Research Article

Diversity Analysis of Distributed Space-Time Codes in Relay Networks with Multiple Transmit/Receive Antennas

Yindi Jing[1] and Babak Hassibi[2]

[1] Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697, USA
[2] Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Correspondence should be addressed to Yindi Jing, yjing@uci.edu

The idea 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 $(1/P)^{\min\{M,N\}\mathcal{R}}$ if $M \neq N$ and $(\log^{1/M} P/P)^{M\mathcal{R}}$ 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 maximal diversity. 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.

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 [1–4]. Recently, with the increasing interestin ad hoc networks, researchers have been looking for methods to exploit spatial diversity using the antennas of different users in the network [5–24]. Many cooperative strategies are proposed, for example, amplify-and-forward (AF) [11, 13, 14, 16, 21, 23], decode-and-forward (DF) [9, 10, 14, 16, 22], and coded cooperation [15]. In [7], the authors proposed the use of space-time codes based on Hurwitz-Radon matrices in wireless relay networks.

This work follows the strategy of [5], where the idea of space-time coding devised for multiple-antennasystems is applied to the problem of communication over a wireless relay network. (Though having the same name, the distributed space-time coding idea in [5] is different from that in [14]. Similar ideas for networks with one and two relays have appeared in [6, 11].) In [5], 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 [25] among the relays. It is proved that without any channel knowledge at the relays, a diversity of $R(1 - \log\log P / \log P)$ can be achieved, where R is the 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. After the appearance of [5], code designs for distributed space-time coding have been proposed in [26–31] and the differential use of distributed space-time coding has been introduced in [32–35]. The references [36, 37] analyze the diversity-multiplexing tradeoff of distributed space-time coding. Distributed space-time coding in asynchronous networks is discussed in [38–43]. Other related papers can be found in [44–46].

This paper has two main contributions. First, we extend the idea of distributed space-time coding to wireless relay networks whose nodes have multiple antennas. Second and

more importantly, based on the pairwise error probability (PEP) analysis, we prove lower bounds on the diversity of this scheme. We use the same two-step transmission method in [5], 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 $M\mathcal{R}(1 - (1/M)(\log\log P/\log P))$ 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 errorprobability 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 larger 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 larger 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 $\mathcal{R}(\log\log P/\log P)$. 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 are explained in detail. A training scheme is also proposed. The PEP is first analyzed in Section 3. In Section 4, the diversity for the network with an infinite number of relays is discussed. Then, the diversity for the general case is obtained in Section 5. Section 6 contains the conclusion. Proofs of some of the technical theorems are given in Appendices A–D. In Appendix E, we discuss heterogeneous networks.

2. WIRELESS RELAY NETWORK

2.1. Network model and distributed space-time coding

We first introduce some notation. For a complex matrix A, $\overline{A}$, A^t, and A^* denote the conjugate, the transpose, and the Hermitian of A, respectively. $\det A$, $\text{rank } A$, and $\text{tr } A$ indicate the determinant, rank, and trace of A, respectively. $\vec{A}$ denotes the vectorization of A formed by stacking the columns of X into a single column vector. I_n denotes the $n \times n$ identity matrix and $0_{m,n}$ is the $m \times n$ matrix with all zero entries. 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 $\lim_{x\to\infty}(g(x)/f(x))$ is a constant. $h(x) = o(f(x))$ means that $\lim_{x\to\infty}(h(x)/f(x)) = 0$. $\lceil a \rceil$ is the minimal integer that is not less than a.

Consider a wireless network with $R + 2$ nodes which are placed randomly and independently according to some distribution. As shown in Figure 1, there are one transmit node and one receive node. All the other R nodes work
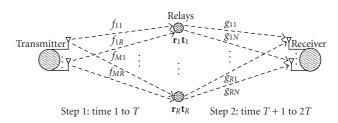

FIGURE 1: Wireless relay network with multiple-antenna nodes.

as relays. The transmitter has M transmit antennas, the receiver has N receive antennas, and the ith relay has R_i antennas. 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. This is one possible scheme. In general, the signal sent by one antenna of a relay can be designed using received signals at all 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. Therefore, to highlight the diversity results by simplifying notation and formulas, in the following, we assume that every relay has a single antenna. Denote the channel vector from the M antennas of the transmitter to the ith relay as $\mathbf{f}_i = [f_{1i} \cdots f_{Mi}]^t$, and the channels from the ith relay to the N antennas at the receiver as $\mathbf{g}_i = [g_{i1} \cdots g_{iN}]$. We use the block-fading model [2] by assuming a coherence interval T. From the two-step protocol that will be discussed in the following, we can see that we only need $\mathbf{f}_i$ to keep constant for the first step of the transmission and $\mathbf{g}_i$ to keep constant for the second step. It is thus good enough to choose T as the minimum of the coherence intervals of $\mathbf{f}_i$ and $\mathbf{g}_i$. Also, perfect symbol-level synchronization is assumed in this network model. For asynchronized networks, please refer to [38–43].

The information bits are encoded into $T \times M$ matrices $\underline{\mathbf{s}}$, whose mth column is the signal sent by the mth transmit antenna. For the power analysis, $\underline{\mathbf{s}}$ is normalized as

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

To send $\underline{\mathbf{s}}$ to the receiver, the same two-step strategy in [5] is used, as shown in Figure 1. In step one, the transmitter sends $\sqrt{P_1 T/M}\underline{\mathbf{s}}$. The average total power used at the transmitter for the T transmissions is $P_1 T$. The received signal vector and the noise vector at the ith relay are denoted as $\mathbf{r}_i$ and $\mathbf{v}_i$. In step two, the ith relay sends $\mathbf{t}_i$. The received signal and noise at the receiver are denoted as X and $\underline{\mathbf{w}}$. The noises are assumed to be i.i.d. $\mathcal{CN}(0, 1)$. Clearly,

$$\begin{aligned} \mathbf{r}_i &= \sqrt{P_1 T/M}\,\underline{\mathbf{s}}\mathbf{f}_i + \mathbf{v}_i, \\ X &= \begin{bmatrix} \mathbf{t}_1 & \cdots & \mathbf{t}_R \end{bmatrix} G + \underline{\mathbf{w}}, \end{aligned} \tag{2}$$

where $G = [\mathbf{g}_1^t \cdots \mathbf{g}_R^t]^t$.

We use distributed space-time coding proposed in [5] by designing the transmit signal at relay i as a linear function of its received signal:

$$\mathbf{t}_i = \sqrt{\frac{P_2}{P_1 + 1}} A_i \mathbf{r}_i, \tag{3}$$

where A_i is a predetermined $T \times T$ unitary matrix known to both the ith relay and the receiver. It is fixed during training and data transmissions. For various methods on how to design the A_i, see [26–31]. P_2 can be proved to be the average transmit power for one transmission at every relay. After some calculation, the system equation can be written as

$$X = \sqrt{\frac{P_1 P_2 T}{M(P_1 + 1)}} SH + W, \tag{4}$$

where

$$S = \begin{bmatrix} A_1 \underline{\mathbf{s}} & \cdots & A_R \underline{\mathbf{s}} \end{bmatrix}, \qquad H = \begin{bmatrix} (\mathbf{f}_1 \mathbf{g}_1)^t & \cdots & (\mathbf{f}_R \mathbf{g}_R)^t \end{bmatrix}^t, \tag{5}$$

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

The received signal matrix, X, is $T \times N$. S, which is $T \times MR$, is the linear distributed space-time code. Since $\mathbf{f}_i$ is $M \times 1$ and $\mathbf{g}_i$ is $1 \times N$, the equivalent channel matrix H is $RM \times N$. W, which is $T \times N$, is the equivalent noise matrix.

Define

$$R_W = I + \frac{P_2}{1 + P_1} G^* G. \tag{7}$$

The covariance matrix of the equivalent noise matrix can be proved to be $\overline{R_W}$. The diversity analysis in this paper is much more difficult than that in [5] because in networks with single-antenna nodes, the covariance matrix of the equivalent noise is a multiple of the identity matrix. Here, for the diversity result, we need to analyze the eigenvalues of R_W or find bounds on them.

2.2. Assumptions and training

In this paper, we assume that f_{mi} and g_{in} have independent Rayleigh distributions; that is, f_{mi} and g_{in} are independent circulant complex Gaussian random variables with zero mean. For simplicity, we also assume that f_{mi} and g_{in} have the same variance, which is 1. The heterogeneous case, in which every channel has a different variance, is discussed in Appendix E. The same diversity results can be obtained in heterogeneous networks. We make the practical assumption that the relays have no channel information. However, we do assume that the receiver has enough channel information to do coherent detection. Thus, a training process is needed.

For coherence ML decoding at the receiver, the receiver needs to know H and $\overline{R_W}$, or equivalently, H and G. We propose a training process that contains two steps and takes $M_p + 2N_p$ symbol periods (other training methods can also be envisioned, and the one proposed here is one possibility).

Each step mimics the training process of a multiple-antenna system [47] as its system equation has the same structure.

First, we estimate G, which takes M_p symbol periods. Let U_p be a predesigned full-rank $M_p \times R$ pilot matrix. The ith relay sends the ith column of U_p simultaneously. The receiver gets

$$Y_p = \sqrt{\frac{Q_p M_p}{R}} U_p G + \underline{\mathbf{w}}_p, \tag{8}$$

where Q_p is the power used at every relay and $\underline{\mathbf{w}}_p$ is the $M_p \times N$ noise matrix. Since there are RN unknowns (corresponding to the components of G) and $\min\{M_p, R\}N$ independent equations, we need $M_p \geq R$. We could estimate G from U_p using ML, MMSE, or other criteria.

Then, we estimate H using distributed space-time coding discussed in Section 2.1. This takes $2N_p$ symbol periods. The transmitter sends a full-rank $N_p \times M$ pilot signal matrix $\underline{\mathbf{s}}_p$ and the relays perform distributed space-time coding. From (4), the received signal can be written as

$$X_p = \sqrt{\frac{P_{1,p} P_{2,p} N_p}{M(P_{1,p} + 1)}} S_p H + W_p, \tag{9}$$

where $P_{1,p}$ and $P_{2,p}$ are the powers used at the transmitter and every relay and

$$S_p = \begin{bmatrix} A_1 \underline{\mathbf{s}}_p & \cdots & A_R \underline{\mathbf{s}}_p \end{bmatrix} \tag{10}$$

is the carefully designed $N_p \times MR$ pilot space-time codeword. Now, let us discuss the number of training symbols needed in this step. Note that G is known from the first training step. Define

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_1^t & \cdots & \mathbf{f}_R^t \end{bmatrix}^t. \tag{11}$$

By stacking the columns of X into one single column vector, we can rewrite (9) as

$$\begin{aligned}
\vec{X}_p &= \sqrt{\frac{P_{1,p} P_{2,p} N_p}{M(P_{1,p} + 1)}} \begin{bmatrix} S_p \mathrm{diag}\{g_{11} I_M, \ldots, g_{R1} I_M\} \\ \vdots \\ S_p \mathrm{diag}\{g_{1N} I_M, \ldots, g_{RN} I_M\} \end{bmatrix} \mathbf{f} + \vec{W}_p \\
&= \sqrt{\frac{P_{1,p} P_{2,p} N_p}{M(P_{1,p} + 1)}} \begin{bmatrix} g_{11} A_1 \underline{\mathbf{s}}_p & \cdots & g_{R1} A_R \underline{\mathbf{s}}_p \\ \vdots & \ddots & \vdots \\ g_{1N} A_1 \underline{\mathbf{s}}_p & \cdots & g_{RN} A_R \underline{\mathbf{s}}_p \end{bmatrix} \mathbf{f} + \vec{W}_p \\
&= \sqrt{\frac{P_{1,p} P_{2,p} N_p}{M(P_{1,p} + 1)}} \begin{bmatrix} g_{11} I_{N_p} & \cdots & g_{R1} I_{N_p} \\ \vdots & \ddots & \vdots \\ g_{1N} I_{N_p} & \cdots & g_{RN} I_{N_p} \end{bmatrix} \\
&\quad \times \mathrm{diag}\{A_1 \underline{\mathbf{s}}_p, \ldots, A_R \underline{\mathbf{s}}_p\} \mathbf{f} + \vec{W}_p.
\end{aligned} \tag{12}$$

Denote

$$\mathcal{H}_p = \begin{bmatrix} g_{11}I_{N_p} & \cdots & g_{R1}I_{N_p} \\ \vdots & \ddots & \vdots \\ g_{1N}I_{N_p} & \cdots & g_{RN}I_{N_p} \end{bmatrix} \mathrm{diag}\{A_1\underline{s}_p, \ldots, A_R\underline{s}_p\}. \tag{13}$$

The number of independent equations in (9) equals the rank of $\mathcal{H}_p$, which is $\min\{N_pN, N_pR, MR\}$. Since there are MR unknowns (corresponding to the components of $\mathbf{f}$), we need $\min\{N_pN, N_pR, MR\} \geq MR$, which is equivalent to

$$N_p \geq \max\left\{\lceil MR/N \rceil, M\right\}. \tag{14}$$

While this condition is satisfied, we could estimate $\mathbf{f}$ from X_p using ML, MMSE, or other criteria. The overall training process takes at least $R+2\max\{\lceil MR/N \rceil, M\}$ symbol periods. The optimal designs of U_p, Q_p, S_p (or $\underline{s}_p$), and $P_{1,p}$, $P_{2,p}$ are interesting issues. However, they are beyond the scope of this paper.

3. PAIRWISE ERROR PROBABILITY AND OPTIMAL POWER ALLOCATION

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

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

$$S_k = \begin{bmatrix} A_1\underline{s}_k & A_2\underline{s}_k & \cdots & A_R\underline{s}_k \end{bmatrix}. \tag{15}$$

Then conditioned on $\underline{s}_k$, the rows of X are independently Gaussian distributed with the same variance $\overline{R_W}$. The tth row of X has mean $\sqrt{P_1P_2T/M(P_1+1)}[S_k]_t H$ with $[S_k]_t$ being the tth row of S_k. Also,

$$P(X \mid \underline{s}_k)$$
$$= (\pi^N \det R_W)^{-T}$$
$$\times e^{-\mathrm{tr}(X - \sqrt{P_1P_2T/M(P_1+1)}S_kH)R_W^{-1}(X - \sqrt{P_1P_2T/M(P_1+1)}S_kH)^*}. \tag{16}$$

Proof. See Appendix A. ☐

In view of Theorem 1, 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 R_W 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 [5], where X had only a single column.

With $P(X \mid \underline{s}_k)$ 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). *The ML decoding of the relay network is*

$$\arg\min_{\underline{s}_k} \mathrm{tr}\left(X - \sqrt{\frac{P_1P_2T}{M(P_1+1)}}S_kH\right)$$
$$\times R_W^{-1}\left(X - \sqrt{\frac{P_1P_2T}{M(P_1+1)}}S_kH\right)^*. \tag{17}$$

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

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \leq \mathop{E}_{f_{mi}, g_{in}} e^{-(P_1P_2T/4M(1+P_1))\mathrm{tr}(S_k-S_l)^*(S_k-S_l)HR_W^{-1}H^*}. \tag{18}$$

Proof. The proof is omitted since it is the same as the proof of Theorem 1 in [5]. ☐

As both H and R_W are known at the receiver, sphere decoding can be used to perform the ML decoding in (17).

The main purpose of this work is to analyze how the PEP decays with the total transmit power. The total power used in the whole network is $P = P_1 + RP_2$. One natural question is how to allocate power between the transmitter and the relays if P is fixed. Notice that when $R \rightarrow \infty$, according to the law of large numbers, the off-diagonal entries of $(1/R)G^*G$ go to zero while the diagonal entries approach 1 with probability 1. It is thus reasonable to assume $(1/R)G^*G \approx I_N$ for large R. With this approximation, minimizing the PEP is now equivalent to maximizing $P_1P_2T/4M(1 + P_1 + RP_2)$. This is exactly the same power allocation problem in [5]. Therefore, we can conclude that the optimum solution is to set

$$P_1 = \frac{P}{2}, \qquad P_2 = \frac{P}{2R}. \tag{19}$$

That is, the optimum power allocation is such that the transmitter uses half the total power and the relays share the other half. As discussed in Section 2.1, for the general network where the ith relay has R_i antennas, the antennas are treated as R_i different relays. Therefore, in general, the optimum power allocation is such that the transmitter uses half the total power as before, but every relay uses a power that is proportional to its number of antennas, that is, $P_1 = P/2$ and the power used at the ith relay is $R_iP/2\mathcal{R}$.

4. DIVERSITY ANALYSIS FOR $R \rightarrow \infty$

4.1. Basic results

As mentioned earlier, to obtain the diversity, we have to compute the expectations over f_{mi} and g_{in} in (18). We will do this rigorously in Section 5. 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, $R \rightarrow \infty$. As discussed in the previous section, when R is large, we can make the approximation $R_W \approx (1 + P_2R/(P_1+1))I_N$. Denote the nth column of H as $\mathbf{h}_n$. From (5), $\mathbf{h}_n = \mathcal{G}_n\mathbf{f}$, where we

have defined $\mathcal{G}_n = \text{diag}\{g_{1n}I_M,\ldots,g_{Rn}I_M\}$. Therefore, from (18) and using the optimal power allocation in (19),

$$
\begin{aligned}
P(\underline{\mathbf{s}}_k \longrightarrow \underline{\mathbf{s}}_l) &\lesssim \underset{f_{mi},g_{in}}{\mathrm{E}}\ e^{-(PT/16MR)\mathrm{tr}H^*(S_k-S_l)^*(S_k-S_l)H} \\
&= \underset{f_{mi},g_{in}}{\mathrm{E}}\ e^{-(PT/16MR)\sum_{n=1}^N \mathbf{h}_n^*\left(S_k-S_l\right)^*(S_k-S_l)\mathbf{h}_n} \\
&= \underset{f_{mi},g_{in}}{\mathrm{E}}\ e^{-(PT/16MR)\mathbf{f}^*[\sum_{n=1}^N \mathcal{G}_n^*\left(S_k-S_l\right)^*(S_k-S_l)\mathcal{G}_n]\mathbf{f}}.
\end{aligned}
\tag{20}
$$

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

$$
\begin{aligned}
&P(\underline{\mathbf{s}}_k \longrightarrow \underline{\mathbf{s}}_l) \\
&\lesssim \underset{g_{in}}{\mathrm{E}}\det^{-1}\left[I_{RM} + \frac{PT}{16MR}\sum_{n=1}^N \mathcal{G}_n^*(S_k-S_l)^*(S_k-S_l)\mathcal{G}_n\right].
\end{aligned}
\tag{21}
$$

Similar to the multiple-antenna case [4, 48] and the case of wireless relay networks with single-antenna nodes [5], to achieve full diversity, $S_k - S_l$ must be full rank. Since the distributed space-time codes S_k and S_l are $T \times MR$, in the following, we will assume $T \geq MR$ and the code is fully diverse.

Denote the minimum singular value of $(S_k - S_l)^*(S_k - S_l)$ by $\sigma_{\min}^2$. From the full diversity of the code, $\sigma_{\min}^2 > 0$. Therefore, the right side of (21) can be further upper bounded as

$$
\begin{aligned}
P(\underline{\mathbf{s}}_k \longrightarrow \underline{\mathbf{s}}_l) &\lesssim \underset{g_{in}}{\mathrm{E}}\det^{-1}\left[I_{RM} + \frac{PT\sigma_{\min}^2}{16MR}\sum_{n=1}^N \mathcal{G}_n^*\mathcal{G}_n\right] \\
&= \underset{g_{in}}{\mathrm{E}}\prod_{i=1}^R\left(1 + \frac{PT\sigma_{\min}^2}{16MR}\sum_{n=1}^N |g_{in}|^2\right)^{-M}.
\end{aligned}
\tag{22}
$$

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

$$
\begin{aligned}
&P(\underline{\mathbf{s}}_k \longrightarrow \underline{\mathbf{s}}_l) \\
&\lesssim \frac{1}{(N-1)!^R}\left[\int_0^\infty \left(1 + \frac{PT\sigma_{\min}^2}{16MR}x\right)^{-M} x^{N-1}e^{-x}dx\right]^R.
\end{aligned}
\tag{23}
$$

By defining $y = 1 + (PT\sigma_{\min}^2/16MR)x$, we have

$$
\begin{aligned}
&P(\underline{\mathbf{s}}_k \longrightarrow \underline{\mathbf{s}}_l) \\
&\lesssim \frac{1}{(N-1)!^R}\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{NR} e^{16MR^2/PT\sigma_{\min}^2} \\
&\quad \times \left[\int_1^\infty \frac{(y-1)^{N-1}}{y^M} e^{-(16MR/PT\sigma_{\min}^2)y}dy\right]^R
\end{aligned}
$$

$$
\begin{aligned}
&\lesssim \frac{1}{(N-1)!^R}\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{NR} \\
&\quad \times \left[\sum_{l=0}^{N-1}\binom{N-1}{l}\int_1^\infty y^{l-M} e^{-(16MR/PT\sigma_{\min}^2)y}dy\right]^R.
\end{aligned}
\tag{24}
$$

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 $\underline{\mathbf{s}}_k$ by $\underline{\mathbf{s}}_l$ has the following upper bound:*

$$
P(\underline{\mathbf{s}}_k \longrightarrow \underline{\mathbf{s}}_l) \lesssim \frac{1}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{min}^2}\right)^{\min\{M,N\}R}
$$

$$
\times \begin{cases}
\left(\dfrac{2^{N-1}}{M-N}\right)^R P^{-NR} & \text{if } M > N, \\[2ex]
\left(\dfrac{\log^{1/M}P}{P}\right)^{MR} & \text{if } M = N, \\[2ex]
(N-M-1)!^R P^{-MR} & \text{if } M < N.
\end{cases}
\tag{25}
$$

Therefore, the diversity of the wireless relay network is

$$
d = \begin{cases}
\min\{M,N\}R & \text{if } M \neq N, \\[1ex]
MR\left(1 - \dfrac{1}{M}\dfrac{\log\log P}{\log P}\right) & \text{if } M = N.
\end{cases}
\tag{26}
$$

Proof. See Appendix B. □

4.2. 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 not 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 $R(\log\log P/\log P)$, which is negligible when P is high.

If we can use the diversity definition in [49], since $\lim_{P\to\infty}(\log\log P/\log P) = 0$, diversity $\min\{M,N\}R$ can be obtained.

The results in Theorem 6 are obtained by considering only the highest-order term of P in the PEP formula. In brief, we call the rth highest-order term of P in the PEP formula

the rth term. When analyzing the diversity, not only is the first term important, but also how dominant it is. Therefore, we should analyze the contributions of the second and also other terms of P compared to those of the first one. This is equivalent to analyzing how large the total transmit power P should be for the terms in (25) to dominate. The following remarks are on this issue. They can be observed from the proof of Theorem 3 in Appendix B.

Remark 1. (1) If $|M - N| > 1$, from (B.13) and (B.22), the second term behaves as $P^{-\min\{M,N\}R+1}$. The difference between the first and second terms is a P factor. Therefore, the first term is dominant when $P \gg 1$. In other words, contributions of the second and other terms are negligible when $P \gg 1$.

(2) If $M = N$, from (B.16), the second term is

$$\frac{2^{M-1}R}{(M-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\frac{\log^{R-1}P}{P^{MR}}, \tag{27}$$

which has one less $\log P$ than the first one. Therefore, the first term, $(1/(M-1)!^R)(16MR/T\sigma_{\min}^2)^{MR}(\log^{1/M}P/P)^{MR}$, 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 and even other terms are not negligible.

(3) If $|M - N| = 1$, from (B.11) and (B.24), the second term behaves as $P^{-\min\{M,N\}R}(\log P/P)$. The difference between the first and second terms is $\log P/P$ factor. Therefore, the first term given in (25) 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 $P \gg 1$.

5. DIVERSITY ANALYSIS FOR THE GENERAL CASE

5.1. 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 diversity for networks with any number of relays is 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 (18), we can see that one way of upper bounding the PEP is to upper bound R_W. Since $R_W \geq 0$,

$$R_W \leq (\text{tr}\, R_W)I_N = \left(N + \frac{P_2}{P_1+1}\sum_{n=1}^{N}\sum_{i=1}^{R}|g_{in}|^2\right)I_N. \tag{28}$$

Therefore, from (18) and using the power allocation given in (19),

$$P(\underline{s}_k \longrightarrow \underline{s}_l)$$
$$\lesssim \mathop{\mathrm{E}}_{f_{mi},g_{in}} e^{-(PT/8MNR(1+(1/NR)\sum_{n=1}^{N}\sum_{i=1}^{R}|g_{in}|^2))\text{tr}\, H^*(S_k-S_l)^*(S_k-S_l)H}$$
$$\tag{29}$$

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

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \mathop{\mathrm{E}}_{g_{in}}\prod_{i=1}^{R}\left(1 + \frac{PT\sigma_{\min}^2}{8MNR}\frac{g_i}{1+(1/NR)\sum_{i=1}^{R}g_i}\right)^{-M}, \tag{30}$$

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 (diversity 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 $\underline{s}_k$ by $\underline{s}_l$ satisfies*

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{1}{(N-1)!^R}\left(\frac{8MNR}{T\sigma_{\min}^2}\right)^{\min\{M,N\}R}$$
$$\times \begin{cases} \left[\dfrac{M}{N(M-N)}\right]^R P^{-NR} & \text{if } M > N, \\[2ex] \left(1+\dfrac{1}{N}\right)^R\left(\dfrac{\log^{1/M}P}{P}\right)^{MR} & \text{if } M = N, \\[2ex] \left[\dfrac{1}{N}+(N-M-1)!\right]^R P^{-MR} & \text{if } M < N. \end{cases} \tag{31}$$

Therefore, the same diversity as in (26) *is obtained.*

Proof. See Appendix C. □

Although the same diversity is obtained as in the $R \to \infty$ case, there is a factor of N in (31), which does not appear in (25). This is because we upper bound R_W by $(\text{tr}\, R_W)I_N$, whose expectation is N times the expectation of R_W, while in the previous subsection we approximate R_W by its expectation. This factor of N can be avoided by tighter upper bounds of R_W. In the following subsection, we analyze the maximum eigenvalue of R_W. Then in Section 5.3, a PEP upper bound using the maximum eigenvalue of R_W is obtained.

5.2. The maximum eigenvalue of Wishart matrix

Denote the maximum eigenvalue of $(1/R)G^*G$ 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 1/2, $(1/R)G^*G$ is known as the Wishart matrix. While there exists explicit formula for the distribution of the minimum eigenvalue of a Wishart matrix, we could not find nonasymptotic formula for the maximum eigenvalue. Therefore, we calculate the PDF and CDF of $\lambda_{\max}$ from the joint distribution of all the eigenvalues of $(1/R)G^*G$ in this section. The following theorem has been proved.
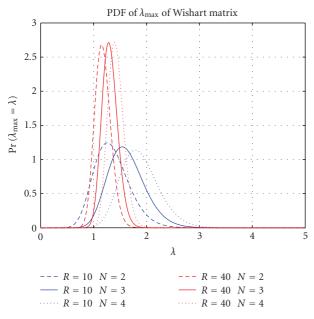
FIGURE 2: PDF of the maximum eigenvalue of $(1/R)G^*G$.

FIGURE 3: CDF of the maximum eigenvalue of $(1/R)G^*G$.

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

(1) The PDF of the maximum eigenvalue of $(1/R)G^*G$ is

$$p_{\lambda_{\max}}(\lambda) = \frac{R^{RN}\lambda^{R-N}e^{-R\lambda}}{\prod_{n=1}^{N}\Gamma(R-n+1)\Gamma(n)}\det F, \qquad (32)$$

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 $(1/R)G^*G$ is

$$P(\lambda_{\max} \leq \lambda) = \frac{R^{RN}}{\prod_{n=1}^{N}\Gamma(R-n+1)\Gamma(n)}\det F', \qquad (33)$$

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. See Appendix D. $\qquad\square$

A theoretical analysis of the PDF and CDF from (32) and (33) appears to be 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 becomes very close to 1 soon after. The larger R is, 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 $G^*G \approx RI_N$ in Section 4 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.
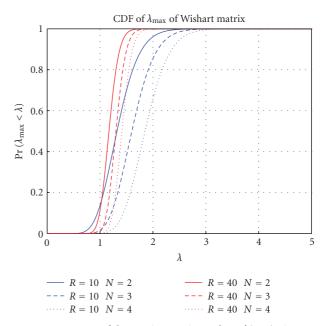
Corollary 1. *When $R \geq N$, the PDF of the maximum eigenvalue of $(1/R)G^*G$ can be upper bounded as*

$$p_{\lambda_{\max}}(\lambda) \leq C_1 \lambda^{RN-1}e^{-R\lambda}, \qquad (34)$$

where

$$C_1 = \frac{1}{\prod_{n=1}^{N}\Gamma(R-n+1)\Gamma(n)}$$
$$\times \frac{2^{N-1}R^{RN}}{\prod_{n=1}^{N-1}(R-N+2n-1)(R-N+2n)(R-N+2n+1)} \qquad (35)$$

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 (32), f_{nn} can be upper bounded as

$$f_{nn} \leq \int_0^\lambda (\lambda-t)^2 t^{R-N+2n-2}dt$$
$$= \frac{2}{(R-N+2n-1)(R-N+2n)(R-N+2n+1)}$$
$$\times \lambda^{R-N+2n+1}, \qquad (36)$$

then we have

$$\det F$$
$$\leq \frac{2^{N-1}}{\prod_{n=1}^{N-1}(R-N+2n-1)(R-N+2n)(R-N+2n+1)}$$
$$\times \lambda^{RN-R+N-1}. \qquad (37)$$

Thus, (34) is obtained. $\qquad\square$

5.3. *Bound on PEP from bound on eigenvalues*

If the maximum eigenvalue of $(1/R)G^*G$ is $\lambda_{\max}$, the maximum eigenvalue of R_W is $1+(P_2R/(P_1+1))\lambda_{\max}$, and therefore $R_W \leq (1 + (P_2R/(P_1 + 1))\lambda_{\max})I_N$. From (20) and using the power allocation given in (19), we have

$$
\begin{aligned}
&\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l \mid \lambda_{\max} = c) \\
&\qquad \leq \mathop{\mathrm{E}}_{f_{mr}, g_{rn}} e^{-(P_1 P_2 T/4M(1+P_1+P_2 R\lambda_{\max}))\mathrm{tr}(S_k-S_l)^*(S_k-S_l)HH^*} \\
&\qquad \lesssim \mathop{\mathrm{E}}_{f_{mr}, g_{rn}} e^{-(PT/8(1+\lambda_{\max})MR)\mathrm{tr}(S_k-S_l)^*(S_k-S_l)HH^*}.
\end{aligned}
\tag{38}
$$

The only difference of the above formula with formula (20) 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,

$$
\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l \mid \lambda_{\max} = c) \lesssim \frac{1}{(N-1)!^R} \left[\frac{8(1+c)MR}{T\sigma_{\min}^2} \right]^{\min\{M,N\}R}
$$
$$
\times \begin{cases} \left(\dfrac{2^{N-1}}{M-N} \right)^R P^{-NR} & \text{if } M > N, \\[2ex] \left(\dfrac{\log^{1/M} P}{P} \right)^{MR} & \text{if } M = N, \\[2ex] (N-M-1)!^R P^{-MR} & \text{if } M < N. \end{cases}
\tag{39}
$$

The following theorem can thus be obtained.

Theorem 6 (diversity 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 $\underline{s}_k$ by $\underline{s}_l$ can be upper bounded as*

$$
\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{\hat{C}}{(N-1)!^R} \left(\frac{8MR}{T\sigma_{\min}^2} \right)^{\min\{M,N\}R}
$$
$$
\times \begin{cases} \left(\dfrac{2^{N-1}}{M-N} \right)^R P^{-NR} & \text{if } M > N, \\[2ex] \left(\dfrac{\log^{1/M} P}{P} \right)^{MR} & \text{if } M = N, \\[2ex] (N-M-1)!^R P^{-MR} & \text{if } M < N, \end{cases}
\tag{40}
$$

where

$$
\hat{C} = \begin{cases} C_1 \displaystyle\sum_{i=0}^{\min\{M,N\}R} \binom{\min\{M,N\}R}{i} \dfrac{(RN+i-1)!}{R^{RN+i}} & \text{if } R \geq N, \\[3ex] C_2 \displaystyle\sum_{i=0}^{\min\{M,N\}R} \binom{\min\{M,N\}R}{i} \dfrac{(RN+i-1)!}{R^i N^{RN}} & \text{if } R < N, \end{cases}
$$

$$
C_2 = \frac{1}{\prod_{r=1}^{R} \Gamma(N - r + 1)\Gamma(r)}
$$
$$
\times \frac{2^{R-1}N^{RN}}{\prod_{r=1}^{R-1}(N - R + 2r - 1)(N - R + 2r)(N - R + 2r + 1)}.
\tag{41}
$$

Therefore, the same diversity as in (26) is obtained.

Proof. When $R \geq N$,

$$
\begin{aligned}
\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l) &= \int_0^\infty \mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l \mid \lambda_{\max} = c) \mathrm{p}_{\lambda_{\max}}(c)\,dc \\
&\leq \int_0^\infty C_1 c^{RN-1} e^{-Rc} \mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l \mid \lambda_{\max} = c)\,dc
\end{aligned}
\tag{42}
$$

using (34) in Corollary 1. From (39),

$$
\mathrm{P}(\underline{s}_i \longrightarrow \underline{s}_i) \lesssim \frac{C_1}{(N-1)!^R} \left(\frac{8MR}{T\sigma_{\min}^2} \right)^{\min\{M,N\}R}
$$
$$
\times \int_0^\infty c^{RN-1} e^{-Rc}(1 + c)^{\min\{M,N\}R}\,dc
$$
$$
\times \begin{cases} \left(\dfrac{2^{N-1}}{M-N} \right)^R P^{-NR} & \text{if } M > N, \\[2ex] \left(\dfrac{\log P}{P^M} \right)^R & \text{if } M = N, \\[2ex] (N-M-1)!^R P^{-MR} & \text{if } M < N. \end{cases}
\tag{43}
$$

Since

$$
\begin{aligned}
&\int_0^\infty c^{RN-1} e^{-Rc}(1 + c)^{\min\{M,N\}R}\,dc \\
&\qquad = \sum_{i=0}^{\min\{M,N\}R} \binom{\min\{M,N\}R}{i} \frac{(RN+i-1)!}{R^{RN+i}},
\end{aligned}
\tag{44}
$$

(40) is obtained.

For the case of $R < N$, G^* is an $N \times R$ ($N > R$) matrix whose entries are i.i.d. $\mathcal{CN}(0,1)$. Denote the maximal eigenvalue of $(1/N)GG^*$ as $\lambda'_{\max}$. Its PDF and CDF are given in Theorem 5 with R and N being switched. Using the facts that the maximal eigenvalue of $(1/R)G^*G$ is $(N/R)\lambda'_{\max}$ and

$$
\begin{aligned}
&\int_0^\infty c^{RN-1} e^{-Nc}\left(1 + \frac{N}{R}c\right)^{\min\{M,N\}R}\,dc \\
&\qquad = \sum_{i=0}^{\min\{M,N\}R} \binom{\min\{M,N\}R}{i} \frac{(RN+i-1)!}{R^i N^{RN}},
\end{aligned}
\tag{45}
$$

we can finish the proof of this theorem. $\qquad\square$

6. SIMULATION RESULTS

In this section, we show simulated block error rates of three networks with multiple transmit/receive antennas and compare them with the three PEP bounds we derived in (25), (31), and (40). These bounds are also addressed as PEP bound 1, PEP bound 2, and PEP bound 3 for the sake of presentation. The main purpose of this section is to verify the diversity results in (26). The optimal code design is not an issue. In the simulations, we use the power allocation in (19) and the ML decoding in (17). It is known that with ML metric, a factor of 1/2 can be applied to Chernoff bounds on the two-signal error rate, which is the block error rate when there are two possible transmit signals. Thus, the PEP bounds shown in Figures 4–6 are calculated from (25), (31), and (40) with a factor of 1/2. In all figures, the horizontal axis indicates P, the total transmit power used in the whole network.

Our first example, whose performance is shown in Figure 4, is a network with one transmit antenna, two relay antennas, and two receive antennas, that is, $M = 1$, $R = 2$, $N = 2$. We set $T = MR = 2$. The transmit signal is designed as

$$\underline{\mathbf{s}} = \begin{bmatrix} s_1 & s_2 \end{bmatrix}^t, \qquad (46)$$

where s_1 and s_2 are chosen as BPSK signals (normalized according to (1)). The matrices used at relays are designed as

$$A_1 = I_2, \quad A_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \qquad (47)$$

The distributed space-time codeword formed at the receiver S is thus a 2×2 real orthogonal design [50]. Then, we show performance of a network with $M = 2$, $R = 2$, $N = 1$ in Figure 5. We set $T = MR = 4$. The transmit signal is deigned as

$$\underline{\mathbf{s}} = \begin{bmatrix} s_1 & -s_2 \\ s_2 & s_1 \\ s_3 & -s_4 \\ s_4 & s_3 \end{bmatrix}, \qquad (48)$$

where s_1, s_2, s_3, s_4 are also BPSK signals (normalized according to (1)). The matrices used at relays are designed as

$$A_1 = I_4, \qquad A_2 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \qquad (49)$$

The distributed space-time codeword formed at the receiver S is thus a 4×4 real orthogonal design [50]. Finally, in Figure 6, we show performance of a network with $M = 2$, $R = 1$, $N = 2$. We set $T = MR = 2$. The transmit signal is
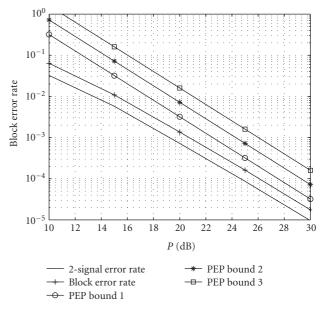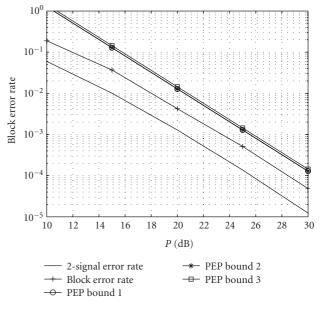

FIGURE 4: $M = 1$, $R = 2$, $N = 2$, $T = 2$.

FIGURE 5: $M = 2$, $R = 2$, $N = 1$, $T = 4$.

designed as

$$\underline{\mathbf{s}} = \begin{bmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{bmatrix}, \qquad (50)$$

where s_1 and s_2 are BPSK signals (normalized according to (1)). The matrices used at the relay are set to be I_2. The distributed space-time codeword formed at the receiver S is again a 2×2 real orthogonal design [50]. The transmission rate of all three networks can be calculated to be 1/2. For comparison, we also show the 2-signal error rates of the three networks by fixing $s_2, \ldots, s_T$.
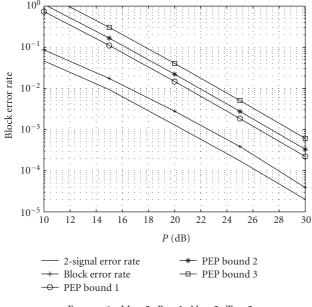
FIGURE 6: $M = 2$, $R = 1$, $N = 2$, $T = 2$.

Legend:
— 2-signal error rate
+ Block error rate
-⊖- PEP bound 1
-∗- PEP bound 2
-□- PEP bound 3

Figures 4–6 indicate that when the transmit power is high, all three networks achieve the diversities shown by the PEP bounds. This verifies our diversity result in (26). PEP bound 1 is the tightest of the three. This is because PEP bound 1 is obtained by approximating R_W by its asymptotic $(R \to \infty)$ limit, which is also its mean; however, strict lower bounds on R_W are used in the calculations of bound 2 and bound 3. In Figure 5, the three bounds are very close to each other and, actually, bounds 1 and 2 are the same.

7. CONCLUSIONS

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 not less than $M\mathcal{R}$, an achievable diversity is $\min\{M, N\}\mathcal{R}$ if $M \neq N$ and $M\mathcal{R}(1 - (1/M)(\log\log P/\log P))$ 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. Simulation results are exhibited to justify our diversity analysis.

APPENDICES

A. PROOF OF THEOREM 1

Proof. It is 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 tth row are $\sqrt{(P_1 P_2 T/(P_1 + 1)M)}[S_k]_t H$ and $\overline{R_W}$, respectively.

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

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

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 $\underline{s}_k$. With full channel information at the receiver,

$$\mathrm{E}x_{tn} = \sqrt{\frac{P_1 P_2 T}{M(P_1 + 1)}} \sum_{i=1}^{R} \sum_{m=1}^{M} \sum_{\tau=1}^{T} f_{mi} g_{in} a_{i,t\tau} s_{k,\tau m}. \tag{A.2}$$

Therefore, the mean of the tth row is then represented by $\sqrt{(P_1 P_2 T/M(P_1 + 1))}[S_k]_t H$. Since $\mathbf{v}_i$, $\mathbf{w}_n$, and $\underline{s}_k$ are independent,

$$\mathrm{Cov}(x_{t_1 n_1}, x_{t_2 n_2})$$

$$= \mathrm{E}(x_{t_1 n_1} - \mathrm{E}x_{t_1 n_1})\overline{(x_{t_2 n_2} - \mathrm{E}x_{t_2 n_2})}$$

$$= \frac{P_2}{P_1 + 1} \sum_{i_1=1}^{R} \sum_{\tau_1=1}^{T} \sum_{i_2=1}^{R} \sum_{\tau_2=1}^{T}$$

$$\times \mathrm{E}g_{i_1 n_1} a_{i_1,t_1 \tau_1} v_{r_1 \tau_1} \overline{g}_{i_2 n_2} \overline{a}_{i_2,t_2 \tau_2} \overline{v}_{i_2 \tau_2} + \mathrm{E}w_{t_1 n_1} \overline{w}_{t_2 n_2}$$

$$= \frac{P_2}{P_1 + 1} \sum_{i=1}^{R} \sum_{\tau=1}^{T} a_{i,t_1 \tau} \overline{a}_{i,t_2 \tau} g_{in_1} \overline{g}_{in_2} + \delta_{n_1 n_2} \delta_{t_1 t_2}$$

$$= \delta_{t_1 t_2} \left(\frac{P_2}{P_1 + 1} \sum_{r=1}^{R} g_{in_1} \overline{g}_{in_2} + \delta_{n_1 n_2} \right)$$

$$= \delta_{t_1 t_2} \left(\frac{P_2}{P_1 + 1} \begin{bmatrix} g_{1n_1} & \cdots & g_{Rn_1} \end{bmatrix} \begin{pmatrix} \overline{g}_{1n_2} \\ \vdots \\ \overline{g}_{Rn_2} \end{pmatrix} + \delta_{n_1 n_2} \right). \tag{A.3}$$

The fourth equality is true since A_i are unitary. Therefore, the rows of X are independent since the covariance of $x_{t_1 n_1}$ and $x_{t_2 n_2}$ is zero when $t_1 \neq t_2$. It is also easy to see that the variance matrix of each row is $I_N + (P_2/(P_1 + 1))G^t \overline{G}$, which equals $\overline{R_W}$. Therefore,

$$\mathrm{P}([X]_t \mid \underline{s}_k)$$

$$= \left(\pi^N \det \overline{R_W} \right)^{-T}$$

$$\times e^{-\mathrm{tr}\overline{[X - \sqrt{(P_1 P_2 T/M(P_1+1))} S_k H]}_t \overline{R_W^{-1}} [X - \sqrt{(P_1 P_2 T/M(P_1+1))} S_k H]_t^t}$$

$$= \left(\pi^N \det R_W \right)^{-T}$$

$$\times e^{-\mathrm{tr}[X - \sqrt{(P_1 P_2 T/M(P_1+1))} S_k H]_t R_W^{-1} [X - \sqrt{(P_1 P_2 T/M(P_1+1))} S_k H]_t^*}. \tag{A.4}$$

Since $\mathrm{P}(X \mid \underline{s}_k) = \prod_{t=1}^{T} \mathrm{P}([X]_t \mid \underline{s}_k)$, (16) can be obtained. □

B. PROOF OF THEOREM 3

Proof. Define

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

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

$$\int_u^\infty x^n e^{-\mu x} dx$$
$$= e^{-u\mu} \sum_{k=0}^n \frac{n!}{k!} \frac{u^k}{\mu^{n-k+1}}, \quad u > 0, \ \Re\mu > 0, \ n = 0, 1, 2, \dots, \tag{B.2}$$

$$\int_u^\infty \frac{e^{-\mu x}}{x^{n+1}} dx = (-1)^{n+1} \frac{\mu^n \mathbf{Ei}(-\mu u)}{n!}$$
$$+ \frac{e^{-\mu u}}{u^n} \sum_{k=0}^{n-1} \frac{(-1)^k \mu^k u^k}{n \cdots (n-k)}, \quad \mu > 0, \ n = 1, 2, \dots, \tag{B.3}$$

$$\int_u^\infty \frac{e^{-\mu x}}{x} dx = -\mathbf{Ei}(-\mu u), \quad \Re\mu > 0, \ u \ge 0, \tag{B.4}$$

where

$$\mathbf{Ei}(\chi) = \int_{-\infty}^\chi \frac{e^t}{t} dt, \quad \chi < 0, \tag{B.5}$$

is the exponential integral function [51]. To calculate I, we discuss the following cases separately.
Case 1 ($M < N$). In this case,

$$I = \sum_{l=M}^{N-1} \binom{N-1}{l} \int_1^\infty y^{l-M} e^{-(16MR/PT\sigma_{\min}^2)y} dy$$
$$+ \binom{N-1}{M-1} \int_1^\infty \frac{e^{-(16MR/PT\sigma_{\min}^2)y}}{y} dy \tag{B.6}$$
$$+ \sum_{l=0}^{M-2} \binom{N-1}{l} \int_1^\infty y^{-(M-l)} e^{-(16MR/PT\sigma_{\min}^2)y} dy.$$

Using equalities (B.2)–(B.4) with $u = 1$, $\mu = (16MR/PT\sigma_{\min}^2)$, and $n = l - M$ or $n = M - l - 1$,

$$I = \sum_{l=M}^{N-1} \binom{N-1}{l} (l-M)! \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(l-M+1)}$$
$$+ \binom{N-1}{M-1} \log P + \sum_{l=0}^{M-2} \binom{N-1}{l} \frac{1}{M-l-1} \tag{B.7}$$
$$+ \text{lower-order terms of } P.$$

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

$$I = (N-M-1)! \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M)} + o\left(P^{-(N-M)}\right). \tag{B.8}$$

Therefore,

$$P(\underline{s}_k \longrightarrow \underline{s}_l)$$
$$\lesssim \frac{1}{(N-1)!^R} \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{NR}$$
$$\times \left[(N-M-1)! \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M)} + o\left(\frac{1}{P^{MR}}\right)\right]^R$$
$$= \left[\frac{(N-M-1)!}{(N-1)!}\right]^R \left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR} \frac{1}{P^{MR}} + o\left(\frac{1}{P^{MR}}\right). \tag{B.9}$$

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 $N = M + 1$,

$$I = \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-1} + M\left[-\mathbf{Ei}\left(-\frac{16MR}{PT\sigma_{\min}^2}\right)\right] + O(1)$$
$$= \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-1} + M\log P + O(1). \tag{B.10}$$

Therefore,

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{1}{M!^R} \left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR} \frac{1}{P^{MR}}$$
$$+ \frac{RM}{M!^R} \left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR+1} \frac{\log P}{P^{MR+1}} + o\left(\frac{\log P}{P^{MR+1}}\right). \tag{B.11}$$

The second highest-order term of P in the PEP behaves as $\log P/P^{MR+1} = P^{-(MR+1-\log\log P/\log P)}$.
If $N > M + 1$,

$$I = (N-M-1)! \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M)} + (N-1)(N-M-2)!$$
$$\times \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M-1)} + o\left(P^{N-M-1}\right)$$
$$= \left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{-(N-M)} \left[(N-M-1)! + (N-1)\right.$$
$$\left. \times (N-M-2)! \frac{16MR}{PT\sigma_{\min}^2} + o\left(\frac{1}{P}\right)\right]. \tag{B.12}$$

Therefore,

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{(N-M-1)!^R}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR}\frac{1}{P^{MR}}$$
$$+ \frac{(N-1)(N-M-2)(N-M-1)!^{R-1}}{(N-1)!^R}$$
$$\times \left(\frac{16MR}{T\sigma_{\min}^2}\right)^{MR+1}\frac{1}{P^{MR+1}} + o\left(\frac{1}{P^{MR+1}}\right). \tag{B.13}$$

Case 2 $(M = N)$. In this case,

$$I = \int_1^\infty \frac{e^{-\left(16MR/PT\sigma_{\min}^2\right)y}}{y}dy$$
$$+ \sum_{l=0}^{N-2}\binom{N-1}{l}\int_1^\infty y^{-(M-l)}e^{-(16MR/PT\sigma_{\min}^2)y}dy. \tag{B.14}$$

Using (B.4) with $\mu = 16MR/PT\sigma_{\min}^2$ and $u = 1$, and (B.3) with $u = 1$ and $n = M-l-1$, we have

$$I = \log P + \sum_{l=0}^{N-2}\binom{N-1}{l}\frac{1}{M-l-1}$$
$$+ \text{lower-order terms of } P \tag{B.15}$$
$$< \log P + 2^{N-1} + \text{lower-order terms of } P.$$

Therefore,

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

Also, the second highest-order term of P in the PEP behaves as $\log^{R-1}P/P^{RM}$ and the next term has one $\log P$ less and so on.
Case 3 $(M > N)$. In this case,

$$I = \sum_{l=0}^{M-2}\binom{N-1}{l}\int_1^\infty y^{-(M-l)}e^{-(16MR/PT\sigma_{\min}^2)y}dy. \tag{B.17}$$

Using (B.3) with $u = 1, \mu = 16MR/PT\sigma_{\min}^2$, and $n = M-l-1$,

$$I = \sum_{l=0}^{N-1}\binom{N-1}{l}\frac{1}{M-l-1} + \text{lower-order terms of } P. \tag{B.18}$$

Thus,

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{1}{(N-1)^R}\left(\frac{16MR}{PT\sigma_{\min}^2}\right)^{NR}$$
$$\times \left[\sum_{l=0}^{N-1}\binom{N-1}{l}\frac{1}{M-l-1} + o(1)\right]^R$$
$$= \left[\frac{1}{(N-1)!}\sum_{l=0}^{N-1}\binom{N-1}{l}\frac{1}{M-l-1}\right]^R$$
$$\times \left(\frac{16MR}{T\sigma_{\min}^2}\right)^{NR}P^{-NR} + o(P^{-NR}). \tag{B.19}$$

We can further upper bound the PEP to get a simpler formula. Notice that $1/(M-l-1) \le 1/(M-N)$. Thus,

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \left[\frac{1}{(M-N)(N-1)!}\sum_{l=0}^{N-1}\binom{N-1}{l}\right]^R$$
$$\times \left(\frac{16MR}{T\sigma_{\min}^2}\right)^{NR}P^{-NR}$$
$$\le \left[\frac{2^{N-1}}{(M-N)(N-1)!}\right]^R\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{NR}P^{-NR}. \tag{B.20}$$

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 (B.3),

$$I < \frac{2^{N-1}}{M-N} - \frac{2^{N-1}}{(M-N)(M-N-1)}\frac{16MR}{PT\sigma_{\min}^2} + o\left(\frac{1}{P}\right). \tag{B.21}$$

Therefore,

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \left[\frac{2^{N-1}}{(M-N)(N-1)!}\right]^R\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{NR}$$
$$\times \left[\frac{1}{P^{NR}} + \frac{R}{M-N-1}\left(\frac{16MR}{T\sigma_{\min}^2}\right)\frac{1}{P^{NR+1}}\right]$$
$$+ o\left(\frac{1}{P^{NR+1}}\right). \tag{B.22}$$

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

$$I < 2^{N-1} + \frac{16MR}{T\sigma_{\min}^2}\frac{\log P}{P} + O\left(\frac{1}{P}\right). \tag{B.23}$$

Therefore,

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{2^{R(N-1)}}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{NR}\frac{1}{P^{NR}}$$
$$+ \frac{2^{(R-1)(N-1)}R}{(N-1)!^R}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{NR+1}\frac{\log P}{P^{NR+1}}$$
$$+ o\left(\frac{\log P}{P^{NR+1}}\right), \tag{B.24}$$

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

C.   PROOF OF THEOREM 4

Proof. Since g_i have PDF $p(g_i) = (1/(N-1)!)g_i^{N-1}e^{-g_i}$,

$$\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l) \le \sum_{r=0}^{R} \sum_{1\le i_1<\cdots<i_r\le R} T_{i_1,\dots,i_r}, \qquad \text{(C.1)}$$

where

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

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

$$T_{1,\dots,r} = \frac{1}{(N-1)!^R}\underbrace{\int_x^\infty\cdots\int_x^\infty}_{r}\underbrace{\int_0^x\cdots\int_0^x}_{R-r}\prod_{i=1}^{R}$$
$$\times \left(1+\frac{PT\sigma_{\min}^2}{8MNR}\frac{g_i}{1+(1/NR)\sum_{i=1}^{R}g_i}\right)^{-M}$$
$$\times g_i^{N-1}e^{-g_i}dg_1\cdots dg_R$$
$$< \frac{1}{(N-1)!^R}\int_x^\infty\cdots\int_x^\infty\prod_{i=1}^{r}$$
$$\times \left(\frac{PT\sigma_{\min}^2}{8MNR}\frac{g_i}{1+((R-r)/NR)x+(1/NR)\sum_{i=1}^{r}g_i}\right)^{-M}$$
$$\times g_i^{N-1}e^{-g_i}dg_1\cdots dg_r$$
$$\times \int_0^x\cdots\int_0^x\prod_{i=r+1}^{R}g_i^{N-1}e^{-g_i}dg_{r+1}\cdots dg_R$$
$$= \frac{1}{(N-1)!^R}\left(\frac{PT\sigma_{\min}^2}{8MNR}\right)^{-rM}\gamma^{R-r}(N,x)$$
$$\times \int_x^\infty\cdots\int_x^\infty\left(1+\frac{R-r}{NR}x+\frac{1}{NR}\sum_{i=1}^{r}g_i\right)^{rM}$$
$$\times \prod_{i=1}^{r}\frac{e^{-g_i}}{g_i^{M-N+1}}dg_1\cdots dg_r, \qquad\qquad \text{(C.3)}$$

where $\gamma(n,x)$ is the incomplete gamma function [51]. 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 does not affect the diversity. Here, to have the PEP result consistent with formula (25) in Section 6, we set $\beta = (T\sigma_{\min}^2/8MNR)^\alpha$. 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

$$\frac{1}{(N-1)!^R}\gamma^R(N,P^\alpha) = 1+o(1). \qquad \text{(C.4)}$$

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, $(1/(N-1)!^R)\gamma^R(N,1)$, is a constant. Therefore, α should be negative. Thus,

$$\gamma(N,x) = \frac{1}{N}x^N + o(x^N) = \frac{1}{N}\beta^N P^{\alpha N} + o(P^{\alpha N}) \qquad \text{(C.5)}$$

We are only interested in the highest-order term of P. When P is large, $((R-r)/NR)x$ is negligible compared with 1. Therefore,

$$T_{1,\dots,r} \lesssim \frac{1}{(N-1)!^R N^{R-r}}\left(\frac{T\sigma_{\min}^2}{8MNR}\right)^{-rM+\alpha N(R-r)}$$
$$\times P^{-rM+\alpha N(R-r)}\Lambda, \qquad\qquad \text{(C.6)}$$

where we have defined

$$\Lambda = \int_x^\infty\cdots\int_x^\infty\left(1+\frac{1}{NR}\sum_{i=1}^{r}g_i\right)^{rM}\prod_{i=1}^{r}\frac{e^{-g_i}}{g_i^{M-N+1}}dg_1\cdots dg_r. \qquad \text{(C.7)}$$

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

$$\left(1+\frac{1}{NR}\sum_{i=1}^{r}g_i\right)^a = \sum_{j=0}^{a}\left(\sum_{1\le l_1<\cdots<l_j\le k}\sum_{\substack{i_1,\dots,i_j\ge 1 \\ \sum i_m\le a}}C(i_1,\dots,i_j)\right.$$
$$\left.\times \frac{1}{(NR)^{i_1+\cdots+i_j}}g_{l_1}^{i_1}g_{l_2}^{i_2}\cdots g_{l_j}^{i_j}\right), \qquad \text{(C.8)}$$

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

$$C(i_1,\dots,i_j) = \binom{k}{i_1}\binom{k-i_1}{i_2}\cdots\binom{k-i_1-\cdots-i_{j-1}}{i_j} \qquad \text{(C.9)}$$

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,

$$\Lambda = \sum_{j=0}^{r}\sum_{1\le l_1<\cdots<l_j\le r}\sum_{\substack{i_1,\dots,i_j\ge 1 \\ \sum i_m\le r}}$$
$$\times C(i_1,\dots,i_j)\Lambda(j;l_1,\dots,l_j;i_1,\dots,i_j), \qquad \text{(C.10)}$$

where

$$\Lambda(j; l_1, \ldots, l_j; i_1, \ldots, i_j)$$

$$= \frac{1}{(NR)^{i_1 + \cdots + i_j}} \left(\prod_{m=1}^{j} \int_x^\infty \frac{e^{-g_{l_m}}}{g_{l_m}^{M-N+1-i_m}} dg_{l_m} \right) \quad \text{(C.11)}$$

$$\times \prod_{i \neq i_1, \ldots, i_j} \int_x^\infty \frac{e^{-g_i}}{g_i^{M-N+1}} dg_i.$$

From (B.2)–(B.4), while $P \to \infty$, $\alpha < 0$, and $n > 0$,

$$\int_x^\infty \lambda^n e^{-\lambda} d\lambda = n! + o(1),$$

$$\int_x^\infty \frac{e^{-\lambda}}{\lambda} d\lambda = (-\alpha) \log P + o(\log P), \quad \text{(C.12)}$$

$$\int_x^\infty \frac{e^{-\lambda}}{\lambda^{n+1}} d\lambda = \frac{1}{n} \beta^{-n} P^{-\alpha n} + o\left(P^{-\alpha n}\right).$$

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

$$\Lambda = \prod_{i=1}^{r} \int_x^\infty \frac{e^{-g_i}}{g_i^{M-N+1}} dg_i$$

$$\approx \begin{cases} \frac{1}{(M-N)^r} \beta^{-r(M-N)} P^{-r\alpha(M-N)} & \text{if } M > N, \\ (-\alpha)^r \log^r P & \text{if } M = N, \\ (N-M-1)!^r & \text{if } M < N. \end{cases} \quad \text{(C.13)}$$

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

$$\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l)$$

$$\leq \sum_{r=0}^{R} \binom{R}{r} T_{1, \ldots, r}$$

$$\lesssim \frac{1}{(N-1)!^R} \sum_{r=0}^{R} \binom{R}{r} \frac{1}{N^{R-r}}$$

$$\times \left(\frac{T\sigma_{\min}^2}{8MNR} \right)^{-rM + \alpha N(R-r)} P^{-rM + \alpha N(R-r)}$$

$$\times \begin{cases} \frac{1}{(M-N)^r} \left(\frac{T\sigma_{\min}^2}{8MNR} \right)^{-r\alpha(M-N)} P^{-r\alpha(M-N)} & \text{if } M > N, \\ (-\alpha)^r \log^r P & \text{if } M = N, \\ (N-M-1)!^r & \text{if } M < N. \end{cases} \quad \text{(C.14)}$$

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 rth term as $f(r)$, choose an $\alpha < 0$ 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 $\alpha \leq -1$, $f(r)$ 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 $\alpha \geq -1$, $f(r)$ is a 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 $\alpha = -1$, and

$$\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{(1/N + 1/(M-N))^R}{(N-1)!^R} \left(\frac{8MNR}{T\sigma_{\min}^2} \right)^{NR} P^{-NR}$$

$$= \left[\frac{M/N}{(M-N)(N-1)!} \right]^R \left(\frac{8MNR}{T\sigma_{\min}^2} \right)^{NR} P^{-NR}. \quad \text{(C.15)}$$

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

$$\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{(1/N + (N-M-1)!)^R}{(N-1)!^R} \left(\frac{8MNR}{T\sigma_{\min}^2} \right)^{-MR} P^{-MR}. \quad \text{(C.16)}$$

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

$$\mathrm{P}(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \frac{1}{(N-1)!^R} \left[\left(\frac{8MNR}{T\sigma_{\min}^2} \right)^{\log \log P / \log P} \frac{1}{N} \right. $$

$$\left. + \left(1 - \frac{1}{N} \frac{\log \log P}{\log P} \right) \right]^R$$

$$\times \left(\frac{8MNR}{T\sigma_{\min}^2} \right)^{-MR} \left(\frac{\log^{1/M} P}{P} \right)^{-MR}$$

$$\approx \frac{(1/N + 1)^R}{(N-1)!^R} \left(\frac{8MNR}{T\sigma_{\min}^2} \right)^{-MR} \left(\frac{\log^{1/M} P}{P} \right)^{-MR}. \quad \text{(C.17)}$$

$$\square$$

D.   PROOF OF THEOREM 5

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

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

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

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

$$V_h(\Lambda) = \begin{bmatrix} h_0(\lambda_1) & \cdots & h_0(\lambda_N) \\ \vdots & \ddots & \vdots \\ h_{N-1}(\lambda_1) & \cdots & h_{N-1}(\lambda_N) \end{bmatrix}, \quad \text{(D.1)}$$

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

Define G' as a complex Gaussian matrix whose entries' real and imaginary parts have mean zero and variance one. Denote the ordered eigenvalues 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 [52]:

$$P(\lambda'_1,\ldots,\lambda'_N) = C\prod_{i=1}^{N}\lambda_i'^{R-N}e^{-\lambda'_i/2}\prod_{1\leq i<j\leq N}(\lambda'_i - \lambda'_j)^2, \quad \text{(D.2)}$$

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

$$P(\lambda_1,\ldots,\lambda_N)$$

$$= P(\lambda'_1,\ldots,\lambda'_N)\frac{d\lambda'_1\cdots d\lambda'_N}{d\lambda_1\cdots d\lambda_N}$$

$$= \det[\text{diag}\{2R,\ldots,2R\}](2R\lambda_1,\ldots,2R\lambda_N)$$

$$= (2R)^N C\prod_{i=1}^{N}(2R\lambda_i)^{R-N}e^{-R\lambda_i}\prod_{1\leq i<j\leq N}[2R(\lambda_i-\lambda_j)]^2$$

$$= C(2R)^{RN}\prod_{i=1}^{N}\lambda_i^{R-N}e^{-R\lambda_i}\prod_{1\leq i<j\leq N}(\lambda_i-\lambda_j)^2. \quad \text{(D.3)}$$

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

$$P(\lambda_{\max} = \lambda)$$

$$= P(\lambda_1 = \lambda)$$

$$= \int_{\lambda\geq\lambda_2\geq\cdots\lambda_N}P(\lambda,\lambda_2,\ldots,\lambda_N)d\lambda_2\cdots d\lambda_N$$

$$= \frac{1}{(N-1)!}\int_0^\lambda\cdots\int_0^\lambda P(\lambda,\lambda_2,\ldots,\lambda_N)d\lambda_2\cdots d\lambda_N$$

$$= \frac{C(2R)^{RN}}{(N-1)!}\lambda^{R-N}e^{-R\lambda}\int_0^\lambda\cdots\int_0^\lambda\prod_{i=2}^{N}(\lambda-\lambda_i)^2\lambda_i^{R-N}e^{-R\lambda_i}$$

$$\times\prod_{2\leq i<j\leq N}(\lambda_i-\lambda_j)^2 d\lambda_2\cdots d\lambda_N$$

$$= \frac{C(2R)^{RN}}{(N-1)!}\lambda^{R-N}e^{-R\lambda}\int_0^\lambda\cdots\int_0^\lambda\prod_{i=2}^{N}f(\lambda_i)$$

$$\times\det V_g(\lambda_2,\ldots,\lambda_N)\det V_h(\lambda_2,\ldots,\lambda_N)\lambda_1\cdots\lambda_N$$

$$= \frac{C(2R)^{RN}}{(N-1)!}\lambda^{R-N}e^{-R\lambda}(N-1)!\det F, \quad \text{(D.4)}$$

where in the second equality we have changed the integral space from ordered λ_i to unordered one. From the symmetry

of λ_i, we only need to divide the new value by $(N-1)!$. From Theorem 7,

$$F = \int_0^\lambda g(t)\begin{bmatrix}1\\t\\\vdots\\t^{N-2}\end{bmatrix}\begin{bmatrix}1 & t & \cdots & t^{N-2}\end{bmatrix}dt, \quad \text{(D.5)}$$

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. □

E. DISCUSSION ON HETEROGENEOUS NETWORKS

In Section 2.2, it is assumed that f_{mi} and g_{in} have the same variance. Physically, this means that the distances between the transmitter/receiver and all relays are about the same, which may not be a practical assumption for networks with scattered nodes. In this appendix, we extend our diversity analysis to heterogeneous networks whose channels have different variances. We assume that the distributions of f_{mi} and g_{in} are $\mathcal{CN}(0,\sigma_{f_{mi}}^2)$ and $\mathcal{CN}(0,\sigma_{g_{in}}^2)$, respectively. By following the derivation in Section 4, compared with (21), the PEP for the heterogeneous case can be upper bounded by

$$P(\underline{s}_k \longrightarrow \underline{s}_l)$$

$$\lesssim \underset{g_{in}}{\mathrm{E}}\det^{-1}\left[\Sigma_{\mathbf{f}} + \frac{PT}{16MR}\sum_{n=1}^{N}\mathscr{g}_n^*(S_k-S_l)^*(S_k-S_l)\mathscr{g}_n\right], \quad \text{(E.1)}$$

where $\Sigma_{\mathbf{f}} = \text{diag}\{\sigma_{f11}^2,\ldots,\sigma_{fm1}^2,\ldots,\sigma_{1R}^2,\ldots,\sigma_{fmR}^2\}$ is the covariance matrix of $\mathbf{f}$. Denote $\sigma_{\mathbf{f}_i}^2 = \min_{m=1}^{M}\{\sigma_{f_{mi}}^2\}$ and $\sigma_{\mathbf{g}_i}^2 = \min_{m=1}^{M}\{\sigma_{g_{in}}^2\}$. We have from (E.1) that

$$P(\underline{s}_k \longrightarrow \underline{s}_l) \lesssim \underset{g_{in}}{\mathrm{E}}\prod_{i=1}^{R}\left(\sigma_{\mathbf{f}_i}^2 + \frac{\sigma_{\mathbf{g}_i}^2 PT\sigma_{\min}^2}{16MR}\sum_{n=1}^{N}\frac{|g_{in}|^2}{\sigma_{g_{in}}^2}\right)^{-M}. \quad \text{(E.2)}$$

Since $|g_{in}|^2/\sigma_{g_{in}}^2$ has the exponential distribution with mean 1 and g_{in}'s are independent, $\sum_{n=1}^{N}|g_{in}|^2/\sigma_{g_{in}}^2$ has the i.i.d. gamma distribution $(1/(N-1)!)g_i^{N-1}e^{-g_i}$. Thus, following the derivations in Section 4 and Appendix B, we can show that the PEP of heterogeneous networks has the following upper bound:

$$P(\underline{s}_k \longrightarrow \underline{s}_l)$$

$$\lesssim \frac{1}{(N-1)!^R}\left(\prod_{i=1}^{R}\sigma_{\mathbf{f}_i}^2\right)^{-\min\{M,N\}}\left(\frac{16MR}{T\sigma_{\min}^2}\right)^{\min\{M,N\}R}$$

$$\times\begin{cases}\left(\prod_{i=1}^{R}\sigma_{\mathbf{g}_i}^2\right)^{-(N-M)}\left(\frac{2^{N-1}}{M-N}\right)^R P^{-NR} & \text{if } M>N,\\[12pt]\left(\frac{\log^{1/M}P}{P}\right)^{MR} & \text{if } M=N,\\[12pt](N-M-1)!^R P^{-MR} & \text{if } M<N.\end{cases} \quad \text{(E.3)}$$

Thus, the same diversity results as in (26) can be obtained. Similarly, the rigorous analysis in Section 5 also applies to this heterogeneous case.

ACKNOWLEDGMENTS

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[2] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, 1999.

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

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[5] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3524–3536, 2006.

[6] Y. Chang and Y. Hua, "Application of space-time linear block codes to parallel wireless relays in mobile ad hoc networks," in *Proceedings of the 36th Asilomar Conference on Signals, Systems and Computers (ACSSC '02)*, vol. 1, pp. 1002–1006, Pacific Grove, Calif, USA, November 2003.

[7] Y. Hua, Y. Mei, and Y. Chang, "Wireless antennas-making wireless communications perform like wireline communications," in *Proceedings of IEEE Topical Conference on Wireless Communication Technology*, Honolulu, Hawaii, USA, October 2003.

[8] Y. Tang and M. C. Valenti, "Coded transmit macrodiversity: block space-time codes over distributed antennas," in *Proceedings of IEEE Vehicular Technology Conference (VTC '01)*, vol. 2, pp. 1435–1438, Atlantic City, NJ, USA, May 2001.

[9] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.

[10] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part II: implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, 2003.

[11] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1099–1109, 2004.

[12] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1433–1444, 2006.

[13] A. F. Dana and B. Hassibi, "On the power efficiency of sensory and ad-hoc wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 2890–2914, 2006.

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

[15] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 362–371, 2006.

[16] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4152–4172, 2005.

[17] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[18] M. Katz and S. Shamai, "Transmitting to colocated users in wireless ad hoc and sensor networks," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3540–3563, 2005.

[19] T. E. Hunter, S. Sanayei, and A. Nosratinia, "Outage analysis of coded cooperation," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 375–391, 2006.

[20] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, 2006.

[21] Q. Zhao and H. Li, "Performance of differential modulation with wireless relays in rayleigh fading channels," *IEEE Communications Letters*, vol. 9, no. 4, pp. 343–345, 2005.

[22] S. Yiu, R. Schober, and L. Lampe, "Differential distributed space-time block coding," in *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM '05)*, pp. 53–56, Vancouver, BC, Canada, August 2005.

[23] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: the relay case," in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '02)*, vol. 3, pp. 1577–1586, New York, NY, USA, June 2002.

[24] A. Ribeiro, X. Cai, and G. B. Giannakis, "Symbol error probabilities for general cooperative links," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1264–1273, 2005.

[25] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1804–1824, 2002.

[26] Y. Jing and H. Jafarkhani, "Using orthogonal and quasi-orthogonal designs in wireless relay networks," *IEEE Transactions on Information Theory*, vol. 53, pp. 4106–4118, November 2007.

[27] F. Oggier and B. Hassibi, "An algebraic family of distributed space-time codes for wireless relay networks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, Seattle, Wash, USA, July 2006.

[28] F. Oggier and B. Hassibi, "An algebraic coding scheme for wireless relay networks with multipleantenna nodes," to appear in *IEEE Transactions on Signal Processing*.

[29] T. Kiran and B. S. Rajan, "Distributed space-time codes with reduced decoding complexity," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, Seattle, Wash, USA, July 2006.

[30] G. S. Rajan and B. S. Rajan, "Algebraic distributed space-time codes with low ML decoding complexity," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, Nice, France, June 2007.

[31] G. S. Rajan, A. Tandon, and B. S. Rajan, "On four-group ML decodable distributed space time codes for cooperative communication," submitted to *IEEE Transactions on Information Theory*, http://arxiv.org/abs/cs.IT/0701067v1.

[32] Y. Jing and H. Jafarkhani, "Distributed differential space-time coding in wireless relay networks," to appear in *IEEE Transactions on Communications*.

[33] F. Oggier and B. Hassibi, "Cyclic distributed space-time codes for wireless networks with no channel information," Submitted.

[34] T. Kiran and B. S. Rajan, "Partially-coherent distributed space-time codes with differential encoder and decoder," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 426–433, 2007.

[35] G. S. Rajan and B. S. Rajan, "Noncoherent low-decoding-complexity space-time codes for wireless relay networks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, Nice, France, June 2007.

[36] P. V. Kumar, K. Vinodh, M. Anand, and P. Elia, "Diversity-multiplexing gain tradeoff and DMT-optimal distributed space-time codes for certain cooperative communication protocols: overview and recent results," in *Proceedings of Information Theory and Applications Workshop*, San Diego, Calif, USA, January 2007.

[37] P. Elia and P. V. Kumar, "Approximately universal schemes for cooperative diversity in wireless networks," submitted to *IEEE Transactions on Information Theory*.

[38] X. Guo and X.-G Xia, "A distributed space-time coding in asynchronous wireless relay networks," to appear in *IEEE Transactions on Wireless Communications*.

[39] Y. Li, W. Zhang, and X.-G Xia, "Distributive high-rate space-frequency codes achieving full cooperative and multipath diversities for asynchronous cooperative communications," to appear in *IEEE Transactions on Vehicular Technology*.

[40] Z. Li and X.-G Xia, "A simple Alamouti space-time transmission scheme for asynchronous cooperative systems," *IEEE Signal Processing Letters*, vol. 14, pp. 804–807, November 2007.

[41] Y. Li and X.-G. Xia, "A family of distributed space-time trellis codes with asynchronous cooperative diversity," *IEEE Transactions on Communications*, vol. 55, no. 4, pp. 790–800, 2007.

[42] P. Elia, S. Kittipiyakul, and T. Javidi, "Cooperative diversity schemes for asynchronous wireless networks," *Wireless Personal Communications*, vol. 43, no. 1, pp. 3–12, 2007.

[43] P. Elia and P. V. Kumar, "Constructions of cooperative diversity schemes for asynchronous wireless networks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 2724–2728, Seattle, Wash, USA, July 2006.

[44] G. S. Rajan and B. S. Rajan, "Distributed space-time codes for cooperative networks with partial CSI," submitted to *IEEE Transactions on Information Theory*, http://arxiv.org/abs/cs/0701068v1.

[45] G. S. Rajan and B. S. Rajan, "A non-orthogonal distributed space-time coded protocol part I: signal model and design criteria," submitted to *IEEE Transactions on Information Theory*, http://arxiv.org/abs/cs/0610161v1.

[46] G. S. Rajan and B. S. Rajan, "A non-orthogonal distributed space-time coded protocol part II: code construction and DM-G tradeoff," submitted to *IEEE Transactions on Information Theory*, http://arxiv.org/abs/cs/0610160.

[47] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[48] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.

[49] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.

[50] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[51] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, New York, NY, USA, 6th edition, 2000.

[52] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*, Ph.D. thesis, Massachusetts Institute of Technology (MIT), Department of Mathematics, Cambridge, Mass, USA, 1989.

[53] Y. Jing and B. Hassibi, "Cooperative diversity in wireless relay networks with multiple-antenna nodes," in *Proceedings of the International Symposium on Information Theory (ISIT '05)*, pp. 815–819, Adelaide, South Australia, September 2005.