Cooperative Diversity in Wireless Relay Networks with Multiple-Antenna Nodes

YINDI JING AND BABAK HASSIBI*

Department of Electrical Engineering

California Institute of Technology

Pasadena, CA 91125

Abstract

In [1], the idea of of space-time coding devised for multiple-antenna systems is applied to the problem of communication over a wireless relay network, a strategy called *distributed space-time coding*, to achieve the cooperative diversity provided by antennas of the relay nodes. In this paper, we extend the idea of distributed space-time coding to wireless relay networks with multiple-antenna nodes and fading channels. We show that for a wireless relay network with M antennas at the transmit node, N antennas at the receive node, and a total of $\mathcal{R}$ antennas at all the relay nodes, provided that the coherence interval is long enough, the high SNR *pairwise error probability (PEP)* behaves as $\left(\frac{1}{P}\right)^{\min\{M,N\}\mathcal{R}}$ if $M \neq N$ and $\left(\frac{\log^{1/M} P}{P}\right)^{\mathcal{M}\mathcal{R}}$ if $M = N$, where P is the total power consumed by the network. Therefore, for the case of $M \neq N$, distributed space-time coding achieves the same diversity as decode-and-forward without any rate constraint on the transmission. For the case of $M = N$, the penalty is a factor of $\log^{1/M} P$ which, compared to P, becomes negligible when P is very high. We also show that for a fixed total transmit power across the entire network, the optimal power allocation is for the transmitter to expend half the power and for the relays to share the other half with the power used at every relay proportional to the number of antennas it has.

1 Introduction

It is known that multiple antennas can greatly increase the capacity and reliability of a wireless communication link in a fading environment using space-time coding [2, 3, 4, 5]. Recently, with the increasing interest in ad hoc

*This work is supported in part by the National Science Foundation under grant nos. CCR-0133818 and CCR-0326554, by the David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking.

networks, researchers have been looking for methods to exploit spatial diversity using the antennas of different users in the network [6, 7, 8, 9, 10, 11, 1]. In [8], the authors exploit spatial diversity using the repetition-based and space-time cooperative algorithms. The mutual information and outage probability of the network are analyzed. However, in their model, the relay nodes need to decode their received signals. In [9], a network with a single relay under different protocols is analyzed and second order spatial diversity is achieved. In [10], the authors use space-time codes based on the Hurwitz-Radon matrices and conjecture a diversity factor around $R/2$ from their simulations. Also, the simulations in [11] show that the use of Khatri-Rao codes lowers the average bit error rate.

This work follows the strategy of [1], where the idea of space-time coding devised for multiple-antenna systems is applied to the problem of communication over a wireless relay network.[1] In [1], the authors consider wireless relay networks in which every node has a single antenna and the channels are fading, and use a cooperative strategy called distributed space-time coding by applying a linear dispersion space-time code [12] among the relays. It is proved that without any channel knowledge at the relays, a diversity of $R\left(1-\frac{\log\log P}{\log P}\right)$ can be achieved, where R is number of relays and P is the total power consumed in the whole network. This result is based on the assumption that the receiver has full knowledge of the fading channels. Therefore, when the total transmit power P is high enough, the wireless relay network achieves the diversity of a multiple-antenna system with R transmit antennas and one receive antenna, asymptotically. That is, antennas of the relays work as antennas of the transmitter although they cannot fully cooperate and do not have full knowledge of the transmit signal. Compared with the other widely used cooperative strategy, decode-and-forward, since no decoding is needed at the relays, distributed space-time coding saves both time and energy, and more importantly, there is no rate constraint on the transmission.

In this paper, we extend the idea of distributed space-time coding to wireless relay networks whose nodes have multiple antennas, and analyze the achievable diversity. As in [1], the focus of this paper is on the pairwise error probability (PEP) analysis. We investigate the achievable diversity gain in a wireless relay network by having the relays cooperate distributively. By diversity gain, or diversity in brief, we mean the negative of the exponent of the SNR or transmit power in the PEP formula at the high SNR regime. This definition is consistent with the diversity definition in multiple-antenna systems [5, 13][2]. It determines how fast the PEP decreases with

[1]Although having the same name, the distributed space-time coding idea in [1] is different from that in [8]. Similar ideas for networks with one and two relays have appeared in [9, 11].

[2]However, this definition is different from the formal diversity definition, $-\lim_{P\to\infty}\frac{\log PEP}{\log P}$, given in [14]. It is more precise in the sense that it can describe the PEP behavior in more detail. For example, if $PEP\sim P^{-[c+f(P)]}$, the same diversity c will be

the SNR or transmit power.

We use the same two-step transmission method in [1, 15, 16], where in one step the transmitter sends signals to the relays and in the other the relays encode their received signals into a linear dispersion space-time code and transmit to the receiver. For a wireless relay network with M antennas at the transmitter, N antennas at the receiver, and a total of $\mathcal{R}$ antennas at all the relay nodes, our work shows that when the coherence interval is long enough, a diversity of $\min\{M,N\}\mathcal{R}$, if $M \neq N$, and $M\mathcal{R}\left(1-\frac{1}{M}\frac{\log\log P}{\log P}\right)$, if $M=N$ can be achieved, where P is the total power used in the network. With this two-step protocol, it is easy to see that the error probability is determined by the worse of the two steps: the transmission from the transmitter to the relays and the transmission from the relays to the receiver. Therefore, when $M \neq N$, distributed space-time coding is optimal since the diversity of the first stage cannot be better than $M\mathcal{R}$, the diversity of a multiple-antenna system with M transmit antennas and $\mathcal{R}$ receive antennas, and the diversity of the second stage cannot be better than $N\mathcal{R}$. When $M = N$, the penalty on the diversity because the relays cannot fully cooperate and do not have full knowledge of the signal is $\mathcal{R}\frac{\log\log P}{\log P}$. When P is very high, it is negligible. Therefore, with distributed space-time coding, wireless relay networks achieve the same diversity of multiple-antenna systems, asymptotically.

The paper is organized as follows. In the following section, the network model and the generalized distributed space-time coding is explained in detail. The PEP is first analyzed in Section 3 based on which an optimum power allocation between the transmitter and the relays with respect to minimizing the PEP are given in Section 4. In Section 5, the diversity for the network with an infinite number of relays is discussed. Then the diversity for the general case is obtained in Section 6. Section 7 is the conclusion and discussion. Proofs of some of the technical theorems are given in the appendices.

2 Wireless Relay Network

2.1 System Model

We first introduce some notation. For a complex matrix A, $\bar{A}$, A^t, and A^* denote the conjugate, the transpose, and the conjugate transpose of A, respectively. $\det A$, $\operatorname{rank} A$, and $\operatorname{tr} A$ indicate the determinant, rank, and trace of A, respectively. I_n denotes the $n \times n$ identity matrix and $0_{m,n}$ is the $m \times n$ matrix with all zero entries.

obtained for any function f satisfying $\lim_{P\to\infty} f(P) = 0$. Also, this more precise definition is used to emphasize that the slope of the logPEP vs. logSNR curve changes with the SNR.

We often omit the subscripts when there is no confusion. $\log$ indicates the natural logarithm. $\|\cdot\|$ indicates the Frobenius norm. P and E indicate the probability and the expected value. $g(x) = O(f(x))$ means that $\lim_{x\to\infty} \frac{g(x)}{f(x)}$ is a constant. $h(x) = o(f(x))$ means that $\lim_{x\to\infty} \frac{g(x)}{f(x)} = 0$.

Consider a wireless network with $R+2$ nodes which are placed randomly and independently according to some distribution. There is one transmit node and one receive node. All the other R nodes work as relays. This is a practical appropriate model for many sensor networks. By using a straight-forward TDMA technique, it can be generalized to ad hoc wireless networks with multiple pairs of transmitters and receivers.[3] The transmitter has M transmit antennas, the receiver has N receive antennas, and the i-th relay has R_i antennas, which can be used for both transmission and reception. Since the transmit and received signals at different antennas of the same relay can be processed and designed independently, the network can be transformed to a network with $\mathcal{R} = \sum_{i=1}^{R} R_i$ single-antenna relays by designing the transmit signal at every antenna of every relay according to the received signal at that antenna only.[4] Therefore, without loss of generality, in the following, we assume that every relay has a single antenna.[5]

Therefore, the network can be depicted by Figure 1. Denote the channels from the M antennas of the transmitter to the i-th relay as $f_{1i}, f_{2i}, \cdots, f_{Mi}$, and the channels from the i-th relay to the N antennas at the receiver as $g_{i1}, g_{i2}, \cdots, g_{iN}$. Here, we only consider the fading effect of the channels by assuming that f_{mi} and g_{in} are independent complex Gaussian with zero-mean and unit-variance. This is a common assumption for networks in urban areas or indoors when there is no line-of-sight, or situations where the distances between the relays and the transmitter/receiver are about the same. We make the practical assumption that the channels f_{mi} and g_{in} are not known at the relays.[6] What every relay knows is only the statistical distribution of its local connections. However, we do assume that the receiver knows all the fading coefficients f_{mi} and g_{in}. Its knowledge of the channels can be obtained by sending training signals from the relays and the transmitter.

We use the block-fading model [3] by assuming a coherence interval T, that is, the time during which f_{mi} and g_{in} keep constant.[7] The information bits are thus encoded into $T \times M$ matrices $\underline{\mathrm{s}} = [\mathrm{s}_1, \cdots, \mathrm{s}_M]$,

[3] However, this straightforward TDMA scheme may not be optimal.

[4] This is one possible scheme. In general, the signal sent by one antenna of a relay can be designed using received signals at all the antennas of the relay. However, as will be seen later, this simpler scheme achieves the optimal diversity asymptotically although a general design may improve the coding gain of the network.

[5] The main reason is to highlight the diversity results by simplifying notation and formulas.

[6] For the i-th relay to know f_{mi}, training from the transmitter and estimation at the i-th relay are needed. It is even less practical for the i-th relay to know g_{in}, which needs feedback from the receiver.

[7] From the two-step protocol that will be discussed in the following, we can see that we only need f_{mi} to keep constant for the

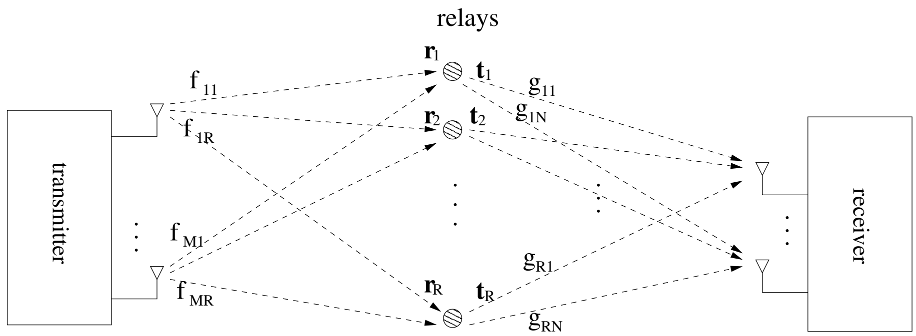

Figure 1: Wireless relay network with M antennas at the transmitter and N antennas at the receiver

where $\mathbf{s}_m$, a T-dimensional vector, is the signal sent by the m-th transmit antenna. For the power analysis, $\underline{\mathbf{s}}$ is normalized as

$$\mathrm{E}\,\mathrm{tr}\,\underline{\mathbf{s}}^*\underline{\mathbf{s}} = M. \tag{1}$$

To send $\underline{\mathbf{s}}$ to the receiver, the same two-step strategy in [1] is used here. As shown in Figure 1, in step one, which is from time 1 to T, the transmitter sends $\sqrt{P_1 T/M}\underline{\mathbf{s}}$ with $\sqrt{P_1 T/M}\mathbf{s}_m$ being the signal sent by the m-th antenna from time 1 to T. Based on (1), the average total power used at the transmitter for the T transmissions is $P_1 T$. The received signal at the i-th relay at time τ is denoted as $r_{i,\tau}$. We denote the additive noise at the i-th relay at time τ as $v_{i,\tau}$. In step two, which is from time $T+1$ to $2T$, the i-th relay sends $t_{i,1}, \cdots, t_{i,T}$. We denote the received signal and noise at the n-th receive antenna of the receiver at time $T+\tau$ by $x_{\tau n}$ and $w_{\tau n}$. Assume that the noises are independent complex Gaussian with zero-mean and unit-variance, that is, the distribution of $v_{i,\tau}, w_{\tau,n}$ is i.i.d. $\mathcal{CN}(0,1)$.

We use the following notation:

$$\mathbf{v}_i = \begin{bmatrix} v_{i,1} \\ v_{i,2} \\ \vdots \\ v_{i,T} \end{bmatrix} \quad \mathbf{r}_i = \begin{bmatrix} r_{i,1} \\ r_{i,2} \\ \vdots \\ r_{i,T} \end{bmatrix} \quad \mathbf{t}_i = \begin{bmatrix} t_{i,1} \\ t_{i,2} \\ \vdots \\ t_{i,T} \end{bmatrix} \quad \mathbf{w}_n = \begin{bmatrix} w_{1n} \\ w_{2n} \\ \vdots \\ w_{Tn} \end{bmatrix} \quad \mathbf{x}_n = \begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{Tn} \end{bmatrix},$$

where $\mathbf{v}_i$, $\mathbf{r}_i$, and $\mathbf{t}_i$ are the noise, the received signal, and the transmit signal at the i-th relay, $\mathbf{w}_n$ is the noise at the n-th antenna of the receiver, and $\mathbf{x}_n$ is the received signal at the n-th antenna of the receiver. $\mathbf{v}_i$, $\mathbf{r}_i$, $\mathbf{t}_i$, $\mathbf{w}_n$, and $\mathbf{x}_n$ are all T-dimensional column vectors.

first step of the transmission and g_{in} to keep constant for the second step. It is thus good enough to choose T as the minimum of the coherence intervals of f_{mi} and g_{in}.

Since f_{mi} and g_{in} keep constant for T transmissions, clearly

$$\mathbf{r}_i = \sqrt{P_1 T/M} \sum_{m=1}^{M} f_{mi} \mathbf{s}_m + \mathbf{v}_i = \sqrt{P_1 T/M} \underline{\mathbf{s}} \mathbf{f}_i + \mathbf{v}_i \tag{2}$$

and

$$\mathbf{x}_n = \sum_{i=1}^{R} g_{in} \mathbf{t}_i + \mathbf{w}_n \tag{3}$$

where we have defined $\mathbf{f}_i = \begin{bmatrix} f_{1i} & f_{2i} & \cdots & f_{Mi} \end{bmatrix}^t$.

2.2 Distributed Space-Time Coding

We want the relays in the network cooperate in a way such that their antennas work as transmit/receive antennas of the same user to obtain diversity. There are two main differences between the wireless network in Figure 1 and a point-to-point multiple-antenna communication system analyzed in [5, 13]. The first is that in a multiple-antenna system, antennas of the transmitter can cooperate fully while in the network, the relays do not communicate with each other and can only cooperate in a distributive fashion. The other difference is that in a wireless network, every relay just has a noisy version of the transmit signal $\underline{\mathbf{s}}$.

Therefore, a crucial issue is how should every relay help the transmission, or more specifically, how should the i-th relay design its transmit signal $\mathbf{t}_i$ based on its received signal $\mathbf{r}_i$. One of the most widely used strategies is called decode-and-forward, (see, e.g., [6, 8]), in which the i-th relay fully decodes its received signal, and then encodes the information again and transmits the newly encoded signal. If the transmission rate is sufficiently low so that all the R relays are able to successfully decode, the system can act as a multiple-antenna system with R transmit antennas and N receive antennas. Therefore the communication from the relays to the receiver can achieve a diversity of NR. However, if some nodes decode incorrectly, they will forward incorrect information to the receiver, which can significantly harm the decoding at the receiver. Therefore, to use decode-and-forward and obtain the maximum diversity NR requires that the transmission rate is low enough so that all the relays can decode correctly. Thus, the decode-and-forward requires a substantial reduction of rate, especially for large R, and so we will therefore not consider it here. There are other disadvantages of decode-and-forward. Because of the decoding complexity, it causes both extra time delay and energy consumption.

In this paper, we will use the cooperative strategy called distributed space-time coding proposed in [1]. Thus, design the transmit signal at relay i as

$$\mathbf{t}_i = \sqrt{\frac{P_2}{P_1 + 1}} A_i \mathbf{r}_i, \tag{4}$$

a linear function of its received signal.[8] A_i is a $T \times T$ unitary matrix. As in [1], while within the framework of linear dispersion codes A_i can be arbitrary, to have a protocol that is equitable among different users and among different time instants we set A_i to be unitary. This simplifies the analysis of the PEP considerably, as will be seen in the following section.

Because $\mathrm{E}\,\mathrm{tr}\,\underline{\mathbf{s}}\underline{\mathbf{s}}^* = M$, $f_{mr}, v_{r,j}$ are $\mathcal{CN}(0,1)$, and $f_{mr}, \underline{\mathbf{s}}, v_{r,j}$ are independent, the average received power at relay i can be calculated to be

$$\mathrm{E}\,\mathbf{r}_i^*\mathbf{r}_i = \mathrm{E}\,(\sqrt{P_1T/M}\underline{\mathbf{s}}\mathbf{f}_i + \mathbf{v}_i)^*(\sqrt{P_1T/M}\underline{\mathbf{s}}\mathbf{f}_i + \mathbf{v}_i) = (P_1+1)T.$$

Therefore, the average transmit power at relay i is

$$\mathrm{E}\,\mathbf{t}_i^*\mathbf{t}_i = \frac{P_2}{P_1+1}\mathrm{E}\,(A_i\mathbf{r}_i)^*(A_i\mathbf{r}_i) = \frac{P_2}{P_1+1}\mathrm{E}\,\mathbf{r}_i^*\mathbf{r}_i = P_2T,$$

which explains our normalization in (4): P_2 is the average transmit power for one transmission at every relay.

Let us now focus on the received signal. Clearly from (2) and (3),

$$\mathbf{x}_n = \sqrt{\frac{P_1P_2T}{(P_1+1)M}}\begin{bmatrix} A_1\underline{\mathbf{s}} & A_2\underline{\mathbf{s}} & \cdots & A_R\underline{\mathbf{s}} \end{bmatrix}\begin{bmatrix} \mathbf{f}_1 g_{1n} \\ \mathbf{f}_2 g_{2n} \\ \vdots \\ \mathbf{f}_R g_{Rn} \end{bmatrix} + \sqrt{\frac{P_2}{P_1+1}}\sum_{i=1}^R g_{in}A_i\mathbf{v}_i + \mathbf{w}.$$

By defining

$$\begin{aligned} X &= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{bmatrix}, \quad S = \begin{bmatrix} A_1\underline{\mathbf{s}} & A_2\underline{\mathbf{s}} & \cdots & A_R\underline{\mathbf{s}} \end{bmatrix}, \\ \mathbf{g}_i &= \begin{bmatrix} g_{i1} & g_{i2} & \cdots & g_{iN} \end{bmatrix}, \quad H = \begin{bmatrix} \mathbf{f}_1\mathbf{g}_1 \\ \mathbf{f}_2\mathbf{g}_2 \\ \vdots \\ \mathbf{f}_R\mathbf{g}_R \end{bmatrix}, \end{aligned} \tag{5}$$

and

$$W = \begin{bmatrix} \sqrt{\frac{P_2}{P_1+1}}\sum_{i=1}^R g_{i1}A_i\mathbf{v}_i + \mathbf{w}_1 & \cdots & \sqrt{\frac{P_2}{P_1+1}}\sum_{i=1}^R g_{iN}A_i\mathbf{v}_i + \mathbf{w}_N \end{bmatrix}, \tag{6}$$

[8]In general, the transmit signal at the relays can be designed as any function of their received signals. Although only linear functions are used here, we conjecture that the diversity result cannot be improved using other more complicated designs. Also, more complicated designs make the noise analysis intractable.

the system equation can be written as

$$X = \sqrt{\frac{P_1 P_2 T}{M(P_1+1)}} S H + W. \tag{7}$$

Let us examine the dimension of each matrix. The received signal matrix X is $T \times N$. S which is actually a linear space-time code is $T \times MR$ since the A_i are $T \times T$, $\underline{\mathbf{s}}$ is $T \times M$, and there are R of them. $\mathbf{f}_i$ is $M \times 1$ and $\mathbf{g}_i$ is $1 \times N$. Therefore, the equivalent channel matrix H is $RM \times N$. W which is the equivalent noise matrix is $T \times N$. As in [1], S works like the space-time code in a multiple-antenna system. It is called the distributed space-time code since it has been generated distributively by the relays without having access to the transmit signal. The power constraint on S is:

$$\mathrm{E}\,\mathrm{tr}\, S^* S = \mathrm{E} \sum_{i=1}^{R} \mathrm{tr}\, \underline{\mathbf{s}}^* A_i^* A_i \underline{\mathbf{s}} = R \mathrm{E}\,\mathrm{tr}\, \underline{\mathbf{s}}^* \underline{\mathbf{s}} = RM.$$

Therefore, the code has the same normalization as a $T \times MR$ unitary matrix.

3 Pairwise Error Probability

To analyze the PEP, we have to determine the *maximum-likelihood (ML)* decoding rule. This requires the conditional *probability density function (PDF)* $\mathrm{P}\left(X | \underline{\mathbf{s}}_k\right)$, where $\underline{\mathbf{s}}_k \in \mathcal{S}$ and $\mathcal{S}$ is the set of all possible transmit signal matrices.

Theorem 1. *Given that $\underline{\mathbf{s}}_k$ is transmitted, define*

$$S_k = \left[\begin{array}{cccc} A_1 \underline{\mathbf{s}}_k & A_2 \underline{\mathbf{s}}_k & \cdots & A_R \underline{\mathbf{s}}_k \end{array} \right].$$

Then conditioned on $\underline{\mathbf{s}}_k$, the rows of X are independently Gaussian distributed with the same variance $I_N + \frac{P_2}{P_1+1} G G^*$, where

$$G = \left[\begin{array}{ccc} g_{11} & \cdots & g_{R1} \\ \vdots & \ddots & \vdots \\ g_{1N} & \cdots & g_{RN} \end{array} \right].$$

The t-th row of X has mean $\sqrt{\frac{P_1 P_2 T}{M(P_1+1)}} [S_k]_t H$ with $[S_k]_t$ the t-th row of S_k. In other words,

$$\mathrm{P}\left(X | \underline{\mathbf{s}}_k\right) = \frac{1}{\pi^{NT} \det{}^T \left(I_N + \frac{P_2}{P_1+1} G G^* \right)} e^{-\mathrm{tr}\left(X - \sqrt{\frac{P_1 P_2 T}{M(P_1+1)}} S_k H \right) \left(I_N + \frac{P_2}{P_1+1} G G^* \right)^{-1} \left(X - \sqrt{\frac{P_1 P_2 T}{M(P_1+1)}} S_k H \right)^*}. \tag{8}$$

Proof: See Appendix A. □

In view of the above theorem, we should emphasize that for a wireless relay network with multiple antennas at the receiver, the columns of X are not independent although the rows of X are. (The covariance matrix of each row $I_N + \frac{P_2}{P_1+1}GG^*$ is not diagonal in general.) That is, the received signals at different *antennas* are not independent, whereas the received signals at different *times* are. This is the main reason that the PEP analysis in the new model is much more difficult than that of the network in [1], where X had only a single column.

With $\mathrm{P}\left(X|\underline{\mathbf{s}}_i\right)$ in hand, we can obtain the ML decoding and thereby analyze the PEP. The result follows.

Theorem 2 (ML decoding and the PEP Chernoff bound). *Define* $R_W = I_N + \frac{P_2}{P_1+1}GG^*$. *The ML decoding of the relay network is*

$$\arg\min_{\underline{\mathbf{s}}_k} \operatorname{tr}\left(X - \sqrt{\frac{P_1P_2T}{M(P_1+1)}}S_kH\right) R_W^{-1} \left(X - \sqrt{\frac{P_1P_2T}{M(P_1+1)}}S_kH\right)^* . \tag{9}$$

With this decoding, the PEP of mistaking $\underline{\mathbf{s}}_k$ *by* $\underline{\mathbf{s}}_l$*, averaged over the channel realization, has the following upper bound:*

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \quad \leq \quad \mathop{\mathrm{E}}_{f_{mi}, g_{in}} e^{-\frac{P_1P_2T}{4M(1+P_1)}\operatorname{tr}(S_k-S_l)^*(S_k-S_l)HR_W^{-1}H^*} . \tag{10}$$

Proof: See Appendix B. □

4 Power Allocation

The main purpose of this work is to analyze how the PEP decays with transmit power or with the receive SNR at the high SNR regime. The total power used in the whole network is $P = P_1 + RP_2$. Therefore, one natural question is how to allocate power between the transmitter and the relays if P is fixed. In this section, we find the optimum power allocation such that the PEP is minimized. Because of the expectations over f_{mi} and g_{in}, and also the dependency of R_W on g_{in}, the exact minimization of formula (10) is very difficult. Therefore, similar to the argument in [1], we recourse to an asymptotic argument for $R \to \infty$.

Notice that the (i,j)-th entry of $\frac{1}{R}GG^*$ is $\frac{1}{R}\sum_{r=1}^R g_{ri}\bar{g}_{rj}$. When $R \to \infty$, according to the law of large numbers, the off-diagonal entries of $\frac{1}{R}GG^*$ goes to zero while the diagonal entries approach 1 with probability 1 since $\sum_{r=1}^R |g|_{ri}^2$ has a gamma distribution with both mean and variance R. Therefore, it is reasonable to assume $\frac{1}{R}GG^* \approx I_N$ for large R, which is the same as $R_W \approx \left(1 + \frac{P_2R}{P_1+1}\right) I_N$.

Therefore, from (10),

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \rightarrow \underline{\mathbf{s}}_l\right) \lesssim \mathop{\mathrm{E}}_{f_{mi}, g_{in}} e^{-\frac{P_1 P_2 T}{4M(1+P_1+RP_2)} \mathrm{tr}\,(S_k - S_l)^*(S_k - S_l) H H^*}.$$

Since S_k, S_l, and H are independent of P_1 and P_2, minimizing the PEP is now equivalent to maximizing $\frac{P_1 P_2 T}{4M(1+P_1+RP_2)}$. This is exactly the same power allocation problem that appeared in [1]. Therefore, with the same argument, we can conclude that the optimum solution is to set

$$P_1 = \frac{P}{2} \quad \text{and} \quad P_2 = \frac{P}{2R}. \tag{11}$$

That is, the optimum power allocation is such that the transmitter uses half the total power and the relays share the other half. When the number of relay nodes are large, which is the case for many sensor networks, every relay spends only a very small amount of power to help the transmitter.

Note that as discussed in Section 2.1, for the general network where the i-th relay has R_i antennas, the antennas are treated as R_i different relays in Figure 1. Therefore, it is easy to see that for this multiple-antenna-relay-node case, the optimum power allocation is such that the transmitter used half the total power as before, but every relay uses power that is proportional to its number of antennas. That is $P_1 = \frac{P}{2}$ and the power used at the i-th relay is $\frac{R_i P}{2\sum_{j=1}^{R} R_j}$.

With this power allocation, at high P, $\frac{P_1 P_2 T}{4M(1+P_1+RP_2)} \approx \frac{PT}{16MR}$. Therefore, the PEP satisfies

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \rightarrow \underline{\mathbf{s}}_l\right) \lesssim \mathop{\mathrm{E}}_{f_{mi}, g_{in}} e^{-\frac{PT}{16MR} \mathrm{tr}\,(S_k - S_l)^*(S_k - S_l) H H^*}. \tag{12}$$

It is easy to see that the expected receive SNR of the system is $\frac{P_1 P_2 T}{4(1+P_1+RP_2)}$. Therefore, this optimal power allocation also maximizes the expected receive SNR. We should emphasis that this result is only valid for the wireless relay network described in Section 2, in which all channels are assumed to be i.i.d. Rayleigh fading. It may not be optimal if these assumptions are not met.

5 Diversity Analysis for $R \rightarrow \infty$

5.1 Basic Results

As mentioned earlier, to obtain the diversity, we have to compute the expectations over f_{mi} and g_{in} in (12). We shall do this rigorously in Section 6. However, since the calculation is detailed and gives little insight, in this section, we give a simple asymptotic derivation for the case where the number of relay nodes approaches infinity, that is $R \to \infty$. As discussed in the previous section, when R is large, we can make the approximation $R_W \approx \left(1+\frac{P_2R}{P_1+1}\right) I_N$ and then obtain formula (12). Denote the n-th column of H as $\mathbf{h}_n$. From (5), $\mathbf{h}_n = \mathcal{G}_n\mathbf{f}$, where we have defined $\mathcal{G}_n = \operatorname{diag}\{g_{1n}I_M, \cdots, g_{Rn}I_M\}$ and $\mathbf{f} = \left[\begin{array}{ccc} \mathbf{f}_1 & \cdots & \mathbf{f}_R \end{array}\right]^t$. Therefore, from (12),

$$\begin{aligned} \mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) &\lesssim \mathop{\mathrm{E}}_{f_{mi},g_{in}} e^{-\frac{PT}{16MR}\operatorname{tr} H^*(S_k-S_l)^*(S_k-S_l)H} \\ &= \mathop{\mathrm{E}}_{f_{mi},g_{in}} e^{-\frac{PT}{16MR}\sum_{n=1}^N \mathbf{h}_n^*(S_k-S_l)^*(S_k-S_l)\mathbf{h}_n} \\ &= \mathop{\mathrm{E}}_{f_{mi},g_{in}} e^{-\frac{PT}{16MR}\mathbf{f}^*\left[\sum_{n=1}^N \mathcal{G}_n^*(S_k-S_l)^*(S_k-S_l)\mathcal{G}_n\right]\mathbf{f}}. \end{aligned}$$

Since $\mathbf{f}$ is white Gaussian with mean zero and variance I_{RM},

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \mathop{\mathrm{E}}_{g_{in}} \det{}^{-1}\left[I_{RM} + \frac{PT}{16MR}\sum_{n=1}^N \mathcal{G}_n^*(S_k-S_l)^*(S_k-S_l)\mathcal{G}_n\right]. \tag{13}$$

Similar to the multiple-antenna case [5, 13] and the case of wireless relay networks with single-antenna nodes [1], the "full diversity" condition can be obtained from (13). It is easy to see that if $S_k - S_l$ drops rank, the exponent of P in the right side of (13) increases. That is, the diversity of the system decreases. Therefore, at high transmit power, the Chernoff bound is minimized, which is equivalent to saying that the diversity is maximized, when $S_k - S_l$ is full-rank, or equivalently, $\det(S_k - S_l)^*(S_k - S_l) \neq 0$ for all $S_k \neq S_l \in \mathcal{S}$. Since the distributed space-time code S_k and S_l are $T \times MR$, there is no point in having MR larger than the coherence interval T. Thus, in the following, we will always assume $T \geq MR$.

Assume that the code is fully-diverse. From (13), roughly speaking, the larger the positive matrix $(S_k - S_l)^*(S_k - S_l)$, the smaller the upper bound. This improvement in the PEP because of the code design is called coding gain. Since the focus of this paper is on the diversity provided by the independent transmission routes from the antennas of the transmitter to the antennas of the receiver via the relays, the coding gain design, or the code optimization, is not an issue. Denote the minimum singular value of $(S_k - S_l)^*(S_k - S_l)$ by $\sigma^2_{\min}$. From the full diversity of the code, $\sigma^2_{\min} > 0$. Therefore, the right side of (13) can be further upper bounded as:

$$\begin{aligned} \mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) &\lesssim \mathop{\mathrm{E}}_{g_{in}} \det{}^{-1}\left[I_{RM} + \frac{PT\sigma^2_{\min}}{16MR}\sum_{n=1}^N \mathcal{G}_n^*\mathcal{G}_n\right] \\ &= \mathop{\mathrm{E}}_{g_{in}} \prod_{i=1}^R \left(1 + \frac{PT\sigma^2_{\min}}{16MR}\sum_{n=1}^N |g_{in}|^2\right)^{-M}. \end{aligned}$$

Define $g_i = \sum_{n=1}^N |g_{in}|^2$. Since g_{in} are i.i.d. $\mathcal{CN}(0,1)$, g_i are i.i.d. gamma distributed with PDF $p(g_i) = \frac{1}{(N-1)!}g_i^{N-1}e^{-g_i}$. Therefore,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{1}{(N-1)!^R}\left[\int_0^\infty \left(1 + \frac{PT\sigma^2_{\min}}{16MR}x\right)^{-M} x^{N-1}e^{-x}dx\right]^R.$$

By defining $y = 1 + \frac{PT\sigma_{\min}^2}{16MR}x$, which is equivalent to $x = (y-1)\frac{16MR}{PT\sigma_{\min}^2}$, and changing the variable in the integral from x to y, we have

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{1}{(N-1)!^R}\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{NR} e^{\frac{16MR^2}{PT\sigma_{\min}^2}}\left[\int_1^\infty \frac{(y-1)^{N-1}}{y^M} e^{-\frac{16MR}{PT\sigma_{\min}^2}y}dy\right]^R.$$

When the transmit power is high, that is, $P \gg 1$, $e^{\frac{16MR^2}{PT\sigma_{\min}^2}} \approx 1$. Therefore,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{1}{(N-1)!^R}\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{NR}\left[\sum_{l=0}^{N-1}\begin{pmatrix} N-1 \\ l \end{pmatrix}\int_1^\infty y^{l-M} e^{-\frac{16MR}{PT\sigma_{\min}^2}y}dy\right]^R.$$

The following theorem can be obtained by calculating the integral.

Theorem 3 (Diversity for $R \to \infty$). *Assume that $R \to \infty$, $T \geq MR$, and the distributed space-time code is full diverse. For large total transmit power P, by looking at only the highest order term of P, the PEP of mistaking $\underline{\mathbf{s}}_k$ with $\underline{\mathbf{s}}_l$ has the following upper bound:*

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{1}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{min}^2}\right)^{\min\{M,N\}R}\begin{cases} \left(\frac{2^{N-1}}{M-N}\right)^R P^{-NR} & \text{if } M > N \\ \left(\frac{\log^{1/M} P}{P}\right)^{MR} & \text{if } M = N \\ (N-M-1)!^R P^{-MR} & \text{if } M < N \end{cases}. \tag{14}$$

Therefore, the diversity of the wireless relay network is

$$d = \begin{cases} \min\{M,N\}R & \text{if } M \neq N \\ MR\left(1 - \frac{1}{M}\frac{\log\log P}{\log P}\right) & \text{if } M = N \end{cases}. \tag{15}$$

Proof: See Appendix C. □

5.2 Discussion

With the two-step protocol, it is easy to see that regardless of the cooperative strategy used at the relay nodes, the error probability is determined by the worse of the two transmission stages: the transmission from the transmitter to the relays and the transmission from the relays to the receiver. The PEP of the first stage cannot be better than the PEP of a multiple-antenna system with M transmit antennas and R receive antennas, whose optimal diversity is MR, while the PEP of the second stage can have diversity no larger than NR. Therefore, when $M \neq N$, according to the decay rate of the PEP, distributed space-time coding is optimal. For the case of $M = N$, the penalty on the decay rate is just $R\frac{\log\log P}{\log P}$, which is negligible when P is high. Therefore, distributed space-time coding is better than decode-and-forward since it achieves the optimum diversity gain without the rate constraint needed for decode-and-forward.

If we use the diversity definition in [14], since $\lim_{P\to\infty}\frac{\log\log P}{\log P}=0$, diversity $\min\{M,N\}R$ can be obtained.

The results in Theorem 3 are obtained by considering only the highest order term of P in the PEP formula. When analyzing the diversity gain, it is not only the highest order term that is important, but also how dominant it is. Therefore, we should analyze the contributions of the second highest order term and also other terms of P compared to that of the highest one. This is equivalent to analyzing how large the total transmit power P should be to have the terms given in (14) to dominate. The following remarks are on this issue. They can be observed from the proof of Theorem 3 in Appendix C.

Remarks:

1. If $|M-N|>1$, from (27) and (29), the second highest order term of P in the PEP formula behaves as $P^{-\min\{M,N\}R+1}$. The difference between the highest and the second highest order terms of P in the PEP is a P factor. Therefore, the highest order term is dominant when $P\gg 1$. In other words, contributions of the second highest order term and other lower order terms are negligible when $P\gg 1$.

2. If $M=N$, from (28), the second highest order term of P in the PEP formula is
$$\frac{2^{M-1}R}{(M-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\frac{\log^{R-1}P}{P^{MR}},$$
which has one less $\log P$ compared with the the highest order term. Therefore, the highest order term, $\frac{1}{(M-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\left(\frac{\log^{1/M}P}{P}\right)^{MR}$, is dominant if and only if $\log P\gg 1$, which is a much stronger condition than $P\gg 1$. When P is not very large, contributions of the second highest order term and even other lower order terms are not negligible.

3. If $|M-N|=1$, from (26) and (30), the second highest order term of P in the PEP formula behaves as $P^{-\min\{M,N\}R}\frac{\log P}{P}$. The difference between the highest and the second highest order terms is a $\frac{\log P}{P}$ factor. Therefore, the highest order term given in (14) is dominant if and only if $P\gg\log P$. This condition is weaker than the condition $\log P\gg 1$ in the previous case, however, it is still stronger than the normally used condition $P\gg 1$. Thus, although the situation is better than the case of $M=N$, still when P is not large enough, the second highest order term and even other lower order terms in the PEP formula are not negligible.

6 Diversity Analysis for the General Case

6.1 A Simple Derivation

The diversity analysis in the previous section is based on the assumption that the number of relays is very large. In this section, analysis on the PEP and the diversity results for the general case, networks with any number of relays, are given.

As discussed in Section 3, the main difficulty of the PEP analysis lies in the fact that the noise covariance matrix R_W is not diagonal. From (10), we can see that one way of upper bounding the PEP is to upper bound R_W. Since $R_W = I_N + \frac{P_2}{P_1+1}GG^* \geq 0$,

$$R_W \leq (\mathrm{tr}\, R_W) I_N = \sum_{n=1}^{N}\left(1+\frac{P_2}{P_1+1}\sum_{i=1}^{R}|g_{in}|^2\right) I_N = \left(N+\frac{P_2}{P_1+1}\sum_{n=1}^{N}\sum_{i=1}^{R}|g_{in}|^2\right) I_N.$$

Therefore, from (10) and using the power allocation given in (11),

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \rightarrow \underline{\mathbf{s}}_l\right) \lesssim \mathop{\mathrm{E}}_{f_{mi}, g_{in}} e^{-\frac{PT}{8MNR\left(1+\frac{1}{NR}\sum_{n=1}^{N}\sum_{i=1}^{R}|g_{in}|^2\right)}\mathrm{tr}\, H^*(S_k-S_l)^*(S_k-S_l)H}$$

when $P \gg 1$. If the space-time code is fully diverse, using similar argument in the previous section,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \rightarrow \underline{\mathbf{s}}_l\right) \lesssim \mathop{\mathrm{E}}_{g_{in}} \prod_{i=1}^{R}\left(1+\frac{PT\sigma_{\min}^2}{8MNR}\frac{g_i}{1+\frac{1}{NR}\sum_{i=1}^{R} g_i}\right)^{-M},$$

where, as before, $\sigma_{\min}^2$ is the minimum singular value of $(S_k-S_l)^*(S_k-S_l)$ and $g_i = \sum_{n=1}^{N}|g_{in}|^2$. Calculating this integral, the following theorem can be obtained.

Theorem 4 (Divesity for wireless relay network). *Assume that $T \geq MR$ and the distributed space-time code is full diverse. For large total transmit power P, by looking at the highest order terms of P, the PEP of mistaking $\underline{\mathbf{s}}_k$ by $\underline{\mathbf{s}}_l$ satisfies:*

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \rightarrow \underline{\mathbf{s}}_l\right) \lesssim \frac{1}{(N-1)!^R}\left(\frac{8MNR}{T\sigma_{\min}^2}\right)^{\min\{M,N\}R} \begin{cases} \left[\frac{M}{N(M-N)}\right]^R P^{-NR} & \text{if } M>N \\ \left(1+\frac{1}{N}\right)^R\left(\frac{\log^{1/M} P}{P}\right)^{MR} & \text{if } M=N \\ \left[\frac{1}{N}+(N-M-1)!\right]^R P^{-MR} & \text{if } M<N \end{cases}. \quad (16)$$

Therefore, the same diversity as in (15) is obtained.

Proof: See Appendix D. □

Although the same diversity is obtain as in the $R \rightarrow \infty$ case. There is a factor of N in formula (16), which does not appear in (14). This is because we upper bound R_W by $(\mathrm{tr}\, R_W)I_N$, whose expectation is N times the expectation of R_W, while in the previous section we approximate R_W by its expectation. This factor of N can be avoided by finding tighter upper bounds of R_W. In the following subsection, we analyze the maximum eigenvalue of R_W. Then in the Subsection 6.3, a PEP upper bound using the maximum eigenvalue of R_W is obtained.

6.2 The Maximum Eigenvalue of Wishart Matrix

Denote the maximum eigenvalue of $\frac{1}{R}GG^*$ as $\lambda_{\max}$. Since G is a random matrix, $\lambda_{\max}$ is a random variable. We first analyze the PDF and the *cumulative distribution function (CDF)* of $\lambda_{\max}$.

If entries of G are independent Gaussian distributed with mean zero and variance one, or equivalently, both the real and imaginary parts of every entry in G are Gaussian with mean zero and variance $\frac{1}{2}$, $\frac{1}{R}GG^*$ is known as the Wishart matrix. While there exists explicit formula for the distribution of the minimum eigenvalue of a Wishart matrix, surprisingly we could not find non-asymptotic formula for the maximum eigenvalue. Therefore, we calculate the PDF and CDF of $\lambda_{\max}$ from the joint distribution of all the eigenvalues of $\frac{1}{R}GG^*$ in this section. The following theorem has been proved.

Theorem 5. *G is an $N \times R$ matrix whose entries are i.i.d. $\mathcal{CN}(0,1)$.*

1. *The PDF of the maximum eigenvalue of $\frac{1}{R}GG^*$ is*

$$\mathrm{p}_{\lambda_{\max}}(\lambda) = \frac{R^{RN}\lambda^{R-N}e^{-R\lambda}}{\prod_{n=1}^{N}\Gamma(R-n+1)\Gamma(n)}\det F, \tag{17}$$

where F is an $(N-1)\times(N-1)$ Hankel matrix whose (i,j)-th entry equals $f_{ij} = \int_0^\lambda (\lambda-t)^2 t^{R-N+i+j-2}e^{-Rt}dt$.

2. *The CDF of the maximum eigenvalue of $\frac{1}{R}GG^*$ is*

$$\mathrm{P}\left(\lambda_{\max} \leq \lambda\right) = \frac{R^{RN}}{\prod_{n=1}^{N}\Gamma(R-n+1)\Gamma(n)}\det F', \tag{18}$$

where F' is an $N \times N$ Hankel matrix whose (i,j)-th entry equals $f_{ij} = \int_0^\lambda t^{R-N+i+j-2}e^{-Rt}dt$.

Proof: In Appendix E. □

A theoretical analysis of the PDF and CDF from (17) and (18) appears quite difficult. To understand $\lambda_{\max}$, we plot the two functions in figures 2 and 3 for different R and N. Figure 2 shows that the PDF has a peak at a value a bit larger than 1. As R increases, the peak becomes sharper. An increase in N shifts the peak right. However, the effect is smaller for larger R. From Figure 3, the CDF of $\lambda_{\max}$ grows rapidly around $\lambda = 1$ and

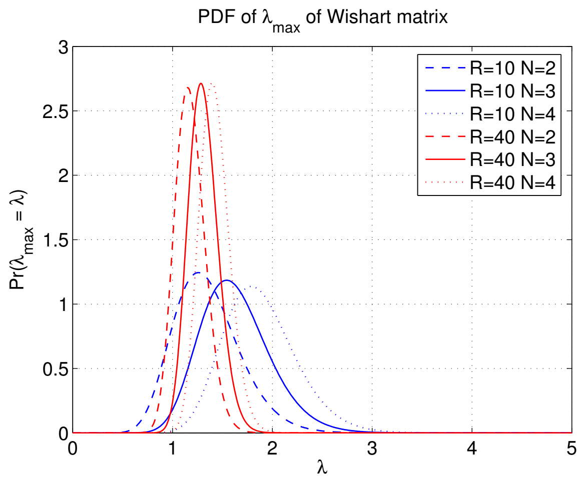

Figure 2: PDF of the maximum eigenvalue of $\frac{1}{R}GG^*$.

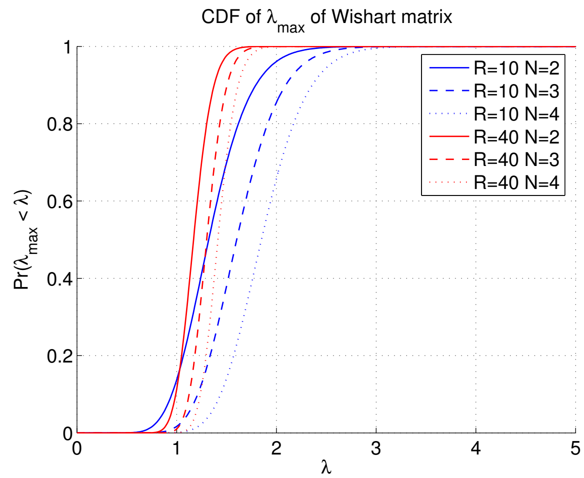

Figure 3: CDF of the maximum eigenvalue of $\frac{1}{R}GG^*$.

becomes very close to 1 soon after. The larger R, the faster the CDF grows. Similar to the PDF, an increase in N results in a right shift of the CDF. However, as R grows, the effect diminishes. This verifies the validity of the approximation $GG^* \approx RI_N$ in Section 6 for large R.

In the following corollary, we give an upper bound on the PDF. This result is used to derive the diversity result for general R in the next subsection.

Corollary 1. *The PDF of the maximum eigenvalue of* $\frac{1}{R}GG^*$ *can be upper bounded as*

$$\mathrm{p}_{\lambda_{\max}}(\lambda) \leq C'\lambda^{RN-1}e^{-R\lambda}, \tag{19}$$

where

$$C' = \frac{2^{N-1}R^{RN}}{\prod_{n=1}^{N}\Gamma(R-n+1)\Gamma(n)\prod_{n=1}^{N-1}(R-N+2n-1)(R-N+2n)(R-N+2n+1)} \tag{20}$$

is a constant that depends only on R *and* N.

Proof: From the proof of Theorem 5, F is a positive semidefinite matrix. Therefore $\det F \leq \prod_{n=1}^{N-1} f_{nn}$. From (17), f_{nn} can be upper bounded as

$$f_{nn} \leq \int_0^{\lambda}(\lambda-t)^2 t^{R-N+2n-2}dt = \frac{2}{(R-N+2n-1)(R-N+2n)(R-N+2n+1)}\lambda^{R-N+2n+1},$$

we have

$$\det F' \leq \frac{2^{N-1}}{\prod_{n=1}^{N-1}(R-N+2n-1)(R-N+2n)(R-N+2n+1)}\lambda^{RN-R+N-1}.$$

Thus, (19) is obtained. $\square$

6.3 Diversity Results for the General Case

If the maximum eigenvalue of $\frac{1}{R}GG^*$ is $\lambda_{\max}$, the maximum eigenvalue of $R_W = I_N + \frac{P_2}{P_1+1}GG^*$ is $1 + \frac{P_2 R}{P_1+1}\lambda_{\max}$, and therefore $R_W \leq \left(1 + \frac{P_2 R}{P_1+1}\lambda_{\max}\right) I_N$. From (2) and using the power allocation given in (11), we have

$$\begin{aligned} \mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l | \lambda_{\max} = c\right) &\leq \mathop{\mathrm{E}}_{f_{mr}, g_{rn}} e^{-\frac{P_1 P_2 T}{4M(1+P_1+P_2 R\lambda_{\max})}\mathrm{tr}\,(S_k - S_l)^*(S_k - S_l)HH^*} \\ &\lesssim \mathop{\mathrm{E}}_{f_{mr}, g_{rn}} e^{-\frac{PT}{8(1+\lambda_{\max})MR}\mathrm{tr}\,(S_k - S_l)^*(S_k - S_l)HH^*}. \end{aligned}$$

The only difference of the above formula with formula (12) is that the coefficient in the constant in the denominator of the exponent is $8(1+\lambda_{\max})$ now instead of 16. This makes sense since $c \to 1$ as $R \to \infty$. Therefore, using an argument similar to the proof of Theorem 3, at high total transmit power, by looking at the highest order terms of P,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l | \lambda_{\max} = c\right) \lesssim \frac{1}{(N-1)!^R}\left[\frac{8(1+c)MR}{T\sigma^2_{min}}\right]^{\min\{M,N\}R} \begin{cases} \left(\frac{2^{N-1}}{M-N}\right)^R P^{-NR} & \text{if } M > N \\ \left(\frac{\log^{1/M} P}{P}\right)^{MR} & \text{if } M = N \\ (N-M-1)!^R P^{-MR} & \text{if } M < N \end{cases}. \tag{21}$$

The following theorem can thus be obtained.

Theorem 6 (Divesity for wireless relay network). *Assume that $T \geq MR$ and the distributed space-time code is full diverse. For large total transmit power P, by looking at the highest order terms of P, the PEP of mistaking $\underline{\mathbf{s}}_k$ by $\underline{\mathbf{s}}_l$ can be upper bounded as:*

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{\hat{C}}{(N-1)!^R}\left(\frac{8MR}{T\sigma^2_{min}}\right)^{\min\{M,N\}R} \begin{cases} \left(\frac{2^{N-1}}{M-N}\right)^R P^{-NR} & \textit{if } M > N \\ \left(\frac{\log^{1/M} P}{P}\right)^{MR} & \textit{if } M = N \\ (N-M-1)!^R P^{-MR} & \textit{if } M < N \end{cases}, \tag{22}$$

where

$$\hat{C} = C' \sum_{i=0}^{\min\{M,N\}R} \begin{pmatrix} \min\{M,N\}R \\ i \end{pmatrix} \frac{(\min\{M,N\}R + i - 1)!}{R^{\min\{M,N\}R+i-1}}.$$

Therefore, the same diversity as in (15) is obtained.

Proof:

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) = \int_0^\infty \mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l | \lambda_{\max} = c\right)\mathrm{p}_{\lambda_{\max}}(c)dc \leq \int_0^\infty C' c^{RN-1} e^{-Rc}\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l | \lambda_{\max} = c\right)dc$$

using (19) in Corollary 1. From (21),

$$\mathrm{P}(\underline{\mathbf{s}}_i \to \underline{\mathbf{s}}_i) \lesssim \frac{C'}{(N-1)!^R}\left(\frac{8MR}{T\sigma^2_{min}}\right)^{\min\{M,N\}R}\int_0^\infty c^{RN-1}e^{-Rc}(1+c)^{\min\{M,N\}R}dc$$
$$\begin{cases} \left(\frac{2^{N-1}}{M-N}\right)^R P^{-NR} & \text{if } M>N \\ \left(\frac{\log P}{P^M}\right)^R & \text{if } M=N \\ (N-M-1)!^R P^{-MR} & \text{if } M<N \end{cases}.$$

Since

$$\int_0^\infty c^{RN-1}e^{-Rc}(1+c)^{\min\{M,N\}R}dc = \sum_{i=0}^{\min\{M,N\}R}\left(\begin{array}{c}\min\{M,N\}R \\ i\end{array}\right)\frac{(R\min\{M,N\}+i-1)!}{R^{R\min\{M,N\}+i-1}},$$

(22) is obtained. □

7 Conclusion and Discussion

In this paper, we generalize the idea of distributed space-time coding to wireless relay networks whose transmitter, receiver, and/or relays can have multiple antennas. We assume that the channel information is only available at the receiver. The ML decoding at the receiver and PEP of the network are analyzed. We have shown that for a wireless relay network with M antennas at the transmitter, N antennas at the receiver, a total of $\mathcal{R}$ antennas at all the relay nodes, and a coherence interval no less than $M\mathcal{R}$, an achievable diversity is $\min\{M,N\}\mathcal{R}$, if $M \neq N$, and $M\mathcal{R}\left(1-\frac{1}{M}\frac{\log\log P}{\log P}\right)$, if $M=N$, where P is the total power used in the whole network. This result shows the optimality of distributed space-time coding according to the diversity gain. It also shows the superiority of distributed space-time coding to decode-and-forward.

We also show that for a fixed total transmit power across the entire network, the optimal power allocation is for the transmitter to expend half the power and for the relays to share the other half such that the power used by every relay is proportional to the number of antennas it has.

There are several directions for future work that can be envisioned. In this paper, We have obtained an achievable diversity for the wireless relay network using the idea of distributed space-time coding. To achieve this diversity, the distributed space-time code

$$\left\{S_k = \left[\begin{array}{cccc} A_1\underline{\mathbf{s}}_k & A_2\underline{\mathbf{s}}_k & \cdots & A_R\underline{\mathbf{s}}_k\end{array}\right] | \underline{\mathbf{s}}_k \in \mathcal{S}\right\}$$

should be full diverse. That is, $\det(S_k-S_l)^*(S_k-S_l) \neq 0$ for all $\underline{\mathbf{s}}_k \neq \underline{\mathbf{s}}_l \in \mathcal{S}$. In addition to the diversity gain, the coding gain design or the code optimization is also an important issue since it effects the actually performance of the system. From (13), the larger $(S_k-S_l)^*(S_k-S_l)$, the better the coding gain. Therefore, possible design criterion are $\max\min_{\underline{\mathbf{s}}_k\neq\underline{\mathbf{s}}_l\in\mathcal{S}}\det(S_k-S_l)^*(S_k-S_l)$ and $\max\min_{\underline{\mathbf{s}}_k\neq\underline{\mathbf{s}}_l\in\mathcal{S}}\sigma(S_k-S_l)^*(S_k-S_l)$, where $\sigma(A)$ indicates the minimum singular value of A. More analysis is needed for this problem. However, it is beyond the scope of this paper.

Another important problem is the non-coherent case. In this work, we assume that the receiver knows all the channel information, which needs training from both the transmitter and the relay nodes. For networks with high mobility, this is not a practical assumption. Therefore, it should be interesting to see whether differential space-time coding technique can be generalized to this network.

The network model can be generalized in many ways too. The network analyzed in this paper has only one transmitter and receiver pair. When there are multiple transmitter-and-receiver pairs, interference plays an important role. Smart detection and/or interference cancellation techniques will be needed. Also, in our model, only the fading effect of the channels is considered. The achievable diversity when channel path-loss is in consideration is another important problem.

A Proof of Theorem 1

Proof: It obvious that since H is known and W is Gaussian, the rows of X are Gaussian. We only need to show that the rows of X are uncorrelated and that the mean and variance of the t-th row are $\sqrt{\frac{P_1P_2T}{(P_1+1)M}}[S_k]_tH$ and $I_N+\frac{P_2}{P_1+1}GG^*$, respectively.

The (t,n)-th entry of X can be written as

$$x_{tn}=\sqrt{\frac{P_1P_2T}{M(P_1+1)}}\sum_{i=1}^{R}\sum_{m=1}^{M}\sum_{\tau=1}^{T}f_{mi}g_{in}a_{i,t\tau}s_{k,\tau m}+\sqrt{\frac{P_2}{P_1+1}}\sum_{i=1}^{R}\sum_{\tau=1}^{T}g_{in}a_{i,t\tau}v_{i\tau}+w_{tn},$$

where $a_{i,t\tau}$ is the (t,τ)-th entry of A_i and $s_{k,\tau m}$ is the (τ,m)-th entry of $\underline{\mathbf{s}}_k$. With full channel information at the receiver,

$$\mathrm{E}\,x_{tn}=\sqrt{\frac{P_1P_2T}{M(P_1+1)}}\sum_{i=1}^{R}\sum_{m=1}^{M}\sum_{\tau=1}^{T}f_{mi}g_{in}a_{i,t\tau}s_{k,\tau m}.$$

Therefore, the mean of the t-th row is $\sqrt{\frac{P_1P_2T}{M(P_1+1)}}[S_k]_tH$. Since $\mathbf{v}_i$, $\mathbf{w}_n$, and $\underline{\mathbf{s}}_k$ are independent,

$$\begin{aligned}
\text{Cov}\ (x_{t_1n_1}, x_{t_2n_2}) &= \text{E}\,(x_{t_1n_1} - \text{E}\,x_{t_1n_1})\overline{(x_{t_2n_2} - \text{E}\,x_{t_2n_2})} \\
&= \frac{P_2}{P_1+1}\sum_{i_1=1}^{R}\sum_{\tau_1=1}^{T}\sum_{i_2=1}^{R}\sum_{\tau_2=1}^{T}\text{E}\,g_{i_1n_1}a_{i_1,t_1\tau_1}v_{r_1\tau_1}\bar{g}_{i_2n_2}\bar{a}_{i_2,t_2\tau_2}\bar{v}_{i_2\tau_2} + \text{E}\,w_{t_1n_1}\bar{w}_{t_2n_2} \\
&= \frac{P_2}{P_1+1}\sum_{i=1}^{R}\sum_{\tau=1}^{T}a_{i,t_1\tau}\bar{a}_{i,t_2\tau}g_{in_1}\bar{g}_{in_2} + \delta_{n_1n_2}\delta_{t_1t_2} \\
&= \delta_{t_1t_2}\left(\frac{P_2}{P_1+1}\sum_{r=1}^{R}g_{in_1}\bar{g}_{in_2} + \delta_{n_1n_2}\right) \\
&= \delta_{t_1t_2}\left(\frac{P_2}{P_1+1}\begin{bmatrix} g_{1n_1} & \cdots & g_{Rn_1}\end{bmatrix}\begin{bmatrix}\bar{g}_{1n_2} \\ \vdots \\ \bar{g}_{Rn_2}\end{bmatrix} + \delta_{n_1n_2}\right).
\end{aligned}$$

The fourth equality is true since A_i are unitary. Therefore, the rows of X are independent since the covariance of $x_{t_1n_1}$ and $x_{t_2n_2}$ is zero when $t_1 \neq t_2$. It is also easy to see that the variance matrix of each row is $I_N+\frac{P_2}{P_1+1}GG^*$. Therefore,

$$\text{P}\,([X]_t|\underline{\mathbf{s}}_k) = \frac{1}{\pi^N\det\left(I_N+\frac{P_2}{P_1+1}GG^*\right)}e^{-\text{tr}\left[X-\sqrt{\frac{P_1P_2T}{M(P_1+1)}}S_kH\right]_t\left(I_N+\frac{P_2}{P_1+1}GG^*\right)^{-1}\left[X-\sqrt{\frac{P_1P_2T}{M(P_1+1)}}S_kH\right]_t^*},$$

from which (8) can be obtained. □

B Proof of Theorem 2

Proof: It is straightforward to obtain the ML decoding formula (9) from formula (8). For any $\lambda > 0$, the PEP of mistaking $\underline{\mathbf{s}}_k$ by $\underline{\mathbf{s}}_l$ has the following Chernoff upper bound [18, 13]:

$$\text{P}\,(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l) \le \text{E}\,e^{\lambda\left(\log\text{P}\,(X|\underline{\mathbf{s}}_l) - \log\text{P}\,(X|\underline{\mathbf{s}}_k)\right)}.$$

Since $\underline{\mathbf{s}}_k$ is transmitted, $X = \sqrt{\frac{P_1P_2T}{M(P_1+1)}}S_kH + W$. From (8),

$$\begin{aligned}
&\log\text{P}\,(\mathbf{x}|\underline{\mathbf{s}}_l) - \log\text{P}\,(\mathbf{x}|\underline{\mathbf{s}}_k) \\
= &-\text{tr}\left[\frac{P_1P_2T}{M(P_1+1)}(S_k-S_l)HR_W^{-1}H^*(S_k-S_l)^* + \sqrt{\frac{P_1P_2T}{M(P_1+1)}}(S_k-S_l)HR_W^{-1}W^*\right. \\
&\left. + \sqrt{\frac{P_1P_2T}{M(P_1+1)}}WR_W^{-1}H^*(S_k-S_l)^*\right].
\end{aligned}$$

From the proof of Theorem 1, W is also Gaussian distributed with mean zero and its variance is the same as that of X. Therefore,

$$
\begin{aligned}
& \mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \\
\leq\ & \mathop{\mathrm{E}}_{f_{mi},g_{in},W} e^{-\lambda \mathrm{tr}\left[\frac{P_1P_2T}{M(P_1+1)}(S_k-S_l)HR_W^{-1}H^*(S_k-S_l)^*+\sqrt{\frac{P_1P_2T}{M(P_1+1)}}(S_k-S_l)HR_W^{-1}W^*+\sqrt{\frac{P_1P_2T}{M(P_1+1)}}WR_W^{-1}H^*(S_k-S_l)^*\right]} \\
=\ & \mathop{\mathrm{E}}_{f_{mi},g_{in}} e^{-\lambda(1-\lambda)\frac{P_1P_2T}{M(1+P_1)}\mathrm{tr}\,(S_k-S_l)HR_W^{-1}H^*(S_k-S_l)^*} \int \frac{e^{-\mathrm{tr}\left(\lambda\sqrt{\frac{P_1P_2T}{M(P_1+1)}}(S_k-S_l)H+W\right)R_W^{-1}\left(\lambda\sqrt{\frac{P_1P_2T}{M(P_1+1)}}(S_k-S_l)H+W\right)^*}}{\pi^{NT}\det{}^{-1}R_W}dW \\
=\ & \mathop{\mathrm{E}}_{f_{mi},g_{in}} e^{-\lambda(1-\lambda)\frac{P_1P_2T}{M(1+P_1)}\mathrm{tr}\,(S_k-S_l)^*(S_k-S_l)HR_W^{-1}H^*}.
\end{aligned}
$$

Choosing $\lambda = \frac{1}{2}$, which maximizes $\lambda(1-\lambda)$ and therefore minimizes the right side of the above formula, (10) is obtained. ☐

C Proof of Theorem 3

Proof: Define

$$
I = \sum_{l=0}^{N-1} \binom{N-1}{l} \int_1^\infty y^{l-M} e^{-\frac{16MR}{PT\sigma_{\min}^2}y} dy.
$$

We first give three integral equalities that will be used later.

$$
\int_u^\infty x^n e^{-\mu x} dx = e^{-u\mu} \sum_{k=0}^{n} \frac{n!}{k!} \frac{u^k}{\mu^{n-k+1}}, \qquad u > 0, \Re\mu > 0, n = 0, 1, 2, \cdots \tag{23}
$$

$$
\int_u^\infty \frac{e^{-\mu x}}{x^{n+1}} dx = (-1)^{n+1} \frac{\mu^n \mathbf{Ei}\,(-\mu u)}{n!} + \frac{e^{-\mu u}}{u^n} \sum_{k=0}^{n-1} \frac{(-1)^k \mu^k u^k}{n \cdots (n-k)}, \quad \mu > 0, n = 1, 2, \cdots \tag{24}
$$

$$
\int_u^\infty \frac{e^{-\mu x}}{x} dx = -\mathbf{Ei}\,(-\mu u), \qquad \Re\mu > 0, u \geq 0 \tag{25}
$$

To calculate I, we discuss the following cases separately.

C.1 Case I: $M < N$

In this case,

$$
\begin{aligned}
I = & \sum_{l=M}^{N-1} \binom{N-1}{l} \int_1^\infty y^{l-M} e^{-\frac{16MR}{PT\sigma_{\min}^2}y} dy + \binom{N-1}{M-1} \int_1^\infty \frac{e^{-\frac{16MR}{PT\sigma_{\min}^2}y}}{y} dy \\
& + \sum_{l=0}^{M-2} \binom{N-1}{l} \int_1^\infty y^{-(M-l)} e^{-\frac{16MR}{PT\sigma_{\min}^2}y} dy.
\end{aligned}
$$

Using equalities (23-25) with $u=1$, $\mu=\frac{16MR}{PT\sigma_{\min}^2}$, and $n=l-M$ or $n=M-l-1$,

$$\begin{aligned} I &= \sum_{l=M}^{N-1}\left(\begin{array}{c} N-1 \\ l \end{array}\right)(l-M)!\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(l-M+1)} + \left(\begin{array}{c} N-1 \\ M-1 \end{array}\right)\log P \\ &\quad + \sum_{l=0}^{M-2}\left(\begin{array}{c} N-1 \\ l \end{array}\right)\frac{1}{M-l-1} + \text{lower order terms of } P. \end{aligned}$$

By only looking at the highest order term of P, which is in the first term with $l=N-1$, we have

$$I=(N-M-1)!\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M)}+o\left(P^{-(N-M)}\right).$$

Therefore,

$$\begin{aligned} \mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) &\lesssim \frac{1}{(N-1)!^R}\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{NR}\left[(N-M-1)!\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M)}+o\left(\frac{1}{P^{MR}}\right)\right]^R \\ &= \left[\frac{(N-M-1)!}{(N-1)!}\right]^R\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\frac{1}{P^{MR}}+o\left(\frac{1}{P^{MR}}\right). \end{aligned}$$

While analyzing the performance of the system at high transmit power P, not only is the highest order term of P important but also how fast other terms decay with respect to it. Therefore, we should also look at the second highest order term of P. To do this, we have to consider two different cases.

If $N=M+1$,

$$\begin{aligned} I &= \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-1}+M\left[-\mathbf{Ei}\left(-\frac{16MR}{PT\sigma_{\min}^2}\right)\right]+O(1) \\ &= \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-1}+M\log P+O(1) \end{aligned}$$

Therefore,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{1}{M!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\frac{1}{P^{MR}}+\frac{RM}{M!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR+1}\frac{\log P}{P^{MR+1}}+o\left(\frac{\log P}{P^{MR+1}}\right). \tag{26}$$

The second highest order term of P in the PEP behaves as $\frac{\log P}{P^{MR+1}}=P^{-\left(MR+1-\frac{\log\log P}{\log P}\right)}$.

If $N>M+1$,

$$\begin{aligned} I &= (N-M-1)!\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M)}+(N-1)(N-M-2)!\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M-1)}+o\left(P^{N-M-1}\right) \\ &= \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M)}\left[(N-M-1)!+(N-1)(N-M-2)!\frac{16MR}{PT\sigma_{\min}^2}+o\left(\frac{1}{P}\right)\right]. \end{aligned}$$

Therefore,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{(N-M-1)!^R}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\frac{1}{P^{MR}} + \\ \frac{(N-1)(N-M-2)(N-M-1)!^{R-1}}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR+1}\frac{1}{P^{MR+1}} + o\left(\frac{1}{P^{MR+1}}\right). \qquad (27)$$

C.2 Case II: $M = N$

In this case,

$$I = \int_1^\infty \frac{e^{-\frac{16MR}{PT\sigma_{\min}^2}y}}{y}dy + \sum_{l=0}^{N-2}\begin{pmatrix} N-1 \\ l \end{pmatrix}\int_1^\infty y^{-(M-l)}e^{-\frac{16MR}{PT\sigma_{\min}^2}y}dy$$

Using (25) with $\mu = \frac{16MR}{PT\sigma_{\min}^2}$ and $u = 1$, and (24) with $u = 1$ and $n = M - l - 1$, we have

$$\begin{aligned} I &= \log P + \sum_{l=0}^{N-2}\begin{pmatrix} N-1 \\ l \end{pmatrix}\frac{1}{M-l-1} + \text{lower order terms of } P \\ &< \log P + 2^{N-1} + \text{lower order terms of } P. \end{aligned}$$

Therefore,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{1}{(M-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\frac{\log^R P}{P^{MR}} + \frac{2^{N-1}R}{(M-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\frac{\log^{R-1}P}{P^{MR}} + o\left(\frac{\log^{R-1}P}{P^{MR}}\right). \quad (28)$$

Also, the second highest order term of P in the PEP behaves as $\frac{log^{R-1}P}{P^{RM}}$ and the next term has one $\log P$ less and so on.

C.3 Case III: $M > N$

In this case,

$$I = \sum_{l=0}^{M-2}\begin{pmatrix} N-1 \\ l \end{pmatrix}\int_1^\infty y^{-(M-l)}e^{-\frac{16MR}{PT\sigma_{\min}^2}y}dy.$$

Using(24) with $u = 1$, $\mu = \frac{16MR}{PT\sigma_{min}^2}$, and $n = M - l - 1$,

$$I = \sum_{l=0}^{N-1}\begin{pmatrix} N-1 \\ l \end{pmatrix}\frac{1}{M-l-1} + \text{lower order terms of } P.$$

Thus,

$$\begin{aligned}
\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) &\lesssim \frac{1}{(N-1)^R}\left(\frac{16MR}{PT\sigma_{min}^2}\right)^{NR}\left[\sum_{l=0}^{N-1}\left(\begin{array}{c} N-1 \\ l \end{array}\right)\frac{1}{M-l-1}+o(1)\right]^R \\
&= \left[\frac{1}{(N-1)!}\sum_{l=0}^{N-1}\left(\begin{array}{c} N-1 \\ l \end{array}\right)\frac{1}{M-l-1}\right]^R\left(\frac{16MR}{T\sigma_{min}^2}\right)^{NR}P^{-NR}+o\left(P^{-NR}\right).
\end{aligned}$$

We can further upper bound the PEP to get a simpler formula. Notice that $\frac{1}{M-l-1} \leq \frac{1}{M-N}$. Thus,

$$\begin{aligned}
\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) &\lesssim \left[\frac{1}{(M-N)(N-1)!}\sum_{l=0}^{N-1}\left(\begin{array}{c} N-1 \\ l \end{array}\right)\right]^R\left(\frac{16MR}{T\sigma_{min}^2}\right)^{NR}P^{-NR} \\
&\leq \left[\frac{2^{N-1}}{(M-N)(N-1)!}\right]^R\left(\frac{16MR}{T\sigma_{min}^2}\right)^{NR}P^{-NR}.
\end{aligned}$$

As discussed before, we also want to see how dominant the highest order term of P given in the above formula is. If $M > N+1$, $M-l-2 > N+1-(N-1)-2 = 0$. From (24),

$$I < \frac{2^{N-1}}{M-N} - \frac{2^{N-1}}{(M-N)(M-N-1)}\frac{16MR}{PT\sigma_{min}^2}+o\left(\frac{1}{P}\right).$$

Therefore,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \left[\frac{2^{N-1}}{(M-N)(N-1)!}\right]^R\left(\frac{16MR}{T\sigma_{min}^2}\right)^{NR}\left[\frac{1}{P^{NR}}+\frac{R}{M-N-1}\left(\frac{16MR}{T\sigma_{min}^2}\right)\frac{1}{P^{NR+1}}\right]+o\left(\frac{1}{P^{NR+1}}\right). \quad (29)$$

The second highest order term in the PEP behaves as $\frac{1}{P^{NR+1}}$. If $M = N+1$,

$$I < 2^{N-1} + \frac{16MR}{T\sigma_{min}^2}\frac{\log P}{P}+O\left(\frac{1}{P}\right).$$

Therefore,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{2^{R(N-1)}}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{min}^2}\right)^{NR}\frac{1}{P^{NR}}+\frac{2^{(R-1)(N-1)}R}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{min}^2}\right)^{NR+1}\frac{\log P}{P^{NR+1}}+o\left(\frac{\log P}{P^{NR+1}}\right), \quad (30)$$

which indicates that the second highest order term in the PEP behaves as $\frac{\log P}{P^{NR+1}} = R^{-\left(NR+1-\frac{\log\log P}{\log P}\right)}$. □

D Proof of Theorem 4

Proof: Since g_i have PDF $p(g_i) = \frac{1}{(N-1)!}g_i^{N-1}e^{-g_i}$,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \leq \sum_{r=0}^{R}\sum_{1\leq i_1<\cdots<i_r\leq R} T_{i_1,\cdots,i_r},$$

where

$$T_{i_1,\cdots,i_r} = \frac{1}{(N-1)!^R} \underbrace{\int\cdots\int}_{\substack{\text{the } i_1,\cdots i_r\text{-th integrals} \\ \text{are from } x \text{ to } \infty, \\ \text{others are from } 0 \text{ to } x}} \prod_{i=1}^{R} \left(1+\frac{PT\sigma_{\min}^2}{8MNR}\frac{g_i}{1+\frac{1}{NR}\sum_{i=1}^{R} g_i}\right)^{-M} g_i^{N-1}e^{-g_i}dg_1\cdots dg_R$$

and x is any positive real number. Let us calculate $T_{1,\cdots,r}$ first.

$$\begin{aligned}
T_{1,\cdots,r} &= \frac{1}{(N-1)!^R}\underbrace{\int_x^\infty\cdots\int_x^\infty}_{r}\underbrace{\int_0^x\cdots\int_0^x}_{R-r}\prod_{i=1}^{R}\left(1+\frac{PT\sigma_{\min}^2}{8MNR}\frac{g_i}{1+\frac{1}{NR}\sum_{i=1}^{R} g_i}\right)^{-M} g_i^{N-1}e^{-g_i}dg_1\cdots dg_R \\
&< \frac{1}{(N-1)!^R}\int_x^\infty\cdots\int_x^\infty\prod_{i=1}^{r}\left(\frac{PT\sigma_{\min}^2}{8MNR}\frac{g_i}{1+\frac{R-r}{NR}x+\frac{1}{NR}\sum_{i=1}^{r} g_i}\right)^{-M} g_i^{N-1}e^{-g_i}dg_1\cdots dg_r \\
&\qquad \int_0^x\cdots\int_0^x\prod_{i=r+1}^{R} g_i^{N-1}e^{-g_i}dg_{r+1}\cdots dg_R \\
&= \frac{1}{(N-1)!^R}\left(\frac{PT\sigma_{\min}^2}{8MNR}\right)^{-rM}\gamma^{R-r}(N,x) \\
&\qquad \int_x^\infty\cdots\int_x^\infty\left(1+\frac{R-r}{NR}x+\frac{1}{NR}\sum_{i=1}^{r} g_i\right)^{rM}\prod_{i=1}^{r}\frac{e^{-g_i}}{g_i^{M-N+1}}dg_1\cdots dg_r,
\end{aligned}$$

where $\gamma(n,x)$ is the incomplete gamma function [17]. We should choose x so that the diversity is maximized. Define $x=\beta P^\alpha$, where β is a positive constant and α is any real constant. The value of β doesn't affect the diversity. Here, to have the PEP result consistent with formula (14) in Section 6, we set $\beta=\left(\frac{T\sigma_{\min}^2}{8MNR}\right)^\alpha$. Therefore, choosing the optimal (in the sense of maximizing the diversity) x is equivalent to choosing the optimal α. If $\alpha>0$, the $r=0$ term in the PEP upper bound is

$$\frac{1}{(N-1)!^R}\gamma^R(N,P^\alpha)=1+o(1).$$

Therefore, having α positive is not optimal according to diversity. Similarly, if $\alpha=0$, $x=1$. The $r=0$ term in the PEP upper bound, $\frac{1}{(N-1)!^R}\gamma^R(N,1)$, is a constant. Therefore, α should be negative. Thus,

$$\gamma(N,x)=\frac{1}{N}x^N+o(x^N)=\frac{1}{N}\beta^N P^{\alpha N}+o(P^{\alpha N}).$$

We are only interested in the highest order term of P. When P is large, $\frac{R-r}{NR}x$ is negligible compared with 1. Therefore,

$$T_{1,\cdots,r}\lesssim\frac{1}{(N-1)!^R N^{R-r}}\left(\frac{T\sigma_{\min}^2}{8MNR}\right)^{-rM+\alpha N(R-r)}P^{-rM+\alpha N(R-r)}\Lambda,$$

where we have defined

$$\Lambda=\int_x^\infty\cdots\int_x^\infty\left(1+\frac{1}{NR}\sum_{i=1}^{r} g_i\right)^{rM}\prod_{i=1}^{r}\frac{e^{-g_i}}{g_i^{M-N+1}}dg_1\cdots dg_r.$$

Consider the expansion of $\left(A+\sum_{i=1}^{k}\lambda_i\right)^a$ into monomial terms:

$$\left(1+\frac{1}{NR}\sum_{i=1}^{r} g_i\right)^a=\sum_{j=0}^{a}\left(\sum_{1\le l_1<\cdots<l_j\le k}\ \sum_{\substack{i_1,\ldots,i_j\ge 1\\ \sum i_m\le a}} C(i_1,\ldots,i_j)\frac{1}{(NR)^{i_1+\cdots+i_j}} g_{l_1}^{i_1}g_{l_2}^{i_2}\cdots g_{l_j}^{i_j}\right),$$

where j denotes how many g_i are present, $l_1,\ldots,l_j$ are the subscripts of the g_i that appears, $i_m\ge 1$ indicates that g_{l_m} is taken to the i_m-th power, and finally

$$C(i_1,\ldots,i_j)=\binom{k}{i_1}\binom{k-i_1}{i_2}\cdots\binom{k-i_1-\cdots-i_{j-1}}{i_j}$$

counts how many times the term $g_{l_1}^{i_1}g_{l_2}^{i_2}\cdots g_{l_j}^{i_j}$ appears in the expansion. Thus,

$$\Lambda=\sum_{j=0}^{r}\ \sum_{1\le l_1<\cdots<l_j\le r}\ \sum_{\substack{i_1,\ldots,i_j\ge 1\\ \sum i_m\le r}} C(i_1,\ldots,i_j)\Lambda(j;l_1,\ldots,l_j;i_1,\ldots,i_j)$$

where

$$\Lambda(j;l_1,\ldots,l_j;i_1,\ldots,i_j)\ =\ \frac{1}{(NR)^{i_1+\cdots+i_j}}\left(\prod_{m=1}^{j}\int_x^\infty\frac{e^{-g_{l_m}}}{g_{l_m}^{M-N+1-i_m}}dg_{l_m}\right)\prod_{i\ne i_1,\ldots i_j}\int_x^\infty\frac{e^{-g_i}}{g_i^{M-N+1}}dg_i.$$

From (23)–(25), While $P\to\infty$, $\alpha<0$, and $n>0$,

$$\int_x^\infty\lambda^n e^{-\lambda}d\lambda=n!+o(1),\quad \int_x^\infty\frac{e^{-\lambda}}{\lambda}d\lambda=(-\alpha)\log P+o(\log P),\ \text{ and }\ \int_x^\infty\frac{e^{-\lambda}}{\lambda^{n+1}}d\lambda=\frac{1}{n}\beta^{-n}P^{-\alpha n}+o(P^{-\alpha n}).$$

Therefore, the highest order term of P in Λ is the $j=0$ term. If we only keep the highest oder term of P in Λ,

$$\Lambda=\prod_{i=1}^{r}\int_x^\infty\frac{e^{-g_i}}{g_i^{M-N+1}}dg_i\approx\begin{cases}\frac{1}{(M-N)^r}\beta^{-r(M-N)}P^{-r\alpha(M-N)} & \text{if } M>N\\ (-\alpha)^r\log^r P & \text{if } M=N\\ (N-M-1)!^r & \text{if } M<N\end{cases}.$$

From the symmetry of $g_1,\cdots,g_R$, we have $T_{i_1,\cdots,i_r}=T_{1,\cdots,r}$. Therefore,

$$\begin{aligned}\mathrm{P}\left(\underline{\mathbf{s}}_k\to\underline{\mathbf{s}}_l\right)\ &\le\ \sum_{r=0}^{R}\binom{R}{r}T_{1,\cdots,r}\\ &\lesssim\ \frac{1}{(N-1)!^R}\sum_{r=0}^{R}\binom{R}{r}\frac{1}{N^{R-r}}\left(\frac{T\sigma_{\min}^2}{8MNR}\right)^{-rM+\alpha N(R-r)}P^{-rM+\alpha N(R-r)}\\ &\qquad\begin{cases}\frac{1}{(M-N)^r}\left(\frac{T\sigma_{\min}^2}{8MNR}\right)^{-r\alpha(M-N)}P^{-r\alpha(M-N)} & \text{if } M>N\\ (-\alpha)^r\log^r P & \text{if } M=N\\ (N-M-1)!^r & \text{if } M<N\end{cases}.\end{aligned}$$

We should choose a negative α such that the exponent of the highest order term of P in the above formula is minimized. In other words, if we denote the exponent of the r-th term as $f(r)$, choose a $\alpha < 0$ such that $\max_r f(r)$ is minimized.

If $M > N$, $f(r) = -rM + \alpha N(R-r) - r\alpha(M-N) = \alpha NR - rM(1+\alpha)$. If $\alpha \le -1$, $f(r)$ is an increasing function of r. Thus, $\max_r f(r) = f(R) = -\alpha(M-N)R - MR$, which is minimized when α equals its maximum -1. If $\alpha \ge -1$, $f(r)$ is an decreasing function of r. Thus, $\max_r f(r) = f(0) = \alpha NR$, which is minimized when α equals its minimum -1. Therefore, we should set $\alpha = -1$. Therefore,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{\left(\frac{1}{N}+\frac{1}{M-N}\right)^R}{(N-1)!^R}\left(\frac{8MNR}{T\sigma^2_{\min}}\right)^{NR} P^{-NR} = \left[\frac{M/N}{(M-N)(N-1)!}\right]^R \left(\frac{8MNR}{T\sigma^2_{\min}}\right)^{NR} P^{-NR}.$$

If $M < N$, $f(r) = \alpha NR - rN(\alpha + \frac{M}{N})$. By similar argument, we should set $\alpha = -\frac{M}{N}$. Thus,

$$\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) \lesssim \frac{\left(\frac{1}{N}+(N-M-1)!\right)^R}{(N-1)!^R}\left(\frac{8MNR}{T\sigma^2_{\min}}\right)^{-MR} P^{-MR}.$$

If $M = N$, $f(r) = \alpha NR - rN\left(\alpha + 1 - \frac{1}{N}\frac{\log\log P}{\log P}\right)$. Using similar argument, the optimal choice of α is $1 - \frac{1}{N}\frac{\log\log P}{\log P}$. Therefore,

$$\begin{aligned}\mathrm{P}\left(\underline{\mathbf{s}}_k \to \underline{\mathbf{s}}_l\right) &\lesssim \frac{1}{(N-1)!^R}\left[\left(\frac{8MNR}{T\sigma^2_{\min}}\right)^{\frac{\log\log P}{\log P}}\frac{1}{N} + \left(1 - \frac{1}{N}\frac{\log\log P}{\log P}\right)\right]^R \left(\frac{8MNR}{T\sigma^2_{\min}}\right)^{-MR}\left(\frac{\log^{1/M} P}{P}\right)^{-MR} \\ &\approx \frac{\left(\frac{1}{N}+1\right)^R}{(N-1)!^R}\left(\frac{8MNR}{T\sigma^2_{\min}}\right)^{-MR}\left(\frac{\log^{1/M} P}{P}\right)^{-MR}.\end{aligned}$$

□

E Proof of Theorem 5

Proof: We first give a theorem that will be needed later.

Theorem 7. *Define $\Lambda = (\lambda_1, \cdots, \lambda_N)$. For any function f, g, and h,*

$$\int d\Lambda \prod_{i=1}^{N} f(\lambda_i) \det V_g(\Lambda) \det V_h(\Lambda) = N! \det F_{gh},$$

where $V_g(\Lambda) = \begin{bmatrix} g_0(\lambda_1) & \cdots & g_0(\lambda_N) \\ \vdots & \ddots & \vdots \\ g_{N-1}(\lambda_1) & \cdots & g_{N-1}(\lambda_N) \end{bmatrix}$, $V_h(\Lambda) = \begin{bmatrix} h_0(\lambda_1) & \cdots & h_0(\lambda_N) \\ \vdots & \ddots & \vdots \\ h_{N-1}(\lambda_1) & \cdots & h_{N-1}(\lambda_N) \end{bmatrix}$ *and*

$$F_{gh} = \int f(t) \begin{bmatrix} g_0(t) \\ \vdots \\ g_{N-1}(t) \end{bmatrix} \begin{bmatrix} h_0(t) & \cdots & h_{N-1}(t) \end{bmatrix} dt.$$

Define G' as a complex Gaussian matrix whose entries' real and imaginary parts have mean zero and variance one. Denote the ordered eigenvalue of $G'G'^*$ as $\lambda'_1 \geq \lambda'_2 \cdots \geq \lambda'_N$. It is well-known that the eigenvalues have the following joint distribution [19]:

$$P(\lambda'_1, \cdots, \lambda'_N) = C \prod_{i=1}^{N} \lambda_i'^{R-N} e^{-\frac{\lambda'_i}{2}} \prod_{1 \leq i < j \leq N} (\lambda'_i - \lambda'_j)^2, \tag{31}$$

where $C = \frac{2^{-RN}}{\prod_{n=1}^{N} \Gamma(R-n+1)\Gamma(n)}$ is a constant. Denote the ordered eigenvalues of $\frac{1}{R}GG^*$ as $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$. Therefore, $\lambda'_i = 2R\lambda_i$. The joint distribution of $\lambda_1, \cdots, \lambda_N$ is therefore

$$\begin{aligned} P(\lambda_1, \cdots, \lambda_N) &= P(\lambda'_1, \cdots, \lambda'_N) \frac{d\lambda'_1 \cdots d\lambda'_N}{d\lambda_1 \cdots d\lambda_N} \\ &= \det\left[\text{diag}\left\{2R, \cdots, 2R\right\}\right] P(2R\lambda_1, \cdots, 2R\lambda_N) \\ &= (2R)^N C \prod_{i=1}^{N} (2R\lambda_i)^{R-N} e^{-R\lambda_i} \prod_{1 \leq i < j \leq N} [2R(\lambda_i - \lambda_j)]^2 \\ &= C(2R)^{RN} \prod_{i=1}^{N} \lambda_i^{R-N} e^{-R\lambda_i} \prod_{1 \leq i < j \leq N} (\lambda_i - \lambda_j)^2. \end{aligned}$$

To get the PDF of λ_1, we have to do the integral over $\lambda_2, \cdots, \lambda_N$. Define $f(x) = (\lambda - x)^2 x^{R-N} e^{-Rx}$ and $g_i(x) = h_i(x) = x^{i-1}$.

$$\begin{aligned} & P(\lambda_{\max} = \lambda) = P(\lambda_1 = \lambda) \\ = & \int_{\lambda \geq \lambda_2 \geq \cdots \lambda_N} P(\lambda, \lambda_2, \cdots, \lambda_N) d\lambda_2 \cdots d\lambda_N \\ = & \frac{1}{(N-1)!} \int_0^{\lambda} \cdots \int_0^{\lambda} P(\lambda, \lambda_2, \cdots, \lambda_N) d\lambda_2 \cdots d\lambda_N \end{aligned}$$

$$
\begin{aligned}
&= \frac{C(2R)^{RN}}{(N-1)!}\lambda^{R-N}e^{-R\lambda}\int_0^\lambda\cdots\int_0^\lambda\prod_{i=2}^N(\lambda-\lambda_i)^2\lambda_i^{R-N}e^{-R\lambda_i}\prod_{2\le i<j\le N}(\lambda_i-\lambda_j)^2 d\lambda_2\cdots d\lambda_N \\
&= \frac{C(2R)^{RN}}{(N-1)!}\lambda^{R-N}e^{-R\lambda}\int_0^\lambda\cdots\int_0^\lambda\prod_{i=2}^N f(\lambda_i)\det V_g(\lambda_2,\cdots,\lambda_N)\det V_h(\lambda_2,\cdots,\lambda_N)\lambda_1\cdots\lambda_N \\
&= \frac{C(2R)^{RN}}{(N-1)!}\lambda^{R-N}e^{-R\lambda}(N-1)!\det F.,
\end{aligned}
$$

where in the second equality we have changed the integral space from ordered λ_i to unordered ones. From the symmetry of λ_i, we only need to divide the new value by $(N-1)!$. From Theorem 7,

$$
F = \int_0^\lambda g(t)\begin{bmatrix} 1 \\ t \\ \vdots \\ t^{N-2} \end{bmatrix}\begin{bmatrix} 1 & t & \cdots & t^{N-2} \end{bmatrix} dt.
$$

whose (i,j)-th entry is $f_{ij} = \int_0^\lambda(\lambda-t)^2 t^{R-N+i+j-2}e^{-Rt}dt$. The CDF of λ_1 can be obtained similarly. $\square$

References

[1] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *submitted to IEEE Transactions on Communications*, July 2004.

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

[3] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Info. Theory*, vol. 45, pp. 139–157, Jan. 1999.

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

[5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Info. Theory*, vol. 44, pp. 744–765, 1998.

[6] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part I: System description," *IEEE Transactions on Communications*, vol. 51, pp. 1927–1938, Nov. 2003.

[7] A. Sendonaris, E. Erkip, and B. Aazhang, “User cooperation diversity-part II: Implementation aspects and performance analysis,” *IEEE Transactions on Communications*, vol. 51, pp. 1939–1948, Nov. 2003.

[8] J. N. Laneman and G. W. Wornell, “Distributed space-time-coded protocols for exploiting cooperative diversity in wireless betwork,” *IEEE. Trans. on Info. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.

[9] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, “Fading relay channels: Performance limits and space-time signal design,” *IEEE Journal on Selected Areas in Communications*, pp. 1099 – 1109, Aug. 2004.

[10] Y. Hua, Y. Mei, and Y. Chang, “Wireless antennas-making wireless communications perform like wireline communications,” in *IEEE AP-S Topical Conference on Wireless Communication Technology*, Oct. 2003.

[11] Y. Chang and Y. Hua, “Application of space-time linear block codes to parallel wireless relays in mobile ad hoc networks,” in *the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, Nov. 2003.

[12] B. Hassibi and B. Hochwald, “High-rate codes that are linear in space and time,” *IEEE Trans. Info. Theory*, vol. 48, pp. 1804–1824, July 2002.

[13] B. M. Hochwald and T. L. Marzetta, “Unitary space-time modulation for multiple-antenna communication in Rayleigh flat-fading,” *IEEE Trans. Info. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[14] L. Zheng and D. Tse, “Diversity and multiplexing: a fundamental tradeoff in multiple antenna channels,” *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[15] A. F. Dana, M. Sharif, R. Gowaikar, B. Hassibi, and M. Effros, “Is broadcase plus multi-access optimal for gaussian wireless network?,” in *the 36th Asilomar Conf. on Signals, Systems and Computers*, Nov. 2003.

[16] M. Gasptar and M. Vetterli, “On the capacity of wireless networks: the relay case,” in *IEEE Infocom*, June 2002.

[17] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*. Academic Press, 6nd ed., 2000.

[18] H. L. V. Trees, *Detection, Estimation, and Modulation Theory–Part I*. New York: Wiley, 1968.

[19] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*. PhD Thesis, MIT, Dept. of Math., 1989.