20)), which are shown in the low-spectral-efficiency regime in Fig. 4. When the relay can fully decode the source message and repeat it, i.e., $\mathtt{SNR}_{\mathrm{norm}}|a_{s,r}|^2 \geq 2$, the outage event for selection decode-and-forward is a strict subset of the outage event of amplify-and-forward, with amplify-and-forward approaching that of selection decode-and-forward as $|a_{s,r}|^2 \to \infty$. On the other hand, when the relay cannot fully decode the source message and the source repeats, i.e., $\mathtt{SNR}_{\mathrm{norm}}|a_{s,r}|^2 < 2$, the outage event of amplify-and-forward is neither a subset nor a superset of the outage event for selection decode-and-forward. Apparently, averaging over the Rayleigh-fading coefficients eliminates the differences between amplify-and-forward and selection decode-and-forward, at least in the high-SNR regime.

3) Effects of Geometry: To isolate the effect of network geometry on performance, we compare the high-SNR behavior of direct transmission (12) with that of incremental amplify-and-forward (35). Comparison with fixed and selection relaying is similar, except for the additional impact of SNR loss with increasing spectral efficiency. Using a common model for the path-loss (fading variances), we set $\sigma_{i,j}^2 \propto d_{i,j}^{-\alpha}$, where $d_{i,j}$ isthe distance between terminals i and j, and α is the path-loss exponent [7]. Under this model, comparing (12) with (35), assuming both approximations are good for the $\mathtt{SNR}$ of interest, we prefer incremental amplify-and-forward whenever

$$\left(\frac{d_{s,r}}{d_{s,d}}\right)^{\alpha} + \left(\frac{d_{r,d}}{d_{s,d}}\right)^{\alpha} < 2\,\mathtt{SNR}_{\mathrm{norm}}. \tag{36}$$

Thus, incremental amplify-and-forward is useful whenever the relay lies within a certain normalized ellipse having the source and destination as its foci, with the size of the ellipse increasing in $\mathtt{SNR}_{\mathrm{norm}}$. What is most interesting about the structure of this "utilization region" for incremental amplify-and-forward is that it is symmetric with respect to the source and destination. By comparison, a certain circle about only the source gives the utilization region for fixed decode-and-forward.

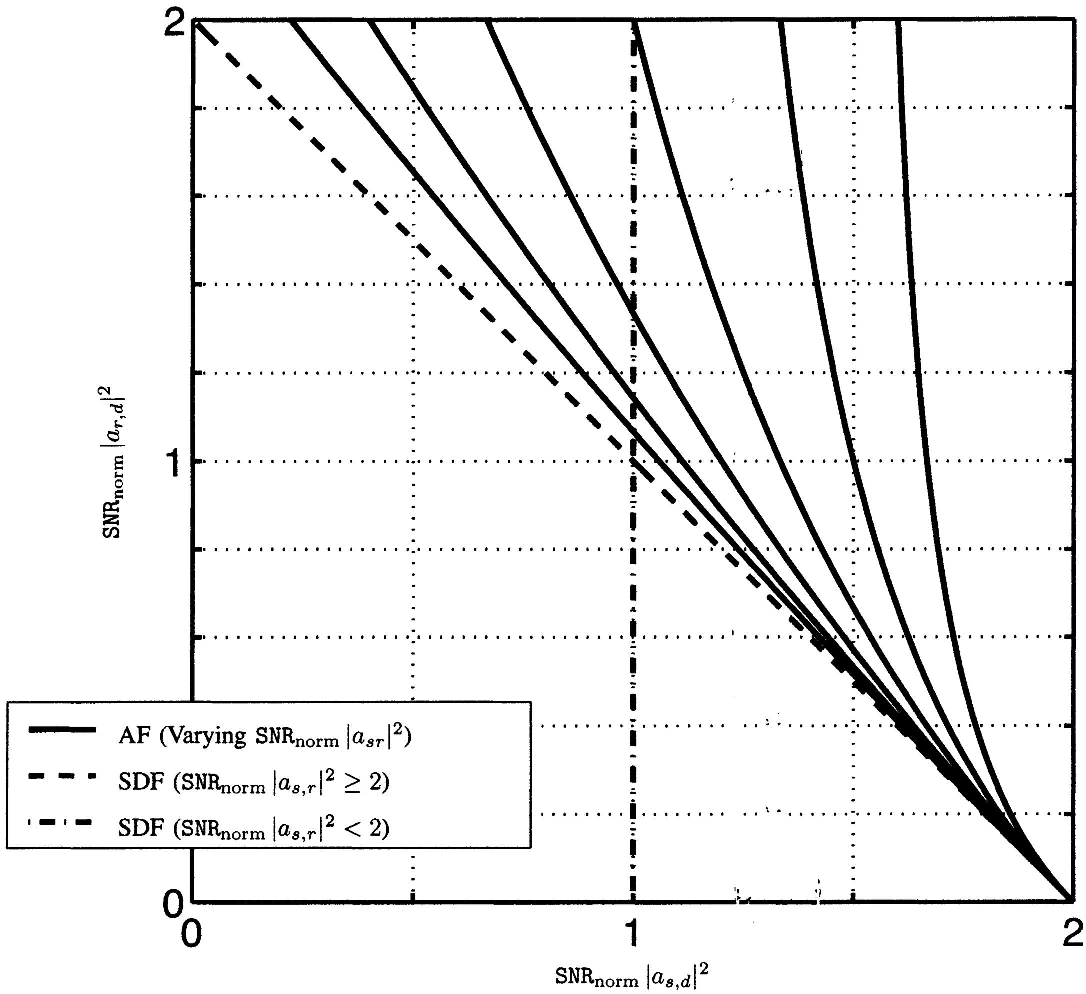

Fig. 4. Outage event boundaries for amplify-and-forward (solid) and selection decode-and-forward (dashed and dash-dotted) as functions of the realized fading coefficient $|a_{s,r}|^2$ between the cooperating terminals. Outage events are to the left and below the respective outage event boundaries. Successively lower solid curves correspond to amplify-and-forward with increasing values of $|a_{s,r}|^2$. The dashed curve corresponds to the outage event for selection decode-and-forward when the relay can fully decode and the relay repeats, i.e., $\mathtt{SNR}_{\mathrm{norm}}|a_{s,r}|^2 \geq 2$, while the dash-dotted curve corresponds to the outage event of selection decode-and-forward when the relay cannot fully decode and the source repeats, i.e., $\mathtt{SNR}_{\mathrm{norm}}|a_{s,r}|^2 < 2$. Note that the dash-dotted curve also corresponds to the outage event for direct transmission.

Utilization regions of the form (36) may be useful in developing higher layer network protocols that select between direct transmission and cooperative diversity using one of a number of potential relays. Such algorithms and their performance represent an interesting area of further research, and a key ingredient for fully incorporating cooperative diversity into wireless networks.

B. Symmetric Networks

We now specialize all of our results to the case of statistically symmetric networks, e.g., $\sigma_{i,j}^2 = 1$ without loss of generality. We develop the results, summarized in Table I, under the two parameterizations $(\mathtt{SNR}_{\mathrm{norm}}, \mathtt{R})$ and $(\mathtt{SNR}, \mathtt{R}_{\mathrm{norm}})$, respectively.

1) Results Under Different Parameterizations: Parameterizing the outage results from Section IV in terms of $(\mathtt{SNR}_{\mathrm{norm}}, \mathtt{R})$ is straightforward because $\mathtt{R}$ remains fixed; we simply substitute $\mathtt{SNR} = \mathtt{SNR}_{\mathrm{norm}}(2^{\mathtt{R}} - 1)$ to obtain the results listed in the second column of Table I. Parameterizing the outage results from Section IV in terms of $(\mathtt{SNR}, \mathtt{R}_{\mathrm{norm}})$ is a bit more involved because $\mathtt{R} = \mathtt{R}_{\mathrm{norm}} \log(1 + \mathtt{SNR})$ increases with $\mathtt{SNR}$.

The results in Appendix I are all general enough to allow this particular parameterization. To demonstrate their application, we consider amplify-and-forward. The outage event under this alternative parameterization is given by

$$|\mathsf{a}_{s,d}|^2 + \frac{1}{\mathtt{SNR}} f\left(\mathtt{SNR}\,|\mathsf{a}_{s,r}|^2, \mathtt{SNR}\,|\mathsf{a}_{r,d}|^2\right) < \frac{(1+\mathtt{SNR})^{2\mathtt{R}_{\mathrm{norm}}} - 1}{\mathtt{SNR}}.$$

For $\mathtt{R}_{\mathrm{norm}} < 1/2$, the outage probability is approximately

$$p_{AF}^{\mathrm{out}}(\mathtt{SNR}, \mathtt{R}_{\mathrm{norm}}) \sim \left[\frac{\mathtt{SNR}}{(1+\mathtt{SNR})^{2\mathtt{R}_{\mathrm{norm}}} - 1}\right]^{-2} \sim 1/\mathtt{SNR}^{2(1-2\mathtt{R}_{\mathrm{norm}})}$$

where we have utilized the results of Claim 1 in Appendix I with

$$\mathsf{u} = |\mathsf{a}_{s,d}|^2, \quad \mathsf{v} = |\mathsf{a}_{s,r}|^2, \quad \mathsf{w} = |\mathsf{a}_{r,d}|^2, \quad \lambda_{\mathsf{u}} = \lambda_{\mathsf{v}} = \lambda_{\mathsf{w}} = 1$$
$$g(\epsilon) = \epsilon\left[(1 + 1/\epsilon)^{2\mathtt{R}_{\mathrm{norm}}} - 1\right], \quad t = \mathtt{SNR}, \quad h(t) = 1/t.$$

The other results listed in the third column of Table I can be obtained in similar fashion using the appropriate results from Appendix I.

TABLE I
SUMMARY OF OUTAGE PROBABILITY APPROXIMATIONS FOR STATISTICALLY SYMMETRIC NETWORKS

Protocol	$p^{\text{out}}(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$, high $\mathtt{SNR}_{\text{norm}}$	$p^{\text{out}}(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$, high $\mathtt{SNR}$
Direct	$1/\mathtt{SNR}_{\text{norm}}$	$1/\mathtt{SNR}^{(1-\mathtt{R}_{\text{norm}})}$
Amplify-and-Forward	$(2^{\mathtt{R}}+1)^2/\mathtt{SNR}_{\text{norm}}^2$	$1/\mathtt{SNR}^{2(1-2\mathtt{R}_{\text{norm}})}$
Decode-and-Forward	$(2^{\mathtt{R}}+1)/\mathtt{SNR}_{\text{norm}}$	$1/\mathtt{SNR}^{(1-2\mathtt{R}_{\text{norm}})}$
Selection Decode-and-Forward	$(2^{\mathtt{R}}+1)^2/\mathtt{SNR}_{\text{norm}}^2$	$1/\mathtt{SNR}^{2(1-2\mathtt{R}_{\text{norm}})}$
Incremental Amplify-and-Forward	$1/\mathtt{SNR}_{\text{norm}}^2$	$1/\mathtt{SNR}^{2(1-\mathtt{R}_{\text{norm}})}$
Transmit Diversity Bound	$2/\mathtt{SNR}_{\text{norm}}^2$	$2/\mathtt{SNR}^{2(1-\mathtt{R}_{\text{norm}})}$
Orthogonal Transmit Diversity Bound	$\left(\frac{2^{2\mathtt{R}}[2\mathtt{R}\ln(2)-1]+1}{(2^{\mathtt{R}}-1)^2}\right)/\mathtt{SNR}_{\text{norm}}^2$	$2[\mathtt{R}_{\text{norm}}\ln(\mathtt{SNR})+1]/\mathtt{SNR}^{2(1-\mathtt{R}_{\text{norm}})}$

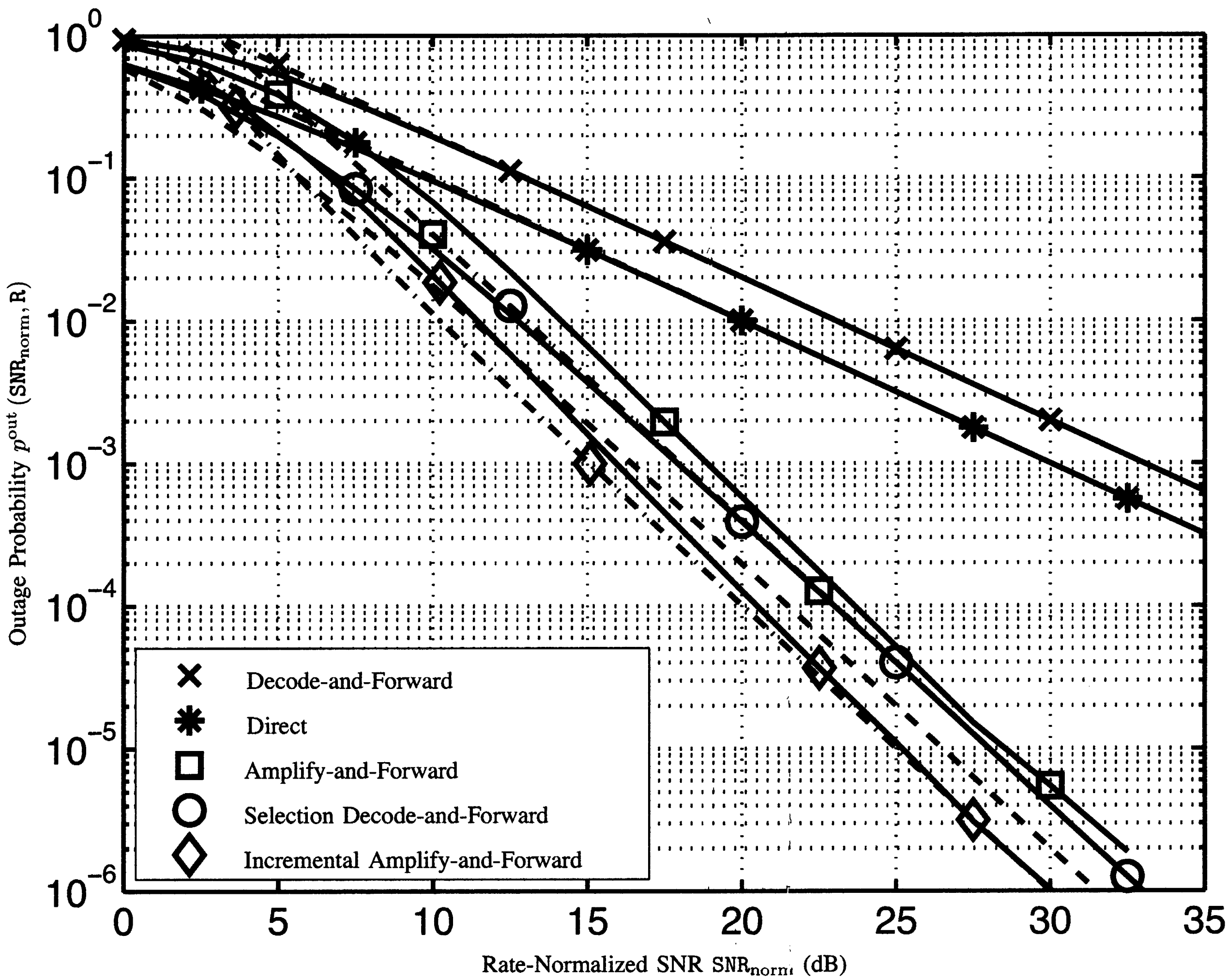

Fig. 5. Outage probabilities versus $\mathtt{SNR}_{\text{norm}}$, small $\mathtt{R}$ regime, for statistically symmetric networks, i.e., $\sigma_{i,j}^2 = 1$. The outage probability curve for amplify-and-forward was obtained via Monte Carlo simulation, while the other curves are computed from analytical expressions. Solid curves correspond to exact outage probabilities, while dash-dotted curves correspond to the high-SNR approximations from Table I. The dashed curve corresponds to the transmit diversity bounds in this low spectral efficiency regime.

2) Fixed $\mathtt{R}$ *Systems:* Fig. 5 shows outage probabilities for the various protocols as functions of $\mathtt{SNR}_{\text{norm}}$ in the small, fixed $\mathtt{R}$ regime. Both exact and high-SNR approximations are displayed, demonstrating the wide range of SNRs over which the high-SNR approximations are useful. The diversity gains of our protocols appear as steeper slopes in Fig. 5, from a factor of 10 decrease in outage probability for each additional 10 dB of SNR in the case of direct transmission, to a factor of 100 decrease in outage probability for each additional 10 dB of SNR in the case of cooperative diversity. The relative loss of 1.5 dB for fixed amplify-and-forward and selection decode-and-forward with respective to the transmit diversity bound is also apparent. The curves for fixed and selection relaying shift to the right by 3 dB for each additional bit/s/Hz of spectral efficiency in the high $\mathtt{R}$ regime. By contrast, the performance of incremental amplify-and-forward is unchanged at high SNR for increasing $\mathtt{R}$. Note that, at outage probabilities on the order of 10^{-3}, cooperative diversity achieves large energy savings over direct transmission—on the order of 12–15 dB.

3) Fixed $\mathtt{R}_{\text{norm}}$ *Families of Systems:* Another way to examine the high spectral efficiency regime as SNR becomes large is to allow $\mathtt{R}$ to grow with increasing $\mathtt{SNR}$. In particular, the choice of $\mathtt{R} = \mathtt{R}_{\text{norm}} \log(1 + \mathtt{SNR})$ is a natural one: for slower growth, the outage results essentially behave like fixed $\mathtt{R}$ systems for sufficiently large $\mathtt{SNR}$, while for faster growth, the outage probabilities all tend to 1. These observations motivate our parameterization in terms of $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$.

Parameterizing performance in terms of $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$ leads to interesting tradeoffs between the diversity order and normalized spectral efficiency of a protocol. Because these tradeoffs arise naturally in the context of multiple-antenna systems [9], [10], it is not surprising that they show up in the context of cooperative diversity. Diversity order can be viewed as the power to which $\mathtt{SNR}^{-1}$ is raised in our outage expressions in the third column of Table I. To be precise, we can define diversity order as

$$\Delta(\mathtt{R}_{\text{norm}}) := \lim_{\mathtt{SNR}\to\infty} \frac{-\log p^{\text{out}}(\mathtt{SNR}, \mathtt{R}_{\text{norm}})}{\log \mathtt{SNR}}. \tag{37}$$

Larger $\Delta(\mathtt{R}_{\text{norm}})$ implies more robustness to fading (faster decay in the outage probability with increasing SNR), but $\Delta(\mathtt{R}_{\text{norm}})$ generally decreases with increasing $\mathtt{R}_{\text{norm}}$. For example, the diversity order of amplify-and-forward is $\Delta_{AF}(\mathtt{R}_{\text{norm}}) = 2(1 - 2\mathtt{R}_{\text{norm}})$; thus, its maximum diversity order 2 is achieved as $\mathtt{R}_{\text{norm}} \to 0$, and maximum normalized spectral efficiency $1/2$ is achieved as $\Delta_{AF} \to 0$. Fig. 6 compares the tradeoffs for direct transmission and cooperative diversity. As we might expect from our previous discussion, among the protocols developed in this paper, incremental amplify-and-forward yields the highest $\Delta(\mathtt{R}_{\text{norm}})$ for each $\mathtt{R}_{\text{norm}}$; this curve also corresponds to the transmit diversity bound in the high-SNR regime. What is most interesting about the results in Fig. 6 is the sharp transition at $\mathtt{R}_{\text{norm}} = 1/3$ between our preference for amplify-and-forward (as well as selection decode-and-forward) for $\mathtt{R}_{\text{norm}} < 1/3$ and our preference for direct transmission for $\mathtt{R}_{\text{norm}} > 1/3$.

VI. Concluding Remarks and Future Directions

We develop in this paper a variety of low-complexity, cooperative protocols that enable a pair of wireless terminals, each with a single antenna, to fully exploit spatial diversity in the channel. These protocols blend different fixed relaying modes, specifically amplify-and-forward and decode-and-forward, with strategies based upon adapting to CSI between cooperating source terminals (selection relaying) as well as exploiting limited feedback from the destination terminal (incremental relaying). For delay-limited and nonergodic environments, we analyze the outage probability performance, in many cases exactly, and in all cases using accurate, high-SNR approximations.

There are costs associated with our cooperative protocols. For one thing, cooperation with half-duplex operation requires twice the bandwidth of direct transmission for a given rate, and leads to larger effective SNR losses for increasing spectral efficiency. Furthermore, depending upon the application, additional receive hardware may be required in order for the sources to relay for one another. Although this may not be the case in emerging *ad hoc* or multihop cellular networks, it would be the case in the uplink of current cellular systems that employ frequency-division duplexing. Finally, although our analysis has not explicitly taken it into account, there may be additional power costs of relays operating instead of powering down. Despite these costs, our analysis demonstrates significant performance enhancements, particularly in the low-spectral-efficiency regime (up to roughly 1 bit/s/Hz) often found in practice. Like other forms of diversity, these performance enhancements take the form of decreased transmit power for the same reliability, increased reliability for the same transmit power, or some combination of the two.

The observations in Section V-B suggest that, among other issues, a key area of further research is exploring cooperative diversity protocols in the high-spectral-efficiency regime. It remains unclear at this point whether our simple protocols are close to optimal in this regime, among all possible cooperative diversity protocols, yet our results indicate that direct transmission eventually becomes preferable. Useful work in this area would develop tighter lower bounds on performance, which is akin to developing tighter converses for the relay channel [5], or demonstrating other protocols that are more efficient for high spectral efficiencies. Some of our own work in this direction appears in [37].

More broadly, there are a number of channel circumstances in addition to those considered here that warrant further investigation. In particular, for scenarios in which the transmitters obtain accurate knowledge of the channel realizations, via feedback or other means, beamforming and power and bandwidth allocation become possible. These options allow the cooperating terminals to adapt to their specific channel conditions and geometry and select appropriate coding schemes for various regimes. Again, better understanding of the relay channel will continue to yield insight on these problems.

We note that we have focused on the case of a pair of terminals cooperating; extension to more than two terminals is straightforward except for the fact that comparatively more options arise. For example, in the case of three cooperating terminals, one of the relays might amplify-and-forward the information, while the other relay might decode-and-forward the information, or *vice versa*. Moreover, as the number of terminals forming a network grows, higher layer protocols for organizing terminals into cooperating groups become increasingly important. Some preliminary work in this direction is reported in [38]. Finally, because cooperative diversity is inherently a network problem, it could be fruitful to take into account additional higher layer network issues such as queuing of bursty data, link layer retransmissions, and routing.

Appendix I
Asymptotic CDF Approximations

To keep the presentation in the main part of the paper concise, we collect in this appendix several results for the limiting behavior of the cumulative distribution function (CDF) of certain combinations of exponential random variables. All our results are of the form

$$\lim_{t\to t_0} \frac{P_{\mathsf{u}(t)}(g_1(t))}{g_2(t)} = c \tag{38}$$

where t is a parameter of interest; $P_{\mathsf{u}(t)}(g_1(t))$ is the CDF of a certain random variable $\mathsf{u}(t)$ that can, in general, depend upon t; $g_1(t)$ and $g_2(t)$ are two (continuous) functions; and t_0 and

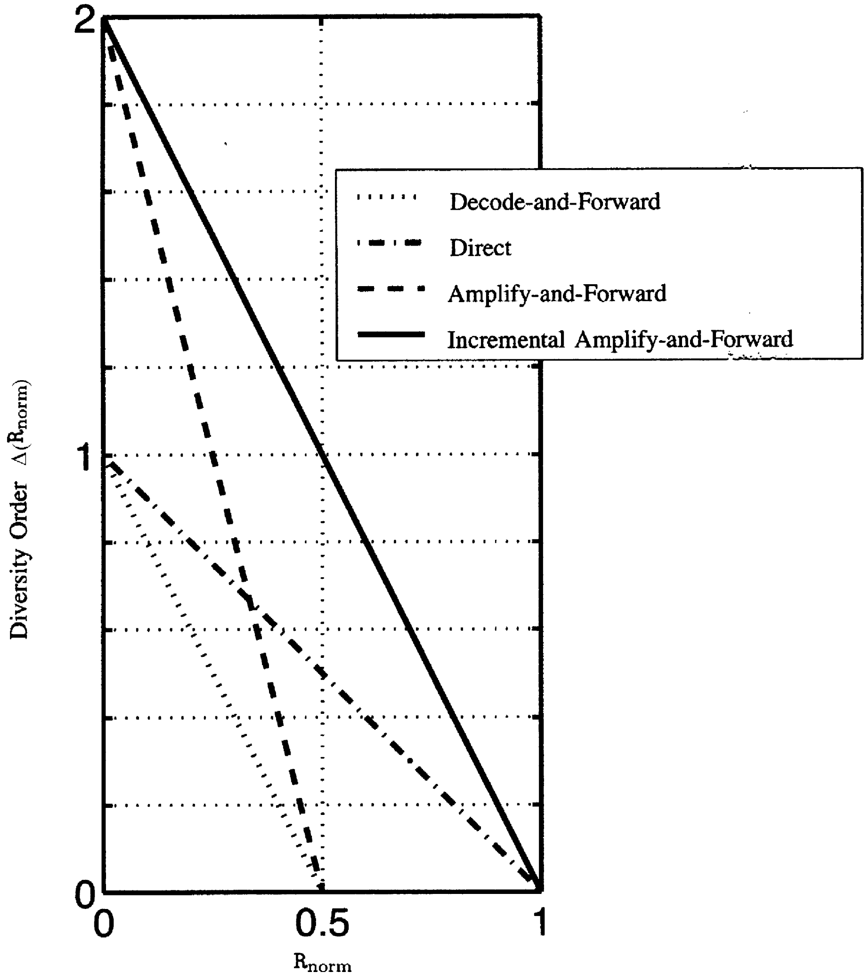

Fig. 6. Diversity order $\Delta(\mathtt{R}_{\mathrm{norm}})$ in (37) versus $\mathtt{R}_{\mathrm{norm}}$ for direct transmission and cooperative diversity.

c are constants. Among other things, for example, (38) implies the approximation $P_{\mathsf{u}(t)}(g_1(t)) \sim cg_2(t)$ is accurate for t close to t_0.

Fact 1: Let u be an exponential random variable with parameter λ_{u}. Then, for a function $g(t)$ continuous about $t = t_0$ and satisfying $g(t) \to 0$ as $t \to t_0$

$$\lim_{t\to t_0} \frac{1}{g(t)} P_{\mathsf{u}}(g(t)) = \lambda_{\mathsf{u}}\,. \tag{39}$$

Fact 2: Let $\mathsf{w} = \mathsf{u} + \mathsf{v}$, where u and v are independent exponential random variables with parameters λ_{u} and λ_{v}, respectively. Then the CDF

$$P_{\mathsf{w}}(w) = \begin{cases} 1 - \left[\left(\frac{\lambda_{\mathsf{v}}}{\lambda_{\mathsf{v}}-\lambda_{\mathsf{u}}}\right) e^{-\lambda_{\mathsf{u}} w} + \left(\frac{\lambda_{\mathsf{u}}}{\lambda_{\mathsf{u}}-\lambda_{\mathsf{v}}}\right) e^{-\lambda_{\mathsf{v}} w}\right], & \lambda_{\mathsf{u}} \neq \lambda_{\mathsf{v}} \\ 1 - (1+\lambda w)e^{-\lambda w}, & \lambda_{\mathsf{u}} = \lambda_{\mathsf{v}} = \lambda \end{cases} \tag{40}$$

satisfies

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2} P_{\mathsf{w}}(\epsilon) = \frac{\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}}{2}\,. \tag{41}$$

Moreover, if a function $g(t)$ is continuous about $t = t_0$ and satisfies $g(t) \to 0$ as $t \to t_0$, then

$$\lim_{t\to t_0} \frac{1}{g^2(t)} P_{\mathsf{w}}(g(t)) = \frac{\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}}{2}\,. \tag{42}$$

Claim 1: Let u, v, and w be independent exponential random variables with parameters λ_{u}, λ_{v}, and λ_{w}, respectively. Let $f(x,y) = (xy)/(x+y+1)$ as in (13). Let ϵ be positive, and let $g(\epsilon) > 0$ be continuous with $g(\epsilon) \to 0$ and $\epsilon/g(\epsilon) \to c < \infty$ as $\epsilon \to 0$. Then

$$\lim_{\epsilon\to 0} \frac{1}{g^2(\epsilon)} \Pr[\mathsf{u} + \epsilon\, f(\mathsf{v}/\epsilon, \mathsf{w}/\epsilon) < g(\epsilon)] = \frac{\lambda_{\mathsf{u}}(\lambda_{\mathsf{v}} + \lambda_{\mathsf{w}})}{2}\,. \tag{43}$$

Moreover, if a function $h(t)$ is continuous about $t = t_0$ and satisfies $h(t) \to 0$ as $t \to t_0$, then

$$\lim_{t\to t_0} \frac{1}{g^2(h(t))} \Pr[\mathsf{u} + h(t)\, f(\mathsf{v}/h(t), \mathsf{w}/h(t)) < g(h(t))] = \frac{\lambda_{\mathsf{u}}(\lambda_{\mathsf{v}} + \lambda_{\mathsf{w}})}{2}\,. \tag{44}$$

The following lemma will be useful in the proof of Claim 1.

Lemma 1: Let δ be positive, and let $\mathsf{r}_\delta := \delta\, f(\mathsf{v}/\delta, \mathsf{w}/\delta)$, where v and w are independent exponential random variables with parameters λ_{v} and λ_{w}, respectively. Let $h(\delta) > 0$ be continuous with $h(\delta) \to 0$ and $\delta/h(\delta) \to d < \infty$ as $\delta \to 0$. Then the probability $\Pr[\mathsf{r}_\delta < \delta]$ satisfies

$$\lim_{\delta\to 0} \frac{1}{h(\delta)} \Pr[\mathsf{r}_\delta < h(\delta)] = \lambda_{\mathsf{v}} + \lambda_{\mathsf{w}}. \tag{45}$$

Proof of Lemma 1: We begin with a lower bound

$$\begin{aligned}\Pr[\mathsf{r}_\delta < h(\delta)] &= \Pr[1/\mathsf{v} + 1/\mathsf{w} + \delta/(\mathsf{vw}) > 1/h(\delta)] \\ &> \Pr[1/\mathsf{v} + 1/\mathsf{w} > 1/h(\delta)] \\ &\geq \Pr[\max(1/\mathsf{v}, 1/\mathsf{w}) > 1/h(\delta)] \\ &= 1 - \Pr[\mathsf{v} \geq h(\delta)] \Pr[\mathsf{w} \geq h(\delta)] \\ &= 1 - \exp[-(\lambda_\mathsf{v} + \lambda_\mathsf{w}) h(\delta)] \end{aligned} \tag{46}$$

so, utilizing Fact 1

$$\liminf_{\delta \to 0} \frac{1}{h(\delta)} \Pr[\mathsf{r}_\delta < h(\delta)] \geq \lambda_\mathsf{v} + \lambda_\mathsf{w}. \tag{47}$$

To prove the other direction, let $l > 1$ be a fixed constant

$$\begin{aligned}\Pr[\mathsf{r}_\delta < h(\delta)] &= \Pr[1/\mathsf{v} + 1/\mathsf{w} + \delta/(\mathsf{vw}) > 1/h(\delta)] \\ &= \int_0^\infty \Pr\left[1/\mathsf{v} > \frac{1/h(\delta) - 1/w}{1 + \delta/w}\right] p_\mathsf{w}(w) dw \\ &\leq \Pr[\mathsf{w} < l h(\delta)] \\ &\quad + \int_{lh(\delta)}^\infty \Pr\left[1/\mathsf{v} > \frac{1/h(\delta) - 1/w}{1 + \delta/w}\right] p_\mathsf{w}(w) dw. \end{aligned} \tag{48}$$

But

$$\Pr[\mathsf{w} < l h(\delta)] / h(\delta) \leq \lambda_\mathsf{w} l \tag{49}$$

which takes care of the first term of (48). To bound the second term of (48), let $k > l$ be another fixed constant, and note that

$$\begin{aligned}&\int_{lh(\delta)}^\infty \Pr[1/\mathsf{v} > (1/h(\delta) - 1/w)/(1 + \delta/w)] p_\mathsf{w}(w) dw \\ &= \int_{kh(\delta)}^\infty \Pr\left[1/\mathsf{v} > \frac{1/h(\delta) - 1/w}{1 + \delta/w}\right] p_\mathsf{w}(w) dw \\ &\quad + \int_{lh(\delta)}^{kh(\delta)} \Pr\left[1/\mathsf{v} > \frac{1/h(\delta) - 1/w}{1 + \delta/w}\right] p_\mathsf{w}(w) dw \\ &\leq \Pr\left[1/\mathsf{v} > \frac{1 - 1/k}{h(\delta) + \delta/k}\right] \\ &\quad + \lambda_\mathsf{w} \int_{lh(\delta)}^{kh(\delta)} \Pr\left[1/\mathsf{v} > \frac{1/h(\delta) - 1/w}{1 + \delta/w}\right] dw \end{aligned} \tag{50}$$

where the first term in the bound of (50) follows from the fact that

$$\Pr[1/\mathsf{v} > (1/h(\delta) - 1/w)/(1 + \delta/w)]$$

is nonincreasing in w, and the second term in the bound of (50) follows from the fact that $p_\mathsf{w}(w) = \lambda_\mathsf{w} \exp(-\lambda_\mathsf{w} w) \leq \lambda_\mathsf{w}$.

Now, the first term of (50) satisfies

$$\Pr\left[1/\mathsf{v} > \frac{1 - 1/k}{h(\delta) + \delta/k}\right] / h(\delta) \leq \lambda_\mathsf{v} \frac{1 + \delta/(kh(\delta))}{1 - 1/k} \tag{51}$$

and, by a change of variable $w' = w/h(\delta)$, the second term of (50) satisfies

$$\begin{aligned}&\frac{1}{h(\delta)} \int_{lh(\delta)}^{kh(\delta)} \Pr\left[1/\mathsf{v} > \frac{1/h(\delta) - 1/w}{1 + \delta/w}\right] dw \\ &= h(\delta) \int_l^k \frac{1}{h(\delta)} \left(1 - \exp\left[-\frac{\lambda_\mathsf{v}(h(\delta) + \delta/w')}{(1 - 1/w')}\right]\right) dw' \\ &\leq h(\delta) \underbrace{\int_l^k \lambda_\mathsf{v} \left(\frac{1 + \delta/(w' h(\delta))}{1 - 1/w'}\right) dw'}_{B(\delta, h(\delta), k, l)} \end{aligned} \tag{52}$$

where $B(\delta, h(\delta), k, l)$ remains finite for any $k > l > 1$ as $\delta \to 0$.

Combining (49), (51), and (52), we have

$$\begin{aligned}&\frac{1}{h(\delta)} \Pr[\mathsf{r}_\delta < h(\delta)] \\ &\leq \lambda_\mathsf{w} l + \lambda_\mathsf{v} \left(\frac{1 + \delta/(kh(\delta))}{1 - 1/k}\right) + h(\delta) B(\delta, h(\delta), k, l) \end{aligned} \tag{53}$$

and furthermore

$$\limsup_{\delta \to 0} \frac{1}{h(\delta)} \Pr[\mathsf{r}_\delta < h(\delta)] \leq \lambda_\mathsf{w} l + \lambda_\mathsf{v} \left(\frac{1 + d/k}{1 - 1/k}\right)$$

since $\lim_{\delta \to 0} B(\delta, h(\delta), k, l) < \infty$ and, by assumption, $h(\delta) \to 0$ and $\delta/h(\delta) \to d$ as $\delta \to 0$.

The constants $k > l > 1$ are arbitrary. In particular, k can be chosen arbitrarily large, and l arbitrarily close to 1. Hence,

$$\limsup_{\delta \to 0} \frac{1}{h(\delta)} \Pr[\mathsf{r}_\delta < h(\delta)] \leq \lambda_\mathsf{w} + \lambda_\mathsf{v}. \tag{54}$$

Combining (47) with (54), the lemma is proved. □

Proof of Claim 1:

$$\begin{aligned}&\Pr[\mathsf{u} + \epsilon f(\mathsf{v}/\epsilon, \mathsf{w}/\epsilon) < g(\epsilon)] \\ &= \Pr[\mathsf{u} + \mathsf{r}_\epsilon < g(\epsilon)] \\ &= \int_0^{g(\epsilon)} \Pr[\mathsf{r}_\epsilon < g(\epsilon) - u] p_\mathsf{u}(u) du \\ &= g(\epsilon) \int_0^{g(\epsilon)} \Pr[\mathsf{r}_\epsilon < g(\epsilon)(1 - u')] \lambda_\mathsf{u} e^{-\lambda_\mathsf{u} g(\epsilon) u'} du' \\ &= g^2(\epsilon) \int_0^1 (1 - u') \left[\frac{\Pr[\mathsf{r}_\epsilon < g(\epsilon)(1 - u')]}{g(\epsilon)(1 - u')}\right] \\ &\quad \times \lambda_\mathsf{u} e^{-\lambda_\mathsf{u} g(\epsilon) u'} du' \end{aligned} \tag{55}$$

where in the second equality we have used the change of variables $u' = u/g(\epsilon)$. But by Lemma 1 with $\delta = \epsilon$ and $h(\delta) = g(\delta)(1 - u')$, the quantity in brackets approaches $\lambda_\mathsf{v} + \lambda_\mathsf{w}$ as $\epsilon \to 0$, so we expect

$$\begin{aligned}&\lim_{\epsilon \to 0} \frac{1}{g^2(\epsilon)} \Pr[\mathsf{u} + \mathsf{r}_\epsilon < g(\epsilon)] \\ &= \lambda_\mathsf{u}(\lambda_\mathsf{v} + \lambda_\mathsf{w}) \int_0^1 (1 - u) du \\ &= \frac{\lambda_\mathsf{u}(\lambda_\mathsf{v} + \lambda_\mathsf{w})}{2}. \end{aligned} \tag{56}$$

To fully verify (56), we must utilize the lower and upper bounds developed in Lemma 1.

Using the lower bound (47), (55) satisfies

$$\begin{aligned}&\liminf_{\epsilon \to 0} \frac{1}{g^2(\epsilon)} \Pr[\mathsf{u} + \mathsf{r}_\epsilon < g(\epsilon)] \\ &\geq \lim_{\epsilon \to 0} \int_0^1 \left(\frac{1 - \exp[-(\lambda_\mathsf{v} + \lambda_\mathsf{w}) g(\epsilon)(1 - u')]}{g(\epsilon)}\right) \\ &\quad \times \lambda_\mathsf{u} e^{-\lambda_\mathsf{u} g(\epsilon) u'} du' \\ &= \lambda_\mathsf{u}(\lambda_\mathsf{v} + \lambda_\mathsf{w}) \int_0^1 (1 - u') du' = \frac{\lambda_\mathsf{u}(\lambda_\mathsf{v} + \lambda_\mathsf{w})}{2} \end{aligned} \tag{57}$$

where the first equality results from the Dominated Convergence Theorem [41] after noting that the integrand is both bounded by and converges to the function $\lambda_\mathsf{u}(\lambda_\mathsf{v} + \lambda_\mathsf{w})(1 - u')$.

Using the upper bound (54), (55) satisfies

$$\begin{aligned}\limsup_{\epsilon\to 0}\frac{1}{g^2(\epsilon)}&\Pr[\mathsf{u}+\mathsf{r}_\epsilon<g(\epsilon)]\\ &\le\limsup_{\epsilon\to 0}(\lambda_{\mathsf{v}}/(1-1/k)+\lambda_{\mathsf{w}}l)\int_0^1(1-u')\lambda_{\mathsf{u}}e^{-\lambda_{\mathsf{u}}g(\epsilon)u'}du'\\ &\quad+\limsup_{\epsilon\to 0}\,\epsilon/g(\epsilon)\int_0^1\lambda_{\mathsf{v}}\lambda_{\mathsf{u}}e^{-\lambda_{\mathsf{u}}g(\epsilon)u'}/(k-1)du'\\ &\quad+\limsup_{\epsilon\to 0}g(\epsilon)D(\epsilon,g(\epsilon),k,l)\end{aligned}\tag{58}$$

where the last equality results from the fact $\epsilon/g(\epsilon)\to c$ and the fact that

$$D(\epsilon,g(\epsilon),k,l):=\int_0^1(1-u')^2B(\epsilon,g(\epsilon)(1-u'),k,l)\lambda_{\mathsf{u}}e^{-\lambda_{\mathsf{u}}g(\epsilon)u'}du'$$

remains finite for all $k>l>1$ even as $\epsilon\to 0$.

Again, the constants $k>l>1$ are arbitrary. In particular, k can be chosen arbitrarily large, and l arbitrarily close to 1. Hence,

$$\limsup_{\epsilon\to 0}\frac{1}{g^2(\epsilon)}\Pr[\mathsf{u}+\mathsf{r}_\epsilon<g(\epsilon)]\le\frac{\lambda_{\mathsf{u}}(\lambda_{\mathsf{v}}+\lambda_{\mathsf{w}})}{2}.\tag{59}$$

Combining (57) and (59) completes the proof. □

Claim 2: Let u and v be independent exponential random variables with parameters λ_{u} and λ_{v}, respectively. Let ϵ be positive and let $g(\epsilon)>0$ be continuous with $g(\epsilon)\to 0$ as $\epsilon\to 0$. Define

$$h(\epsilon):=\epsilon^2\left[(g(\epsilon)/\epsilon+1)\ln(g(\epsilon)/\epsilon+1)-g(\epsilon)/\epsilon\right].\tag{60}$$

Then

$$\lim_{\epsilon\to 0}\frac{1}{h(\epsilon)}\Pr[\mathsf{u}+\mathsf{v}+\mathsf{uv}/\epsilon<g(\epsilon)]=\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}.\tag{61}$$

Moreover, if $\epsilon(t)$ is continuous about $t=t_0$ with $\epsilon(t)\to 0$ as $t\to t_0$, then

$$\lim_{t\to t_0}\frac{1}{h(\epsilon(t))}\Pr[\mathsf{u}+\mathsf{v}+\mathsf{uv}/\epsilon(t)<g(\epsilon(t))]=\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}.\tag{62}$$

Proof: First, we write CDF in the form

$$\begin{aligned}\Pr[\mathsf{u}+\mathsf{v}+&\mathsf{uv}/\epsilon<g(\epsilon)]\\ &=\int_0^\infty\Pr[\mathsf{u}+\mathsf{v}+\mathsf{uv}/\epsilon<g(\epsilon)|\,\mathsf{v}=v]\,p_{\mathsf{v}}(v)\,dv\\ &=\int_0^{g(\epsilon)}\Pr\left[\mathsf{u}<\frac{g(\epsilon)-v}{1+v/\epsilon}\middle|\,\mathsf{v}=v\right]\lambda_{\mathsf{v}}e^{-\lambda_{\mathsf{v}}v}\,dv\\ &=\int_0^{g(\epsilon)}\left[1-\exp\left(-\lambda_{\mathsf{u}}\left[\frac{g(\epsilon)-v}{1+v/\epsilon}\right]\right)\right]\lambda_{\mathsf{v}}e^{-\lambda_{\mathsf{v}}v}\,dv\\ &=g(\epsilon)\int_0^1\left[1-\exp\left(-\lambda_{\mathsf{u}}\left[\frac{g(\epsilon)(1-w)}{1+g(\epsilon)w/\epsilon}\right]\right)\right]\\ &\quad\times\lambda_{\mathsf{v}}e^{-\lambda_{\mathsf{v}}g(\epsilon)w}\,dw\end{aligned}\tag{63}$$

where the last equality follows from the change of variables $w=v/g(\epsilon)$.

To upper-bound (63), we use the identities $1-e^{-x}\le x$ for all $x\ge 0$ and $e^{-y}\le 1$ for all $y\ge 0$, so that (63) becomes

$$\begin{aligned}\Pr[\mathsf{u}+\mathsf{v}+&\mathsf{uv}/\epsilon<g(\epsilon)]\\ &\le g^2(\epsilon)\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}\int_0^1\frac{1-w}{1+g(\epsilon)w/\epsilon}\,dw\\ &=\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}\,g^2(\epsilon)\,\frac{(g(\epsilon)/\epsilon+1)\ln(g(\epsilon)/\epsilon+1)-g(\epsilon)/\epsilon}{(g(\epsilon)/\epsilon)^2}\\ &=\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}h(\epsilon)\end{aligned}$$

whence

$$\limsup_{\epsilon\to 0}\frac{1}{h(\epsilon)}\Pr[\mathsf{u}+\mathsf{v}+\mathsf{uv}/\epsilon<g(\epsilon)]\le\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}.\tag{64}$$

To lower-bound (63), we use the concavity of $1-e^{-x}$, i.e., for any $t>0$,

$$1-e^{-x}\ge\frac{1-e^{-t}}{t}x,\qquad\text{for all }x\le t$$

and the identity $e^{-y}\ge 1-y$ for all $y\ge 0$, so that (63) becomes

$$\begin{aligned}\Pr[\mathsf{u}+\mathsf{v}+&\mathsf{uv}/\epsilon<g(\epsilon)]\\ &\ge g(\epsilon)\int_0^1\left[\left(\frac{1-e^{-\lambda_{\mathsf{u}}g(\epsilon)}}{\lambda_{\mathsf{u}}g(\epsilon)}\right)\frac{\lambda_{\mathsf{u}}g(\epsilon)(1-w)}{1+wg(\epsilon)/\epsilon}\right]\\ &\quad\times\lambda_{\mathsf{v}}(1-\lambda_{\mathsf{v}}g(\epsilon)w)\,dw\\ &=\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}\,g^2(\epsilon)\left(\frac{1-e^{-\lambda_{\mathsf{u}}g(\epsilon)}}{\lambda_{\mathsf{u}}g(\epsilon)}\right)\\ &\quad\times\int_0^1\left[\frac{1-w}{1+wg(\epsilon)/\epsilon}\right](1-\lambda_{\mathsf{v}}g(\epsilon)w)\,dw\\ &\ge\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}g^2(\epsilon)\left(\frac{1-e^{-\lambda_{\mathsf{u}}g(\epsilon)}}{\lambda_{\mathsf{u}}g(\epsilon)}\right)(1-\lambda_{\mathsf{v}}g(\epsilon))\\ &\quad\times\int_0^1\frac{1-w}{1+wg(\epsilon)/\epsilon}\,dw\\ &=\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}\left(\frac{1-e^{-\lambda_{\mathsf{u}}g(\epsilon)}}{\lambda_{\mathsf{u}}g(\epsilon)}\right)(1-\lambda_{\mathsf{v}}g(\epsilon))\,g^2(\epsilon)\\ &\quad\times\frac{(g(\epsilon)/\epsilon+1)\ln(g(\epsilon)/\epsilon+1)-g(\epsilon)/\epsilon}{(g(\epsilon)/\epsilon)^2}\\ &=\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}\left(\frac{1-e^{-\lambda_{\mathsf{u}}g(\epsilon)}}{\lambda_{\mathsf{u}}g(\epsilon)}\right)(1-\lambda_{\mathsf{v}}g(\epsilon))\,h(\epsilon).\end{aligned}$$

Thus,

$$\begin{aligned}\liminf_{\epsilon\to 0}\frac{1}{h(\epsilon)}&\Pr[\mathsf{u}+\mathsf{v}+\mathsf{uv}/\epsilon<g(\epsilon)]\\ &\ge\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}\lim_{\epsilon\to 0}\left(\frac{1-e^{-\lambda_{\mathsf{u}}g(\epsilon)}}{\lambda_{\mathsf{u}}g(\epsilon)}\right)(1-\lambda_{\mathsf{v}}g(\epsilon))\\ &=\lambda_{\mathsf{u}}\lambda_{\mathsf{v}}.\end{aligned}\tag{65}$$

Since the bounds in (64) and (65) are equal, the claim is proved. □

Claim 3: Suppose $f_t(s)\to g(s)$ pointwise as $t\to t_0$, and that $f_t(s)$ is monotone increasing in s for each t. Let $h_t(s)$

be such that $h_t(s) \le s$, $h_t(s) \to s$ pointwise as $t \to t_0$, and $h_t(s)/s$ is monotone decreasing in s for each t. Define $\tilde{h}_t^{-1}(r) := \min h_t^{-1}(r)$. Then

$$\lim_{t\to t_0} f_t(\tilde{h}_t^{-1}(r)) = g(r)\,. \tag{66}$$

Proof: Since $h_t(s) \le s$ for all t, we have $r \le \tilde{h}_t^{-1}(r)$, and consequently $f_t(r) \le f_t(\tilde{h}_t^{-1}(r))$ because $f_t(\cdot)$ is monotone increasing. Thus,

$$\liminf_{t\to t_0} f_t(\tilde{h}_t^{-1}(r)) \ge g(r). \tag{67}$$

The upper bound is a bit more involved. Fix $\delta > 0$. Lemma 2 shows that for each r there exists t^* such that

$$\tilde{h}_t^{-1}(r) \le r/(1-\delta)$$

for all t such that $|t - t_0| < |t^* - t_0|$. Then we have

$$f_t(\tilde{h}_t^{-1}(r)) \le f_t(r/(1-\delta))\,.$$

Thus,

$$\limsup_{t\to t_0} f_t(\tilde{h}_t^{-1}(r)) \le g(r/(1-\delta))$$

and since δ can be made arbitrarily small

$$\limsup_{t\to t_0} f_t(\tilde{h}_t^{-1}(r)) \le g(r). \tag{68}$$

Combining (67) with (68), we obtain the desired result. □

The following Lemma is used in the proof of the upper bound of Claim 3.

Lemma 2: Let $h_t(s)$ be such that $h_t(s) \le s$, $h_t(s) \to s$ pointwise as $t \to t_0$, and $h_t(s)/s$ is monotone decreasing in s for each t. Define $\tilde{h}_t^{-1}(r) := \min h_t^{-1}(r)$. For each $r_0 > 0$ and any $\delta > 0$, there exists t^* such that

$$\tilde{h}_t^{-1}(r_0) \le r_0/(1-\delta)$$

for all t such that $|t - t_0| < |t^* - t_0|$.

Proof: Fix $r_0 > 0$ and $\delta > 0$, and select s_0 such that $s_0 > r_0/(1-\delta)$.

Because $h_t(s)/s \to 1$ point-wise as $t \to t_0$, for each $s > 0$ and any $\delta > 0$, there exists a t^* such that

$$h_t(s) \ge s(1-\delta), \qquad \text{all } t : |t - t_0| < |t^* - t_0|.$$

Moreover, since $h_t(s)/s$ is monotone decreasing in s, if t^* is sufficient for convergence at s_0, then it is sufficient for convergence at all $s \le s_0$. Thus, for any $s_0 > 0$ and $\delta > 0$ there exists a t^* such that

$$h_t(s) \ge s(1-\delta), \qquad \text{all } s \le s_0, t : |t - t_0| < |t^* - t_0|.$$

Throughout the rest of the proof, we only consider $s \le s_0$ and t such that $|t - t_0| < |t^* - t_0|$.

Consider the interval $I = [r_0, r_0/(1-\delta)]$, and note that $s \in I$ implies $s < s_0$. Since $h_t(s) < s$, we have $h_t(r_0) < r_0$. Also, since $h_t(s) > s(1-\delta)$ by the above construction, we have $h_t(r_0/(1-\delta)) > r_0$. By continuity, $h_t(s)$ assumes all intermediate values between $h_t(r_0)$ and $h_t(r_0/(1-\delta))$ on the interval $(r_0, r_0/(1-\delta))$ [42, Theorem 4.23]; in particular, there exists an $s_1 \in (r_0, r_0/(1-\delta))$ such that $h_t(x_1) = r_0$. The result follows from $\tilde{h}_t^{-1}(r_0) \le x_1 \le r_0/(1-\delta)$, where the first inequality follows from the definition of $\tilde{h}_t^{-1}(\cdot)$ and the second inequality follows from the fact that $x_1 \in I$. □

Appendix II
Amplify-and-Forward Mutual Information

For completeness, in this appendix we compute the maximum average mutual information for amplify-and-forward transmission (12). The result borrows substantially from the vector results in [2], [3], aside from taking into account the amplifier power constraint in the relay as well as simplifying manipulations.

We write the equivalent channel (2)–(4), with relay processing (8), in vector form as

$$\underbrace{\begin{bmatrix} \mathsf{y}_d[n] \\ \mathsf{y}_d[n+N/4] \end{bmatrix}}_{\mathbf{y}_d[n]} = \underbrace{\begin{bmatrix} \mathsf{a}_{s,d} \\ \mathsf{a}_{r,d}\beta\mathsf{a}_{s,r} \end{bmatrix}}_{A} \mathsf{x}_s[n] + \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ \mathsf{a}_{r,d}\beta & 0 & 1 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} \mathsf{z}_r[n] \\ \mathsf{z}_d[n] \\ \mathsf{z}_d[n+N/4] \end{bmatrix}}_{\mathbf{z}[n]}$$

where the source signal has power constraint $\mathrm{E}[\mathsf{x}_s] \le P_s$, and relay amplifier has constraint

$$\beta \le \sqrt{\frac{P_r}{|\mathsf{a}_{s,r}|^2 P_s + N_r}} \tag{69}$$

and the noise has covariance $\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right] = \mathrm{diag}(N_r, N_d, N_d)$. Note that we determine the mutual information for arbitrary transmit powers, relay amplification, and noise levels, even though we utilize the result only for the symmetric case. Since the channel is memoryless, the average mutual information satisfies

$$\mathsf{I}_{AF} \le \mathsf{I}(\mathsf{x}_s; \mathbf{y}_d) \le \log\det\left(I + (P_s A A^\dagger)(B\,\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]B^\dagger)^{-1}\right)$$

with equality for x_s zero-mean, circularly symmetric complex Gaussian [2], [3]. Noting that

$$AA^\dagger = \begin{bmatrix} |\mathsf{a}_{s,d}|^2 & \mathsf{a}_{s,d}(\mathsf{a}_{r,d}\beta\mathsf{a}_{s,r})^* \\ \mathsf{a}_{s,d}{}^*\mathsf{a}_{r,d}\beta\mathsf{a}_{s,r} & |\mathsf{a}_{r,d}\beta\mathsf{a}_{s,r}|^2 \end{bmatrix}$$

$$B\,\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]B^\dagger = \begin{bmatrix} N_d & 0 \\ 0 & |\mathsf{a}_{r,d}\beta|^2 N_r + N_d \end{bmatrix}$$

we have

$$\det\left(I_2 + (P_s AA^\dagger)(B\,\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]B^\dagger)^{-1}\right) = 1 + \frac{P_s|\mathsf{a}_{s,d}|^2}{N_d} + \frac{P_s|\mathsf{a}_{r,d}\beta\mathsf{a}_{s,r}|^2}{(|\mathsf{a}_{r,d}\beta|^2 N_r + N_d)}. \tag{70}$$

It is apparent that (70) is increasing in β, so the amplifier power constraint (69) should be active, yielding, after substitutions and algebraic manipulations

$$\mathsf{I}_{AF} = \log(1 + |\mathsf{a}_{s,d}|^2 \mathtt{SNR}_{s,d} + f(|\mathsf{a}_{s,r}|^2\mathtt{SNR}_{s,r}, |\mathsf{a}_{r,d}|^2\mathtt{SNR}_{r,d}))$$

with $f(\cdot,\cdot)$ given by (13).

APPENDIX III
INPUT DISTRIBUTIONS FOR TRANSMIT DIVERSITY BOUND

In this appendix, we derive the input distributions that minimize outage probability for transmit diversity schemes in the high-SNR regime. Our derivation is a slight extension of the results in [2], [3] dealing with asymmetric fading variances.

An equivalent channel model for the two-antenna case can be summarized as

$$\mathsf{y}[n] = \underbrace{\begin{bmatrix} \mathsf{a}_1 & \mathsf{a}_2 \end{bmatrix}}_{\mathbf{a}} \underbrace{\begin{bmatrix} \mathsf{x}_1[n] \\ \mathsf{x}_2[n] \end{bmatrix}}_{\mathbf{x}[n]} + \mathsf{z}[n] \tag{71}$$

where $\mathbf{a}$ represents the fading coefficients and $\mathsf{x}[n]$ the transmit signals from the two transmit antennas, and $\mathsf{z}[n]$ is a zero-mean, white complex Gaussian process with variance N_0 that captures the effects of noise and interference. Let $Q = \mathrm{E}\left[\mathsf{x}\mathsf{x}^\dagger\right]$ be the covariance matrix for the transmit signals. Then the power constraint on the inputs may be written in the form $\mathrm{tr}(Q) \leq P$.

We are interested in determining a distribution on the input vector x, subject to the power constraint, that minimizes outage probability, i.e.,

$$\min_{p_{\mathbf{x}}:\mathrm{tr}(Q)\leq P} \Pr[I(\mathsf{x};\mathsf{y}|\mathbf{a}=\boldsymbol{a}) < \mathtt{R}]. \tag{72}$$

As [2], [3] develops, the optimization (72) can be restricted to optimization over zero-mean, circularly symmetric complex Gaussian inputs, because Gaussian codebooks maximize the mutual information for each value of the fading coefficients $\mathbf{a}$. Thus, (72) is equivalent to maximizing over the covariance matrix of the complex Gaussian inputs subject to the power constraint, i.e.,

$$\min_{Q:\,\mathrm{tr}(Q)\leq P} \Pr\left[\log\left(1 + \frac{\mathbf{a}Q\mathbf{a}^\dagger}{N_0}\right) < \mathtt{R}\right]. \tag{73}$$

We now argue that Q diagonal is sufficient, even if the components of $\mathbf{a}$ are independent but not identically distributed. We note that this argument is a slight extension of [2], [3], in which i.i.d. fading coefficients are treated. Although we treat the case of two transmit antennas, the argument extends naturally to more than two antennas.

We write $\mathbf{a} = \tilde{\mathbf{a}}\Sigma$, where $\tilde{\mathbf{a}}$ is a zero-mean, i.i.d. complex Gaussian vector with unit variances and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2)$. Thus, the outage probability in (73) may be written as

$$\Pr\left[\log\left(1 + \frac{\tilde{\mathbf{a}}\Sigma Q\Sigma^\dagger\tilde{\mathbf{a}}^\dagger}{N_0}\right) < \mathtt{R}\right].$$

Now consider an eigendecomposition of the matrix $\Sigma Q\Sigma^\dagger = UDU^\dagger$, where U is unitary and D is diagonal. Using the fact that the distribution of $\tilde{\mathbf{a}}$ is rotationally invariant, i.e., $\tilde{\mathbf{a}}U$ has the same distribution as $\tilde{\mathbf{a}}$ for any unitary U [2], [3], we observe that the outage probability for covariance matrix $\Sigma Q\Sigma^\dagger$ is the same as the outage probability for the diagonal matrix D.

For $D = \mathrm{diag}(d_1, d_2)$, the outage probability can be written in the form

$$\Pr\left[d_1|\mathsf{a}_1|^2 + d_2|\mathsf{a}_2|^2 < \frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right],$$

which, using Fact 2, decays proportional to $1/(\mathtt{SNR}^2 \det D)$ for large $\mathtt{SNR}$ if $d_1, d_2 \neq 0$. Thus, minimizing the outage probability for large $\mathtt{SNR}$ is equivalent to maximizing

$$\det D = \det \Sigma Q\Sigma^\dagger = {\sigma_1}^2{\sigma_2}^2(Q_{1,1}Q_{2,2} - |Q_{1,2}|^2) \tag{74}$$

such that $Q_{1,1} + Q_{2,2} \leq P$. Clearly, (74) is maximized for $Q_{1,1} = Q_{2,2} = P/2$ and $Q_{1,2} = Q_{2,1} = 0$. Thus, zero-mean, i.i.d. complex Gaussian inputs minimize the outage probability in the high-SNR regime.

REFERENCES

[1] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, Inc., 2001.

[2] İ. E. Telatar. (1995) Capacity of Multi-Antenna Gaussian Channels, Tech. Rep. Bell Labs, Lucent Technologies. [Online]. Available: http://mars.bell-labs.com/papers/proof/proof.pdf

[3] ——, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–596, Nov.-Dec. 1999.

[4] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2418–2433, Nov. 1999.

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

[6] A. Sendonaris, E. Erkip, and B. Aazhang, "Increasing uplink capacity via user cooperation diversity," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Cambridge, MA, Aug. 1998, p. 156.

[7] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[8] V. Hayes, *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 1997.

[9] L. Zheng and D. N. C. Tse, "Diversity and freedom: A fundamental tradeoff in multiple antenna channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Lausanne, Switzerland, June/July 2002, p. 476.

[10] ——, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[11] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[12] E. C. van der Meulen, *Transmission of Information in a T-Terminal Discrete Memoryless Channel*. Berkeley, CA: Dept. Statistics, Univ. California, , 1968.

[13] ——, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, pp. 120–154, 1971.

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

[15] D. S. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[16] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 294–300, May 1975.

[17] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, Jan. 1976.

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

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

[20] P. Gupta and P. R. Kumar, "Toward and information theory of large networks: An achievable rate region," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Washington, DC, June 2001, p. 150.

[21] ——, "Toward an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1877–1894, Aug. 2003.

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

[23] ——, "On the asymptotic capacity of Gaussian relay networks," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Lausanne, Switzerland, July 2002, p. 195.

[24] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Lausanne, Switzerland, July 2002, p. 136.
[25] A. Reznik, S. Kulkarni, and S. Verdú, "Capacity and optimal resource allocation in the degraded Gaussian relay channel with multiple relays," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2002.
[26] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Sorrento, Italy, June 2000, p. 40.
[27] R. C. King, "Multiple access channels with generalized feedback," Ph.D. dissertation, Stanford Univ., Palo Alto, CA, 1978.
[28] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 841–850, Nov. 1982.
[29] F. M. J. Willems, "Informationtheoretical results for the discrete memoryless multiple access channel," Ph.D. dissertation, Katholieke Univ. Leuven, Leuven, Belgium, 1982.
[30] F. M. J. Willems, E. C. van der Meulen, and J. P. M. Schalkwijk, "An achievable rate region for the multiple access channel with generalized feedback," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 1983, pp. 284–292.
[31] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 441–445, May 1983.
[32] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 313–327, May 1985.
[33] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, Part I: System description," *IEEE Trans. Commun.*, vol. 51, pp. 1927–1938, Nov. 2003.
[34] ——, "User cooperation diversity, Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, pp. 1939–1948, Nov. 2003.
[35] J. N. Laneman and G. W. Wornell, "Exploiting distributed spatial diversity in wireless networks," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2000.
[36] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, "An efficient protocol for realizing cooperative diversity in wireless networks," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Washington, DC, June 2001, p. 294.
[37] J. N. Laneman and G. W. Wornell, "Distributed space–time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2415–2525, Oct. 2003.
[38] J. N. Laneman, "Cooperative diversity in wireless networks: algorithms and architectures," Ph.D. dissertation, MIT, Cambridge, MA, 2002.
[39] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2384–2415, Oct. 1998.
[40] J. N. Laneman and G. W. Wornell, "Energy-efficient antenna sharing and relaying for wireless networks," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Chicago, IL, Sept. 2000.
[41] M. Adams and V. Guillemin, *Measure Theory and Probability*. Boston, MA: Birkhäuser, 1996.
[42] W. Rudin, *Principles of Mathematical Analysis*, 2nd ed. New York: McGraw-Hill, 1964.