Abstract

This paper investigates the benefits of Amplify-and-Forward (AF) relaying in the setup of multi-antenna wireless networks. Reference [1] introduced the idea of Random Sequential (RS) relaying and showed that it achieves the maximum diversity gain in a general multi-antenna network. Here, we show that random unitary matrix multiplication at the relay nodes empowers the RS scheme to achieve a better Diversity-Multiplexing Tradeoff (DMT) as compared to the traditional AF relaying. First, we study the case of a multi-antenna full-duplex single-relay two-hop network, for which we show that the RS achieves the optimum DMT. Applying this result, we derive a new achievable DMT for the case of multi-antenna half-duplex parallel relay network. Interestingly, it turns out that the DMT of the RS scheme is optimum for the case of multi-antenna two parallel non-interfering half-duplex relays. Next, we show that random unitary matrix multiplication also improves the DMT of the Non-Orthogonal AF relaying scheme of [2] in the case of a multi-antenna single relay channel. Finally, we study the general case of multi-antenna full-duplex relay networks and derive a new lower-bound on its DMT using the RS scheme.

I. INTRODUCTION

Recently, cooperative schemes have been proposed as candidates to exploit the spatial diversity offered by the relay networks (see for example [2], [5]–[9] and references therein). Decode-and-Forward (DF), Amplify-and-Forward (AF) and Compress-and-Forward (CF) are the primary relaying strategies discussed in the literature. While DF and CF strategies are popular for small-scale networks to achieve or approach capacity values (see for example [10]–[12] and references therein), the AF relaying turns out to be more suitable in realizing cooperative transmission strategies (see for example [1], [2], [5], [8], [13] and references therein).
In AF relaying, the relays do not decode the transmitted messages. Instead, they forward their received signals to the destination after a proper amplification. Hence, the relays require less computing power and the end-to-end system experiences a smaller delay in comparison with the other relaying strategies. Accordingly, AF relaying schemes are suitable for deployment in practical wireless systems. Moreover, in contrast to DF relaying, the performance of AF relaying is not limited by the source-to-relay channel quality. Also, unlike CF relaying, parallel relays can be used in conjunction with the AF strategy to realize the power boosting offered by coherently combining the received signals at the destination [9]. Indeed, it is shown that AF relaying is asymptotically (large number of relays) optimal (capacity achieving) for both single-antenna [14] and multi-antenna [9], [15] parallel relay networks.

While AF relaying is investigated well in single-antenna networks, much less is known about its application in multi-antenna networks. Indeed, unlike the single-antenna scenario, in this case the AF multipliers are matrices rather than scalars. Hence, finding the optimum AF matrices becomes challenging.

A fundamental measure to evaluate the performance of cooperative schemes is the DMT, which was first introduced by Zheng and Tse in the context of point-to-point multi-antenna fading channels [16]. Roughly speaking, the diversity-multiplexing tradeoff identifies the optimal compromise between the transmission reliability and the data rate in the high Signal to Noise Ratio (SNR) regime.

The Non-orthogonal Amplify-and-Forward (NAF) scheme, first proposed by Nabar et al. in [17], has been further studied by Azarian et al. in [2] for the single-antenna multi-relay setup. In addition to analyzing the DMT of the NAF scheme, reference [2] shows that NAF is the best in the class of AF strategies for single-antenna single-relay systems. However, the NAF scheme falls far from the upper-bound in multi-relay systems.

Yang and Belfiore in [13] study the DMT performance of the NAF scheme for the multi-antenna parallel relay setup. Moreover, based on the non-vanishing determinant criterion, the authors constructed a family of space-time codes for the NAF scheme over multi-antenna channels. However, as shown in [13], the NAF scheme falls far from the DMT upper-bound in the multiple-antenna setup, particularly for small values of multiplexing gain. Indeed, even for the case of a multi-antenna two-hop single-relay setup, the NAF scheme is unable to achieve the maximum diversity gain of the system.

Recently, [1] has studied the achievable DMT of the AF relaying for a general multi-relay network. For this purpose, they propose a new AF relaying scheme which is called RS. They show that the RS scheme achieves the optimum diversity gain in a general multi-antenna multi-relay network. However, they established the DMT optimality results only for the single-antenna multi-relay network. Also, the authors in [18] independently obtained the same DMT optimality result using a different approach.

Yuksel et al. in [6] apply CF strategy and show that CF achieves the DMT upper-bound for multi-antenna half-duplex single-relay networks. However, in their proposed scheme, the relay node needs to know the Channel State Information (CSI) of all the channels in the network, which may not be practical. More recently, Avestimehr et al. in [12] show that a variant of the CF relaying achieves the capacity of any general single-antenna Gaussian relay network within a fixed number of bits, which only depends on the number of nodes in the network. Furthermore, the authors in [19] show that the result of [12] is
still valid for both multi-antenna Gaussian relay networks with or without fast (ergodic) Rayleigh fading.

In this paper, we investigate the benefits of AF relaying in multi-antenna multi-relay networks. For this purpose, we study the application of the RS scheme proposed in [1]. The key elements of the proposed scheme are: 1) signal transmission through sequential paths in the network, 2) path timing such that no non-causal interference is caused from the transmitted signal of the future paths on the received signal of the current path, 3) multiplication by a random unitary matrix at each relay node, and 4) no signal boosting in AF relaying at the relay nodes, i.e. the received signal is amplified by a coefficient with the absolute value of at most 1. We derive the DMT of the RS scheme for multi-antenna multi-relay networks. To accomplish this task, we first study a simple structure, namely the multi-antenna full-duplex two-hop single-relay network. We show that unlike the traditional AF relaying, the RS scheme achieves the optimum DMT. This fact can be justified as follows: using the traditional AF relaying, there exists a chance that the eigenvectors corresponding to the largest eigenvalues of the incoming channel matrix of the relay project to the eigenvectors corresponding to the small eigenvalues of the relay’s outgoing channel matrix. This event degrades the performance of traditional AF relaying in the multi-antenna setup. However, in the RS scheme, due to the random independent unitary matrix multiplication at the relay nodes for different time-slots, such an event is much less likely to happen. This fact will be elaborated throughout the paper.

Next, we study the case of multi-antenna half-duplex parallel relay network and, by deriving its DMT, we show that the RS scheme improves the DMT of the traditional AF relaying scheme. Interestingly, it turns out that the DMT of the RS scheme is optimum for the multi-antenna half-duplex parallel two-relay ($K = 2$) setup with no direct link between the relays. We also show that utilizing random unitary matrix multiplication improves the DMT of the NAF relaying scheme of [2] in the case of a multi-antenna single relay channel.

Finally, we study the class of general full-duplex multi-antenna relay networks whose underlying graph is directed acyclic and all nodes are equipped with the same number of antennas. Using the RS scheme, we derive a new lower-bound for the achievable DMT of this class of networks. It turns out that the new DMT lower-bound meets the optimum DMT at the corner points, corresponding to the maximum multiplexing gain and the maximum diversity gain of the network, respectively. Another point worth mentioning is that the RS scheme is robust in the sense that it achieves all points of the DMT curve with no modification of the underlying parameters. In other words, the relay nodes of the network perform the same operation, no matter at which point of the DMT curve the scheme is operating.

The rest of the paper is organized as follows. System model is introduced in section II. Section III is dedicated to Diversity-Multiplexing Tradeoff analysis of the RS scheme in the multi-antenna setup. This section is further divided into four subsections as follows. Subsection III-A studies the multi-antenna single-relay two-hop network and also the multi-antenna multi-hop relay network with one relay in each hop. Subsection III-B is dedicated to the multi-antenna half-duplex parallel relay network. The well-known multi-antenna single-relay channel (with direct link between the source and the destination) is investigated in subsection III-C. Subsection III-D studies the achievable DMT of the RS scheme for the general multi-antenna full-duplex relay networks whose underlying graph is directed acyclic. Finally, section IV concludes the paper.
A. Notations

Throughout the paper, the superscripts $^T$ and $^H$ stand for matrix operations of transposition and conjugate transposition, respectively. Capital bold letters represent matrices, while lowercase bold letters and regular letters represent vectors and scalars, respectively. $\|v\|$ denotes the norm of vector $v$, while $\|A\|$ represents the Frobenius norm of matrix $A$. $|A|$ denotes the determinant of matrix $A$ and $\text{Tr}\{A\}$ denotes the trace of matrix $A$. $\log(.)$ denotes the base-2 logarithm. The notation $A \preceq B$ is equivalent to $B - A$ being a positive semi-definite matrix. Motivated by the definition in [16], we define the notation $f(P) = g(P)$ as $\lim_{P \to \infty} \frac{f(P)}{\log(P)} = \lim_{P \to \infty} \frac{g(P)}{\log(P)}$. Similarly, $f(P) \preceq g(P)$ and $f(P) \succeq g(P)$ are equivalent to $\lim_{P \to \infty} \frac{f(P)}{\log(P)} \leq \lim_{P \to \infty} \frac{g(P)}{\log(P)}$ and $\lim_{P \to \infty} \frac{f(P)}{\log(P)} \geq \lim_{P \to \infty} \frac{g(P)}{\log(P)}$, respectively. Finally, we use $A \approx B$ to denote the approximate equality between $A$ and $B$, such that by substituting $A$ with $B$ the validity of the equations are not compromised.

II. System Model

In this paper, we study the DMT for various setups of multiple-antenna relay networks. The common assumptions for all setups are:

- The network consists of a single source, one or multiple relays and a single destination
- All nodes in the network are assumed be equipped with multiple antennas. However, the number of antennas at each node are denoted differently depending on the scenario being discussed. In the setup of Theorem 1 in subsection III-A, which studies the DMT of a two-hop single-relay network, the number of antennas at the source, relay and destination are denoted by $m, p, \text{and } n$, respectively. However, in Theorem 2 of this section which considers the setup of multi-hop relay network, the number of antennas at nodes are denoted by $N_0, N_1, \cdots, N_h$ where $h$ denotes the total number of hops.
- In subsection III-B, in which the setup of a parallel relay network is considered, the number of antennas at the source and the destination are assumed to be $m$ and $n$. In Theorem 3, $K$ parallel relays with the same number (denoted by $p$) of antennas are considered while Theorem 4 studies the setup of two parallel relays with different number of antennas denoted by $n_1$ and $n_2$ while assuming $n = m$. The general setup of a single-relay network (with direct link between the source and the destination) is studies in subsection III-C (Theorem 5) and the number of antennas at the source, relay and destination are denoted by $m, p, \text{and } n$, respectively. Finally, the general setup of a multi-antenna full-duplex relay network with directed acyclic underlying graph is considered in subsection III-D (Theorem 6) in which the number of antennas at node $i$ is denoted by $N_i$.
- All channels in the network are assumed to be circularly symmetric zero-mean unit-variance Gaussian with gains that remain unchanged during the entire transmission period (quasi-static Rayleigh fading). For each link, the receiver side is assumed to have perfect knowledge about the corresponding channel and the transmitter side is assumed to have no knowledge about the channel.
- The noise vector at each node (relay or destination) is assumed to be additive white with circularly symmetric zero-mean unit-variance Gaussian distribution.
- The relays are assumed to work in full-duplex mode in sub-sections III-A and III-D and half-duplex mode in sub-sections
The source and the relay(s) have power constraints $\mathbb{E}\{x_t^H x_t\} \leq P$ and $\mathbb{E}\{x_r^H x_r\} \leq P$, respectively, where the average is taken with respect to $x_t$ for the source and with respect to $x_r$ and $y_r$ (the received signal at the relay) for the relays. As the main focus of this paper is the analysis of DMT, the source and the relay(s) are assumed to operate in the high SNR regime ($P \to \infty$).

III. DIVERSITY-MULTIPLEXING TRADEOFF

A. Two-Hop Single Relay Network

This setup corresponds to the network consisting of a source, a single full-duplex relay and a destination with no direct link between the source and the destination. The source, relay, and destination are equipped with $m$, $p$, and $n$ antennas, respectively (see Figure 1). The channel between the source and the relay is denoted by $H$ and the channel between the relay and the receiver is denoted by $G$. First, we study the achievable DMT by the traditional AF scheme in Lemmas 1 and 2 and show the achievable DMT in general does not match with the optimal value, which is achievable by the DF scheme. Then, in Theorem 1 we prove that using the proposed RS scheme, which is a modification of the traditional AF scheme, the optimal DMT is indeed achievable. Theorem 2 generalizes the result of Theorem 1 to the case of multi-hop relay network and shows that the proposed RS scheme still achieves the optimum DMT provided that a certain relationship between the number of antennas at nodes is satisfied.

$$y = \alpha G H x_t + \alpha G n_r + n_d,$$

where $x_t$ denotes the transmitted signal from the source and $n_r \sim \mathcal{CN}(0, I_p)$ and $n_d \sim \mathcal{CN}(0, I_n)$ denote the noise vectors at the relay and at the destination, respectively.

$^2$Note that as the channels are assumed to be fixed during the whole transmission period, the expectation are not taken with respect to any of the channel matrices.

$^3$In fact, this configuration is a special case of the degraded relay channel studied by [11]. In [11], the authors show that the DF scheme achieves the capacity of the degraded relay channel.
Lemma 1 The DMT of the system given in (1) is upper-bounded by the DMT of the following system:

\[ y = \alpha \mathbf{G} \mathbf{H} \mathbf{x}_t + \mathbf{n}_d, \]  

and is lower-bounded by the DMT of the following system:

\[ y = \alpha \mathbf{G} \mathbf{H} \mathbf{x}_t + \sqrt{c \log(P) + 1} \mathbf{n}_d, \]  

for some constant \( c \).

**Proof:** See Appendix I. ■

Lemma 2 The DMT of the systems in (2) and (3) are equal.

**Proof:** See Appendix II. ■

A direct conclusion of the Lemmas 1 and 2 is that the DMT of the two-hop network can be expressed as the DMT of the product channel \( \mathbf{G} \mathbf{H} \) which is computed in [13]. Due to the result given in Proposition 1 in [13], assuming \( m, n \geq p \), the DMT of the product channel \( \mathbf{A} = \mathbf{G} \mathbf{H} \) is a piecewise-linear function connecting the points \((r, d_{\mathbf{A}}(r))\), \( r = 0, 1, \ldots, p \), where

\[ d_{\mathbf{A}}(r) = (p - r)(q - r) - \frac{1}{2} \left\lfloor \frac{(p - \Delta - r)^+}{2} \right\rfloor, \]  

\( q = \min(m, n) \) and \( \Delta = |m - n| \). On the other hand, the piecewise-linear function connecting the integer points \((r, (p-r)(q-r))\) can be easily derived as the upper-bound by considering each of the source-relay or the relay-destination cuts. Comparing (4) with the upper-bound, it follows that the traditional AF scheme achieves the optimum DMT only when \( r \geq p - \Delta \). This motivates us to use a variant of AF scheme which achieves the optimum DMT in all cases. In fact, using the traditional AF scheme, there are three sources of outage: (i) the outage in the source-relay link, (ii) the outage in the relay-destination link, and (iii) the projection of the eigenmodes of \( \mathbf{H} \) over the eigenmodes of \( \mathbf{G} \) is very small. More precisely, the matrix \( \mathbf{V}^H(\mathbf{G})\mathbf{U}(\mathbf{H}) \), in which \( \mathbf{V}^H(\mathbf{G}) \) denotes the right eigenvector matrix from the Singular Value Decomposition (SVD) of \( \mathbf{G} \) and \( \mathbf{U}(\mathbf{H}) \) denotes the left eigenvector matrix from the SVD of \( \mathbf{H} \), has very small eigenvalues. The extra term \( \frac{1}{2} \left\lfloor \frac{(p - \Delta - r)^+}{2} \right\rfloor \) in (4) is due to the third source of outage. The first two outage events depend on the distribution of the eigenvalues of \( \mathbf{H} \) and \( \mathbf{G} \), while the third event depends solely on the direction of the eigenvectors of these two matrices. This suggests that in order to eliminate the extra terms in \( \frac{1}{2} \left\lfloor \frac{(p - \Delta - r)^+}{2} \right\rfloor \) one can multiply the received signal at the relay by \( \alpha \mathbf{\Theta} \), for some \( p \times p \) unitary matrix \( \mathbf{\Theta} \) (for preserving the power constraint at the relay). However, it should be noted that if \( \mathbf{\Theta} \) does not change across the transmission block, the performance of the systems does not change. Therefore, we propose that in each transmission slot an independent unitary matrix \( \mathbf{\Theta}_t \) is used and at the destination side the decoding is performed across \( L \) transmission slots.\(^4\) This is exactly what is being done in the proposed RS scheme in [1]. Indeed, this setup is a simple example of the general setup of the relay network studied in [1] in which the source and the destination are connected through a single path. In this case the proposed RS scheme reduces to the following: the source’s message is sent using \( L \) slots through the same path; at the relay

\(^4\)From practical point of view, the transmission slots are assumed to be long enough to make the probability of error solely dominated by the outage event. This fact is more elaborated in [16].
side the received signal is multiplied by a randomly independent (through different slots) unitary matrix, and subsequently, it is multiplied by a scalar $\alpha \leq 1$ such that the power constraint is satisfied, and the result is transmitted in the next slot. At the destination, following receiving the signal of the slots $2, 3, \ldots, L+1$, the source message is decoded. In the following theorem, we show that as long as $L$ is above a certain threshold, the probability of the third outage event is negligible compared to the first two outage events and hence, the optimum DMT is achievable by the RS scheme.

**Theorem 1** Consider a two-hop network consisting of a source with $m$ antennas and a destination with $n$ antennas that are connected through a full-duplex relay with $p$ antennas. Let us define $q = \min(m, n)$. Providing $L$ is large enough such that $L \geq \min\{2(p, q) \max(p, q)\}$, the RS scheme achieves the optimum DMT, which is the piecewise-linear function connecting the points $(k, (p-k)(q-k))$, $k = 0, 1, \ldots, \min(p, q)$.

**Proof:** Using Lemmas 1 and 2 the DMT of the system using the proposed RS scheme is equal to the DMT of the following system:

$$Y = \alpha \Omega X_t + N_d,$$

where $X_t \triangleq [x_t(1), \cdots, x_t(L)]^T$, $Y = [y(1), \cdots, y(L)]^T$, and $N_d \triangleq [n_d(1), \cdots, n_d(L)]^T$, in which $x_t(l)$, $y(l)$ and $n_d(l)$ denote the transmitted signal, received signal and noise in the $l$th slot, respectively, and

$$\Omega \triangleq \begin{bmatrix}
A_1 & 0 & \cdots & 0 \\
0 & A_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & A_L
\end{bmatrix},$$

in which $A_l \triangleq G \Theta_l H$. Hence, the matrix of the end-to-end channel is a block diagonal matrix consisting of $A_l$’s. Assuming that the transmitted signals in each slot are independent of each other, the mutual information between the input and the output of (5) can be written as

$$I(X_t; Y) = \sum_{l=1}^{L} \log \left| I + \alpha^2 \frac{P}{m} A_l A_l^H \right|,$$

in which it is assumed that $x_t(l) \sim \mathcal{CN}(0, \frac{P}{m} I_m)$, $\forall l = 1, \cdots, k$. Using the above equation, the probability of outage can be written as

$$P\{O\} = \mathbb{P} \left\{ \sum_{l=1}^{L} \log \left| I + \alpha^2 \frac{P}{m} A_l A_l^H \right| < Lr \log(P) \right\},$$

or equivalently,

$$P\{O\} = \mathbb{P} \left\{ \sum_{l=1}^{L} \sum_{j=1}^{\min(p, q)} \log \left( 1 + \alpha^2 \frac{P}{m} \lambda_j (A_l) \right) < Lr \log(P) \right\},$$

$^5$Note that, as can be observed from the proof of Lemma 1 in Appendix I, the constraint $\alpha \leq 1$ does not affect the DMT of the system.
where $\lambda_i(A)$ denotes the $i$th ordered eigenvalue of $A^H A$ ($\lambda_1 > \lambda_2 > \cdots > \lambda_{\min}$). Defining $\gamma_j(B) \triangleq \frac{-\log(\lambda_j(B))}{\log(P)}$ and $\delta \triangleq -\frac{\log(\alpha^2)}{\log(P)}$, we have

$$
P\{O\} = \mathbb{P}\left\{ \sum_{l=1}^{L} \sum_{j=1}^{\min(p,q)} \log \left( 1 + \frac{1}{m} p^{1-\delta - \gamma_j(A_l)} \right) < Lr \log(P) \right\}
$$

$$
\leq \mathbb{P}\left\{ \sum_{l=1}^{L} \sum_{j=1}^{\min(p,q)} \left( 1 - \delta - \gamma_j(A_l) \right)^+ < Lr \right\}.
$$

(10)

First, we show that $\alpha = 1$ (or $\delta = 0$), with probability one. For this purpose, we write $\alpha$ as follows:

$$
\alpha^2 = \min \left( 1, \frac{P}{\mathbb{E}_{x_t,n_r} \{ \|Hx_t + n_r\|^2 \}} \right)
$$

$$
= \min \left( 1, \frac{P}{\text{Tr} \{ HQ_xH^H + I \}} \right)
$$

$$
= \min \left( 1, \frac{P}{\text{Tr} \{ \frac{P}{m} HH^H + I \}} \right)
$$

$$
= \min \left( 1, \frac{P}{\frac{P}{m} \|H\|^2 + m} \right).
$$

(11)

From the above equation, we have

$$
P\{\delta \geq \delta_0\} < P\{\|H\|^2 > P^{\delta_0 - \epsilon}\},
$$

for all $\delta_0, \epsilon > 0$. Noting that $\|H\|^2$ has Chi-square distribution with $2mp$ degrees of freedom, it follows that

$$
P\{\|H\|^2 > P^{\delta_0 - \epsilon}\} \sim \frac{P^{mp(\delta_0 - \epsilon)}}{(mp)!} \exp \left\{ -P^{\delta_0 - \epsilon} \right\}.
$$

(13)

Choosing $\epsilon = \frac{\delta_0}{2}$, the above equation implies that for $\delta_0 > 0$, the probability $P\{\delta \geq \delta_0\}$ approaches to zero much faster than polynomially. More precisely, defining the event $\mathcal{F} \equiv \{ \delta > 0 \}$, we have $P\{\mathcal{F}\} = o(P^{-c})$ for any positive constant $c$. Since $P\{O\} \geq P^{-(q-r)(p-r)}$ (lower-bound on the outage probability corresponding to the DMT upper-bound), we can write

$$
P\{O\} = P\{O|\mathcal{F}\}P\{\mathcal{F}\} + P\{O|\mathcal{F}^c\}P\{\mathcal{F}^c\}
$$

$$
\sim (a) P\{O|\mathcal{F}^c\},
$$

(14)

where $(a)$ follows from the fact that $P\{\mathcal{F}\} = o(P\{O\})$. In other words, one can replace $\delta$ with zero in (10), which results in

$$
P\{O\} \equiv P\left\{ \sum_{l=1}^{L} \sum_{j=1}^{\min(p,q)} \left( 1 - \gamma_j(A_l) \right)^+ < Lr \right\}.
$$

(15)

Moreover, we have

$$
\lambda_i(A_l) \leq \|A_l\|^2 \leq (a) \|G\|^2\|H\|^2,
$$

(16)

Note that due to the definition of $\alpha$, we always have $\delta \geq 0$. 
where \((a)\) results from the fact that \(\|AB\|^2 \leq \|A\|^2\|B\|^2\), for any two matrices \(A\) and \(B\). Consider a negative number \(\varepsilon\). From the above equation, it follows that

\[
\mathbb{P}\{\gamma_i(A_t) \leq \varepsilon\} \leq \mathbb{P}\{\|G\|^2\|H\|^2 \geq P^{-\varepsilon}\}
\]

\[
\leq \mathbb{P}\{\|G\|^2 \geq P^{-\frac{\varepsilon}{2}}\} + \mathbb{P}\{\|H\|^2 \geq P^{-\frac{\varepsilon}{2}}\}
\]

\[
\geq \frac{P^{-\frac{\varepsilon}{2}}}{(np)!} \exp\{-P^{-\frac{\varepsilon}{2}}\} + \frac{P^{-\frac{\varepsilon}{2}}}{(mp)!} \exp\{-P^{-\frac{\varepsilon}{2}}\}
\]

\[
= o\left(P^{-(q-r)(p-r)}\right). \quad (17)
\]

As a result, following (14), we can assume that \(\gamma_j(A_t) \geq 0, \forall j = 1, \ldots, \min(p, q)\), in (15).

In order to compute the outage probability in (15), we need to find the statistical behavior of \(\gamma_j(A_t)\). Since we are interested in upper-bounding the outage probability of the RS scheme, finding an upper-bound for \(\gamma_j(A_t)\), or equivalently, a lower-bound for \(\lambda_j(A_t)\) would be sufficient. This is performed in the following lemma.

**Lemma 3** Consider matrices \(G\) and \(H\) with the size of \(m \times p\) and \(p \times n\), respectively, and a \(p \times p\) matrix \(\Theta\). Assume \(G\) and \(H\) are singular value decomposed as \(G = U(G)A_{\frac{1}{2}}(G)V^H(G)\) and \(H = U(H)A_{\frac{1}{2}}(H)V^H(H)\), respectively. We have

\[
\lambda_i(G\Theta H) \geq \lambda_i(G)\lambda_i(H)\lambda_{\min}\left(V^H_{(1,i)}(G)\Theta U_{(1,i)}(H)\right), \quad (18)
\]

where \(\lambda_i(A)\) and \(\lambda_{\min}(A)\) denote the \(i\)'th largest eigenvalue and the minimum eigenvalue of \(A^H A\), respectively, and \(A_{(a,b)}\) denotes the submatrix of \(A\) consisting of the \(a, a+1, \ldots, b\)'th columns of \(A\).

**Proof:** See Appendix [II] \(\blacksquare\)

The above lemma relates \(\lambda_i(A_t)\) to \(\lambda_i(G)\) and \(\lambda_i(H)\), which facilitates the subsequent derivations. A direct consequence of the above lemmas is that

\[
\gamma_i(A_t) \leq \gamma_i(G) + \gamma_i(H) + \gamma_{\min}(\Psi_{i,t}), \quad (19)
\]

where \(\Psi_{i,t} = V^H_{(1,i)}(G)\Theta U_{(1,i)}(H)\). As the statistical behaviors of \(\gamma_i(G)\) and \(\gamma_i(H)\) are known from [16], it is sufficient to derive the asymptotic behavior of \(\gamma_{\min}(\Psi_{i,t})\), or equivalently, \(\lambda_{\min}(\Psi_{i,t})\), which is performed in the following lemma:

**Lemma 4** Assuming small enough \(\varepsilon\), we have

\[
\mathbb{P}\{\lambda_{\min}(\Psi_{i,t}) \leq \varepsilon\} \leq \eta \sqrt{\varepsilon}, \quad (20)
\]

for some constant \(\eta\).

**Proof:** See Appendix [IV] \(\blacksquare\)

A direct consequence of the above lemma is that

\[
\mathbb{P}\{\gamma_{\min}(\Psi_{i,t}) > \theta\} \leq P^{-\frac{\theta}{2}}. \quad (21)
\]
Defining the $L \times 1$ vector $\psi \triangleq [\psi(1), \ldots, \psi(L)]^T$ as $\psi(l) \triangleq \max_i \gamma_{\min}(\Psi_{i,l})$, we have

$$\mathbb{P}\{\psi \geq \psi_0\} \overset{(a)}{=} \prod_{l=1}^{L} \mathbb{P}\{\psi(l) \geq \psi_0(l)\}$$

$$= \prod_{l=1}^{L} \mathbb{P}\left\{\min(p,q) \sum_{i=1}^{\min(p,q)} (\gamma_{\min}(\Psi_{i,l}) \geq \psi_0(l))\right\}$$

$$\overset{(b)}{\leq} P^{-\frac{1}{\min(p,q)}} \tag{22}$$

As $\Theta_{i}$'s are independent isotropic unitary matrices, their products with any possibly correlated set of unitary matrices constructs a set of independent isotropic unitary matrices [20]. Accordingly, $\Psi_{i,l}$'s are independent for different values of $l$, which results in (a). Also, (b) follows from Lemma 4 and the union bound inequality.

Let us define the $1 \times \min(p, q)$ vectors $\chi(H) \triangleq [\gamma_{\min}(p,m)(H), \gamma_{\min}(p,m)-1(H), \ldots, \gamma_{1+\min(p,m)-\min(p,q)}(H)]$ and $\chi(G) \triangleq [\gamma_{\min}(p,n)(G), \gamma_{\min}(p,n)-1(G), \ldots, \gamma_{1+\min(p,n)-\min(p,q)}(G)]$, respectively. Notice that these vectors include the log-values of the corresponding $\min(p, q)$ smallest eigenvalues of $HH^H$ and $GG^H$, respectively. Now, applying the result of Lemma 3 to (15), we can upper-bound the outage probability of the end-to-end channel as

$$\mathbb{P}\{O\} \leq \mathbb{P}\left\{\sum_{l=1}^{L} \sum_{i=1}^{\min(p,q)} (1 - \gamma_i(G) - \gamma_i(H) - \gamma_{\min}(\Psi_{i,l}))^+ < Lr\right\}$$

$$\leq \mathbb{P}\left\{\sum_{l=1}^{L} \sum_{i=1}^{\min(p,q)} (1 - \gamma_i(G) - \gamma_i(H) - \psi(l))^+ < Lr\right\}$$

$$\overset{(a)}{\leq} \mathbb{P}\left\{\sum_{l=1}^{L} \sum_{i=1}^{\min(p,q)} \left(1 - \gamma_{1+\min(p,n)-i}(G) - \gamma_{1+\min(p,m)-i}(H) - \psi(l)\right)^+ < Lr\right\}$$

$$= \mathbb{P}\left\{\sum_{l=1}^{L} \sum_{i=1}^{\min(p,q)} (1 - \chi_i(G) - \chi_i(H) - \psi(l))^+ < Lr\right\} \tag{23}$$

where (a) follows from the fact that the log-values ($\gamma_i$'s) corresponding to the smallest eigenvalues of $HH^H$ and $GG^H$ are greater than the log-values corresponding to the largest eigenvalues of these matrices. According to (22), to upper-bound the outage probability, it is sufficient to upper-bound the probability of the region of $(\psi, \chi(H), \chi(G))$ that satisfies (23). The following lemma gives a general formula for computing such an upper-bound:

**Lemma 5** Consider a fixed region $R \subseteq [0, \infty)^n$. Assume that a uniformly continuous non-negative function $f(x)$ ($f(x) \geq 0$) is defined over $[0, \infty)^n$ such that for all $x \in [0, \infty)^n$ we have $\mathbb{P}\{y \geq x\} \leq P^{-f(x)}$. Then, we have

$$\mathbb{P}\{x \in R\} \leq P^{-\inf_{x \in R} f(x)} \tag{24}$$

**Proof:** See Appendix \[\n\]

\[A uniformly continuous function f : M \rightarrow N where M \subseteq \mathbb{R}^m, N \subseteq \mathbb{R}^n is a function that has the following property: for every \(\epsilon\), there exists a constant \(g(\epsilon) > 0\) such that for all \(x, y \in M, \|x - y\| \leq g(\epsilon)\), we have \(\|f(x) - f(y)\| \leq \epsilon\).\]
According to the upper-bound in \(22\) and the distribution of \(\chi(G), \chi(H)\) derived in [16], we have
\[
\mathbb{P}\left\{ \psi \geq \hat{\psi}, \chi(G) \geq \chi', \chi(H) \geq \chi'' \right\} \lesssim P^{-\min_{p,q}} \sum_{l=1}^{L} \psi(l) - \sum_{i=1}^{\min(p,q)} a_i \chi'' + b_i \chi',
\]
\[
\overset{(a)}{\leq} P^{-\min_{p,q}} \sum_{l=1}^{L} \psi(l) - \sum_{i=1}^{\min(p,q)} (2i-1+|p-q|)(\chi'_i + \chi''),
\]
where \(a_i \triangleq 2|i + \min(p,m) - \min(p,q)| - 1 + |p-m|\) and \(b_i \triangleq 2|i + \min(p,n) - \min(p,q)| - 1 + |p-n|\) and (a) follows from the fact that
\[
a_i = 2[i + \min(p,m) - \min(p,q)] - 1 + |p-m|
= 2i - 1 + m + p - 2 \min(p,q)
\geq 2i - 1 + q + p - 2 \min(p,q)
= 2i - 1 + |p-q|,
\]
and similarly, \(b_i \geq 2i - 1 + |p-q|\). Now, we can apply the result of Lemma 5 to the region defined in (23) and the upper-bound derived in (25). Accordingly, we have
\[
\mathbb{P}\{O\} \lesssim P^{-\min_{p,q}} \sum_{l=1}^{L} \psi(l) + \sum_{i=1}^{\min(p,q)} (2i-1+|p-q|)(\chi_i(G) + \chi_i(H)),
\]
where the region \(\mathcal{R}\) is defined as
\[
\mathcal{R} \triangleq \left\{ (\chi(G), \chi(H), \psi) \mid \psi \geq 0, \chi_1(G) \geq \cdots \geq \chi_{\min(p,q)}(G) \geq 0, \chi_1(H) \geq \cdots \geq \chi_{\min(p,q)}(H) \geq 0, \right. \\
\left. \sum_{i=1}^{\min(p,q)} (1 - \chi_i(G) - \chi_i(H) - \psi(l))^+ \leq L \right\}.
\]
Let us assume \(L \geq \min(p,q)\left(\sum_{i=1}^{\min(p,q)} 2i - 1 + |p-q|\right) = \min^2(p,q) \max(p,q)\). We define \(\min(p,q) \times 1\) vector \(\varphi \triangleq [\varphi_1, \cdots, \varphi_{\min(p,q)}]^{T}\) as \(\varphi_i \triangleq \chi_i(G) + \chi_i(H) + \frac{1}{L} \sum_{l=1}^{L} \psi(l)\). For each \((\chi(G), \chi(H), \psi) \in \mathcal{R}\), we have
\[
L \varphi \geq \sum_{i=1}^{\min(p,q)} \sum_{l=1}^{L} (1 - \chi_i(G) - \chi_i(H) - \psi(l))^+
= \sum_{i=1}^{\min(p,q)} \sum_{l=1}^{L} \max \left\{ 0, 1 - \chi_i(G) - \chi_i(H) - \psi(l) \right\}
\geq \sum_{i=1}^{\min(p,q)} \max \left\{ 0, \sum_{l=1}^{L} 1 - \chi_i(G) - \chi_i(H) - \psi(l) \right\}
= L \sum_{i=1}^{\min(p,q)} (1 - \varphi_i)^+.
\]
On the other hand, according to (27) we have
\[
\mathbb{P}\{O\} \lesssim P^{-\min_{p,q}} \sum_{l=1}^{L} \psi(l) + \sum_{i=1}^{\min(p,q)} (2i-1+|p-q|)(\chi_i(G) + \chi_i(H))
\overset{(a)}{\lesssim} P^{-\min_{p,q}} \sum_{i=1}^{\min(p,q)} (2i-1+|p-q|)\varphi,
\]
\[
\overset{(29)}{\leq} P^{-\min_{p,q}} \sum_{i=1}^{\min(p,q)} (2i-1+|p-q|)\varphi_1,
\]
\[
\overset{(29)}{\leq} P^{-\min_{p,q}} \sum_{i=1}^{\min(p,q)} (2i-1+|p-q|)\varphi_i,
\]
\[
\overset{(29)}{\leq} P^{-\min_{p,q}} \sum_{i=1}^{\min(p,q)} (2i-1+|p-q|)\varphi_1,
\]
Lemma 4 is yet valid. Hence, similar to the proof of Theorem 1, we define the
\( \Psi \) of the
\[ \sum_{i=1}^{\min(p,q)} (1 - \varphi_i)^+ \leq r \], \hspace{1cm} (31)
noting that according to the definition of \( \varphi \) we can easily conclude that \( \varphi_1 \geq \cdots \geq \varphi_{\min(p,q)} \geq 0 \). According to [16], \( \Phi \) defines the probability of outage from the rate \( r \log(P) \) in an equivalent \( p \times q \) point-to-point multi-antenna Rayleigh fading channel. Hence, we have
\[ d_{RS}(r) \geq d_{p \times q}(r). \] \hspace{1cm} (32)
On the other hand, due to the cut-set bound Theorem [21] we know that the DMT of the system is upper-bounded by the minimum of the DMT of the equivalent point-to-point \( p \times m \) and \( n \times p \) multi-antenna channels. Hence,
\[ d_{RS}(r) \leq d_{opt}(r) = d_{p \times q}(r). \] \hspace{1cm} (33)
Comparing (32) and (33) completes the proof.

The statement of Theorem 1 can be generalized to multi-hop networks. However, in a general multi-hop network, the RS scheme does not necessarily achieve the optimum DMT for any number of antennas at the network nodes. The following theorem gives a sufficient condition for the RS scheme to achieve the optimal DMT in a multi-hop network:

**Theorem 2** Consider a multi-antenna multi-hop network consisting of a single source and destination and full-duplex relays, with exactly one relay in each hop. Assume that each relay is connected to the relays in the previous and next hop. Moreover, assume that for a fixed \( 1 \leq m \leq h \), we have \( \max(N_m, N_{m-1}) \leq \min(N_0, N_1, \ldots, N_{m-2}, N_{m+1}, \ldots, N_h) \) where \( h \) denotes the number of hops and \( N_i \) denotes the number of antennas at the relay in the \( i \)’th hop. (\( N_0 \) and \( N_h \) denote the number of antennas at the source and destination, respectively). Providing \( L \) is large enough such that \( L \geq \min^2(N_m, N_{m-1}) \max(N_m, N_{m-1}) \), the RS scheme achieves the optimum DMT, which is the piecewise-linear function connecting the points \( (k, (N_m - k)(N_{m-1} - k)), k = 0, 1, \ldots, \min(N_m, N_{m-1}) \).

**Proof:** Using the same argument as in Theorem 1 we can show that the probability of outage from the rate \( r \log(P) \) is equal to
\[ \mathbb{P}(\mathcal{O}) \triangleq \mathbb{P} \left\{ \sum_{i=1}^{L} \sum_{j=1}^{N_{min}} (1 - \gamma_j(A_i))^+ < Lr \right\}, \] \hspace{1cm} (34)
where \( N_{min} \triangleq \min\{N_m, N_{m-1}\} \), \( A_i \triangleq G_h \Theta_{l,h-1} G_{h-1} \cdots \Theta_{l,1} G_1 \), and \( G_i \) denotes the channel matrix between the nodes of the \( i \)’th hop and \( i-1 \)’th hop. On the other hand, applying the argument of Lemma 3 we have
\[ \gamma_i(A_i) = \gamma_i(G_h) + \gamma_{min}(\Psi_{i,l,h-1}) + \gamma_i(G_{h-1}) + \cdots + \gamma_{min}(\Psi_{i,l,1}) + \gamma_i(G_1), \] \hspace{1cm} (35)
where \( \Psi_{i,l,j} \triangleq \Psi_{i,l,j}^{H}(G_{j+1})\Theta_{l,j} U_{1,i}(G_{j}\Theta_{l,j-1} G_{j-1} \cdots \Theta_{l,1} G_1) \). Moreover, we can easily check that the statement of Lemma 3 is yet valid. Hence, similar to the proof of Theorem 1 we define the \( L \times 1 \) vector \( \psi \triangleq [\psi(1), \cdots, \psi(L)]^T \) as
ψ(l) \triangleq \max_{i,j} \gamma_{\text{min}}(Ψ_{i,l,j}). \quad \text{We have}
\begin{align*}
P \{ \psi \geq \psi_0 \} &= \prod_{l=1}^{L} P \{ \psi(l) \geq \psi_0(l) \} \\
&= \prod_{l=1}^{L} \left\{ \bigcup_{i=1}^{N_{\text{min}}} \bigcup_{j=1}^{h} (\gamma_{\text{min}}(Ψ_{i,l,j}) \geq \psi_0(l)) \right\} \\
&\leq P^{-\frac{\psi_0}{\gamma_{\text{min}}}}. \quad (36)
\end{align*}

As \( Θ_{i,j} \)'s are independent isotropic unitary matrices, their products with any possibly correlated set of unitary matrices constructs a set of independent isotropic unitary matrices [20]. Accordingly, \( Ψ_{i,l,j} \)'s are independent for different values of \( l, j \), which results in \( (a) \). Also, \( (b) \) follows from Lemma 4 and the union bound inequality.

Accordingly, applying (35) to (34) and rewriting inequality series of (23), we can upper-bound the outage probability as
\begin{align*}
P \{ O \} &\leq P \left\{ \sum_{l=1}^{L} \sum_{i=1}^{N_{\text{min}}} \left( 1 - \sum_{j=1}^{h} \chi_i(G_j) - \psi(l) \right) + \right\} < Lr \}, \quad (37)
\end{align*}
where \( \chi_i(G_j) \triangleq \gamma_{N_{\text{min}}+1-i}(G_j) \), i.e., the reverse ordering of \( \gamma_i(G_j) \)'s. Let us define the \( N_{\text{min}} \times 1 \) vectors \( \chi(G_j) \)'s as \( \chi(G_j) \triangleq [\chi_1(G_j), \chi_2(G_j), \ldots, \chi_{N_{\text{min}}}(G_j)]^T \) containing the corresponding log-values of the \( N_{\text{min}} \) smallest eigenvalues of \( G_j G_j^H \). Notice that \( \chi_1(G_j) \geq \chi_2(G_j) \geq \cdots \geq \chi_{N_{\text{min}}}(G_j) \geq 0 \). According to the upper-bound in (36) and the statistical behavior of the eigenvalues of \( G_j G_j^H \) derived in [16], we have
\begin{align*}
P \left\{ \psi \geq \hat{\psi}, \chi(G_j) \geq \hat{\chi}(G_j), j = 1, \ldots, h \right\} &\leq P^{-\frac{\psi_0}{\gamma_{\text{min}}} \sum_{l=1}^{L} \psi(l) - \sum_{j=1}^{h} \sum_{i=1}^{N_{\text{min}}} (2i + \min(N_j, N_{j-1}) - N_{\text{min}}) - 1 + \min(N_j - N_{j-1})} \chi_i(G_j) \\
&\leq P^{-\frac{\psi_0}{\gamma_{\text{min}}} \sum_{l=1}^{L} \psi(l) - \sum_{j=1}^{h} \sum_{i=1}^{N_{\text{min}}} (2i + 1 + \min(N_m - N_{m-1}), \hat{\chi}_i(G_j))} \chi_i(G_j) \}. \quad (38)
\end{align*}

Here, \( (a) \) results from the fact that \( \min(N_j, N_{j-1}) - N_{\text{min}} \geq N_j - N_{j-1} - N_{\text{min}} = N_m - N_{m-1} \) which comes form the assumption of \( \max(N_m, N_{m-1}) \leq \min(N_0, N_1, \ldots, N_{m-2}, N_{m+1}, \ldots, N_h) \). Now, we can apply the result of Lemma 5 to the region defined in (38) and the upper-bound derived in (37). Accordingly, we have
\begin{align*}
P \{ O \} &\leq P^{-\frac{\psi_0}{\gamma_{\text{min}}} \sum_{l=1}^{L} \psi(l) - \sum_{j=1}^{h} \sum_{i=1}^{N_{\text{min}}} (2i + 1 + \min(N_m - N_{m-1}), \hat{\chi}_i(G_j))} \chi_i(G_j) \}, \quad (39)
\end{align*}
where the region \( \mathcal{R} \) is defined as
\begin{align*}
\mathcal{R} &\triangleq \left\{ (\chi(G_1), \chi(G_2), \ldots, \chi(G_h), \psi) \left| \begin{array}{l}
\psi \geq 0, \chi_1(G_j) \geq \chi_2(G_j) \geq \cdots \geq \chi_{\min(p,q)}(G_j) \geq 0, j = 1, 2, \ldots, h \\
\sum_{l=1}^{L} \sum_{i=1}^{\min(p,q)} \left( 1 - \psi(l) - \sum_{j=1}^{h} \chi_i(G_j) \right) \leq Lr \end{array} \right. \right\}, \quad (40)
\end{align*}
Similar to the proof of Theorem 1, we define \( N_{\text{min}} \times 1 \) vector \( \varphi \triangleq [\varphi_1, \cdots, \varphi_{N_{\text{min}}}]^T \) as \( \varphi_i \triangleq \sum_{j=1}^{h} \chi_i(G_j) + \frac{1}{L} \sum_{l=1}^{L} \psi(l) \). Rewriting the inequality series in (29) and (30), we can upper-bound the outage probability as
\begin{align*}
P \{ O \} &\leq P^{-\frac{\psi_0}{\gamma_{\text{min}}} \sum_{i=1}^{\min(p,q)} (2i + 1 + \min(N_m - N_{m-1}), \varphi_i)} \}, \quad (41)
\end{align*}
where \( \hat{\mathcal{R}} \) is defined as
\begin{align*}
\hat{\mathcal{R}} &\triangleq \left\{ \varphi \left| \begin{array}{l}
\varphi_1 \geq \cdots \geq \varphi_{N_{\text{min}}} \geq 0, \sum_{i=1}^{\min(p,q)} (1 - \varphi_i)^+ \leq Lr \end{array} \right. \right\}, \quad (42)
\end{align*}
According to [16], (41) and (42) define the probability of outage from the rate \( r \log(P) \) in an equivalent \( N_m \times N_{m-1} \) point-to-point multi-antenna Rayleigh fading channel. Hence, we have

\[
d_{RS}(r) \geq d_{N_m \times N_{m-1}}(r).
\]  

(43)

On the other hand, due to the cut-set bound Theorem [21], we know that the DMT of the system is upper-bounded by the minimum of the DMT of the channels of different hops. Hence,

\[
d_{RS}(r) \leq d_{opt}(r) = d_{N_m \times N_{m-1}}(r).
\]  

(44)

Comparing (43) and (44) completes the proof.

**Corollary 1** Consider a multi-antenna multi-hop network consisting of a single source, a single destination and full-duplex relays with exactly one relay in each hop and assume that all the nodes are equipped with \( N \) antennas. Providing \( L \) is large enough such that \( L \geq N^3 \), the RS scheme achieves the optimum DMT, which is the piecewise-linear function connecting the points \((k,(N-k)^2)), k = 0, 1, \ldots, N\).

**B. Parallel Relay Network**

In this subsection, we consider the setup of a multi-antenna parallel relay network. In specific, we consider a two-hop network consisting of \( K > 1 \) half-duplex relays with the assumption that there is no direct link between the source and the destination. The source and the destination are shown by nodes 0 and \( K + 1 \), respectively, while the \( K \) parallel relays are denoted by the nodes 1, 2, \ldots, \( K \). Earlier, the optimum DMT of the single-antenna parallel relay network is derived by [1] and independently by [18]. Indeed, it is shown in [1] that the RS scheme can achieve the optimum DMT of the single-antenna multiple-access parallel relay network. However, much less is known regarding the DMT of the multi-antenna parallel relay networks.

Here, we show that the RS scheme achieves a better DMT with respect to the traditional AF relaying and also with respect to the other results reported in the literature. Moreover, we show that the RS scheme achieves the optimum DMT of the 2-relay parallel relay network.

**Theorem 3** Consider a multi-antenna parallel relay network consisting of a source equipped with \( m \) antennas, a destination equipped with \( n \) antennas and \( K \) half-duplex relays each equipped with \( p \) antennas. Assume that there exists no direct link between the source and the destination. For any fixed \( B \geq \min^2(p,q) \max(p,q) \), the RS scheme \(^9\) with \( L = BK \) number of paths, \( S = BK + 1 \) number of slots, the path sequence

\[
Q \equiv (q_1, \ldots, q_K, q_1, \ldots, q_K, \ldots, q_1, \ldots, q_K),
\]

in which \( q_k \equiv (0, k, K + 1) \), and the timing sequence \( s_{i,j} = i + j - 1 \), achieves the diversity gain

\[
d_{RS}(r) \geq Kd_{p \times q} \left( 1 + \frac{1}{BK} \right) r.
\]  

(45)

\(^8\)Note that in this theorem, the relays are not assumed to be isolated from each other, i.e., there may exist some links between the relays.

\(^9\)The reader is encouraged to have a look at the definition of RS scheme and its parameters in [1].
where \( q = \min(m, n) \) and \( d_{p \times q}(r) \) denotes the diversity gain of the point-to-point \( p \times q \) multi-antenna Rayleigh fading channel corresponding to the rate \( r \log(P) \). Moreover, as \( B \to \infty \), the RS scheme achieves the diversity gain \( Kd_{p \times q}(r) \).

**Proof**: The proof steps are the same as the ones presented for Theorem 3 of [1] and Theorem 6.7 of [18]. Let us denote the channel between the \( k \)'th relay and the source and the channel between the \( k \)'th relay and the destination by \( H_k \) and \( G_k \), respectively. Moreover, let us define \( r(i) = (i - 1) \mod K + 1 \). Similar to the proof of Theorem 3 in [1], one can easily show that the end-to-end channel from the source to the destination can be shown by a block lower-triangular matrix. More precisely, we have

\[
y = Fx + Qr + n_d.
\]

Here, \( x \) denotes the vector corresponding to all the paths transmitted by the source, \( y \) denotes the vector corresponding to all the paths received by the destination,

\[
F = \begin{pmatrix}
F_{1,1} & 0 & 0 & \ldots \\
F_{2,1} & F_{2,2} & 0 & \ldots \\
& \vdots & \vdots & \ddots \\
F_{L,1} & F_{L,2} & \ldots & F_{L,L}
\end{pmatrix},
\]

\[
Q = \begin{pmatrix}
Q_{1,1} & 0 & 0 & \ldots \\
Q_{2,1} & Q_{2,2} & 0 & \ldots \\
& \vdots & \vdots & \ddots \\
Q_{L,1} & Q_{L,2} & \ldots & Q_{L,L}
\end{pmatrix},
\]

where \( F_{i,i} = G_{r(i)} \alpha_i \Theta_i H_{r(i)} \), and

\[
Q_{i,i} = G_{r(i)} \alpha_i \Theta_i. \quad \text{The DMT of the end-to-end channel is equal to}
\]

\[
d_{RS}(r) = \lim_{P \to \infty} \frac{\log(P) - \log(\Pr[I(x; y) < (L + 1)r \log(P)])}{\log(P)}
\]

\[
= \lim_{P \to \infty} \frac{\log(P) \log(\Pr[\log(|I_{L,n} + PFFH (I_{L,n} + QQQ^H)^{-1}|) < (L + 1)r \log(P)])}{\log(P)}
\]

Noting \( F \) is block lower-diagonal and applying Theorem 3.3 in [18], we have

\[
|I_{L,n} + PFFFH (I_{L,n} + QQ^H)^{-1}| \geq \prod_{i=1}^{L} |I_n + P_{F,i,F_i^H}^H (I_n + \sum_{i=1}^{L} Q_{i,i}Q_{i,i}^H)^{-1}|
\]

Note that according to the constraint in the RS scheme, we have \( \alpha_i \leq 1 \). Hence, one can apply the same argument as in Lemma and show that

\[
\Pr \left\{ \prod_{i=1}^{L} |I_n + P_{F,i,F_i^H}^H (I_n + \sum_{i=1}^{L} Q_{i,i}Q_{i,i}^H)^{-1}| < P^{(L+1)r} \right\} = \Pr \left\{ \prod_{i=1}^{L} |I_n + P_{F,i,F_i^H}^H| < P^{(L+1)r} \right\}.
\]

\( ^{10} \)As only the value of \( F_{i,i} \) is needed in the proof of Theorem 3, we just give the value of \( F_{i,i} \) here. The formula for \( F_{i,j}, i < j \), is much more complicated and hence, we decide not to bring it. The same thing for \( Q_{i,j} \) defined right after this.
Moreover, using the argument in the proof of Theorem 1, one can show that with probability one, we have \( \alpha_l = 1 \). Hence, defining \( A_{k,b} \triangleq G_k \Theta_{(b-1)K+k} H_k \), we have

\[
\mathbb{P} \left\{ \prod_{l=1}^L |I_n + P F_{l,i} F_{l,i}^H| < P^{(L+1)r} \right\} \leq \mathbb{P} \left\{ \prod_{k=1}^B \prod_{b=1}^L |I_n + PA_{k,b}^H A_{k,b}| < P^{(L+1)r} \right\}. \tag{52}
\]

Let us define the \( 1 \times K \) random vector \( \sigma = [\sigma_1, \sigma_2, \ldots, \sigma_K] \) where \( \sigma_k \triangleq \sum_{l=1}^B \log |I_n + PA_{k,b}^H A_{k,b}| \). Notice that \( \sigma_k \)'s are independent of each other. As \( B \geq \min^2(p,q) \max(p,q) \), we can apply Theorem 1 to \( \sigma_k \)'s. Hence, for any fixed \( 1 \times K \) vector \( \hat{\sigma} \geq 0 \), we have

\[
\mathbb{P} \{ \sigma \geq \hat{\sigma} \} \overset{(a)}{=} P^{-\sum_{k=1}^K d_{p \times q}(\sigma_k)}. \tag{53}
\]

Here, \((a)\) results from Theorem 1 and the fact that \( \sigma_k \)'s are independent of each other. Denoting the outage event as \( \mathcal{O} \), according to (49), (50), (51), and (52), we have

\[
\mathbb{P} \{ \mathcal{O} \} \overset{\text{(a)}}{\leq} \mathbb{P} \left\{ \sum_{k=1}^K |\sigma_k| \leq \left( K + \frac{1}{B} \right) r \right\}. \tag{54}
\]

Let us define the region \( \mathcal{R} \triangleq \left\{ \sigma \geq 0 \left| \sum_{k=1}^K |\sigma_k| \leq \left( K + \frac{1}{B} \right) r \right. \right\} \). We have

\[
\mathbb{P} \{ \mathcal{O} \} \overset{\text{(a)}}{\leq} \mathbb{P} \{ \mathcal{R} \} \overset{\text{(b)}}{\leq} P^{-\min_{\sigma \in \mathcal{X}} \sum_{k=1}^K d_{p \times q}(\sigma_k)} = P^{-Kd_{p \times q}(1 + \frac{1}{B^2} r)}. \tag{55}
\]

Here, \((a)\) results from Lemma 5 \((b)\) results from the fact that \( d_{p \times q}(r) \) is a convex decreasing function and, as a result, we have \( \frac{1}{K} \sum_{k=1}^K d_{p \times q}(\sigma_k) \geq d_{p \times q}\left( \frac{1}{K} \sum_{k=1}^K \sigma_k \right) \geq d_{p \times q}(1 + \frac{1}{B^2} r) \). \(55\) completes the proof of the theorem.

In the following theorem, we show that the RS scheme achieves the optimum DMT for the two-relays half-duplex parallel relay network in which \( m = n \), but the two parallel relays can have a different number of antennas but there is no direct link between them. Figure 2 shows the schematic of such a network.

**Theorem 4** Consider a multi-antenna parallel relay network consisting of a source and a destination each equipped with \( m \) antennas, and \( K = 2 \) half-duplex relays equipped with \( n_k, k = 1, 2 \) antennas. Assume that there exists no direct link between the source and the destination and also between the two relays. Consider the RS scheme with \( L = BK \), \( S = BK + 1 \), and the path and timing sequences defined in Theorem 3 As \( B \to \infty \), the RS scheme achieves the optimum DMT of the network.

---

Fig. 2. A schematic of the MIMO parallel 2 relays network with no direct link between the source and the destination and also between the relays.
Proof: First, notice that according to the argument of Theorem 3 as $B \to \infty$, the RS scheme achieves the DMT $d_{RS,\infty}(r) \triangleq \min_{0 \leq \nu \leq 2r} d_{m \times n_1}(\nu) + d_{m \times n_2}(2r - \nu)$. Now, to prove the Theorem, we just have to show that $d_{RS,\infty}(r)$ is indeed an upper-bound for the optimum DMT. According to the cut-set Theorem [21], we have an upper-bound for the capacity of the network for each channel realization. Hence, we can apply the cut-set Theorem to find an upper-bound for the optimum DMT. In general, for any arbitrary half-duplex relay network with $K$ relays and any set $\{0\} \subseteq S \subseteq \{0, 1, \ldots, K\}$, we say the network is in the state $S$, if the network nodes in $S$ are transmitting and the network nodes in $S^C \triangleq \{0, 1, \ldots, K + 1\} \setminus S$ are receiving. Notice that as the source is always transmitting and the destination is always receiving, we have $0 \in S, K + 1 \in S^C$. Accordingly, we define a $1 \times 2^K$ state vector $\rho$ such that for any set $\{0\} \subseteq S \subseteq \{0, 1, \ldots, K\}$, $\rho_S$ shows the portion of time that the half-duplex relay network spends in the state $S$ ($\sum_{\{0\} \subseteq S \subseteq \{0, 1, \ldots, K\}} \rho_S = 1$). As the channels are assumed to be fixed for the whole transmission period and the relay nodes and the source are assumed to have no channel state knowledge about their forward channels, we can assume that a fixed state vector $\rho$ is associated with the strategy that achieves the optimum DMT. Denoting the outage event by $\mathcal{O}$, for any general half-duplex relay network consisting of $K$ relays, we have

$$\mathbb{P}\{\mathcal{O}\} \stackrel{(a)}{=} \min_{\rho} \max_{0 \leq \nu \leq 2r} \mathbb{P}\left\{ \sum_{\{0\} \subseteq S \subseteq \{0, 1, \ldots, K\}} \rho_S I(X(S \cap T); Y(S^C \cap T^C) | X(S \cap T^C)) < r \log(P) \right\},$$

$$\mathbb{P}\{\mathcal{O}\} \stackrel{(b)}{=} \min_{\rho} \max_{0 \leq \nu \leq 2r} \mathbb{P}\left\{ \sum_{\{0\} \subseteq S \subseteq \{0, 1, \ldots, K\}} \rho_S I(X(S \cap T); Y(S^C \cap T^C) | X(S \cap T^C)) < r \log(P) \right\},$$

$$\mathbb{P}\{\mathcal{O}\} \stackrel{(c)}{=} P^{-d_{RS,\infty}(r)},$$

where $1[x] = 1$ for $x \geq 0$ and is 0 otherwise. Here, (a) results from taking the maximization of the right-hand side of (56) over $T_1, T_2$. (b) results from the facts that i) conditioned on $\mathcal{O}_1$ and assuming $r \geq (\rho_{0,1} + \rho_{0,2}) \hat{\nu} + (\rho_{0,1} + \rho_{0,2}) (\nu - \hat{\nu})$, we have $(\rho_{0,1} + \rho_{0,2}) | I_m + PG_1 G_1^H | + (\rho_{0,1} + \rho_{0,2}) | I_n + PH_2 H_2^H | \leq \nu \log(P)$; and ii) conditioned on $\mathcal{O}_2$ and assuming $r \geq (\rho_{0,1} + \rho_{0,2}) \hat{\nu} + (\rho_{0,1} + \rho_{0,2}) (\nu - \hat{\nu})$, we conclude that $(\rho_{0,2} + \rho_{0,1}) | I_m + PG_2 G_2^H | +
18

Fig. 3. Parallel relay network with $K = 2$ relays, each node with 3 antennas and no direct link between source and destination.

$$\begin{align*}
(r_0 + r_0) \left| I_{n_1} + P H_1 H_f^T \right| &\leq r \log(P). \quad (c) \text{ results from } i) \mathbb{P} \{C_1\} = \mathbb{P} \{C_2\} = P^{-d_{m \times n_1}(\hat{\nu}) - d_{m \times n_2}(2r - \hat{\nu})} = P^{-d_{RS, \infty}(r)} \\
\text{and ii) } r_0 + r_{0,1} + r_{0,2} + r_{0,1,2} = 1 \text{ (due to the definition of } \rho) \text{ which results in having } r - (r_{0,1,2} + r_{0,1}) \hat{\nu} - (r_{0,1} + r_{0}) (2r - \hat{\nu}) = - [r - (r_{0} + r_{0,2}) \hat{\nu} - (r_{0,2} + r_{0,1,2}) (2r - \hat{\nu})] \text{ and consequently,} \\
1 [r - (r_{0,1,2} + r_{0,1}) \hat{\nu} - (r_{0,1} + r_{0}) (2r - \hat{\nu})] + 1 [r - (r_{0} + r_{0,2}) \hat{\nu} - (r_{0,2} + r_{0,1,2}) (2r - \hat{\nu})] = 1.
\end{align*}$$

Equation (57) completes the proof of the theorem.

Figure 3 shows the DMT of various schemes for the parallel relay network with $K = 2$ relays and $m = n = p = 3$, i.e. 3 antennas at each node. As it is shown in Theorem 3 the RS scheme achieves the optimum DMT. However, if we do not apply random unitary matrix multiplication at the relay nodes, applying the steps in the proof of Theorem 3 one can easily show that the RS scheme achieves the DMT of $K d_{GH}(r)$, where $d_{GH}(r)$ denotes the DMT of the product of the channel matrix $H$ from the source to the relay and the channel matrix $G$ from the relay to the destination (see (4)). Finally, applying the NAF scheme of [2], [17], one can easily show that the DMT $K d_{GH}(2r)$ is achievable.

C. Multiple-Antenna Single Relay Channel

In this subsection, we consider the most-studied scenario in the relay network, the single relay setup, in which a direct link exists between the source and the destination. The relay is assumed to be half-duplex. There have been extensive research on this particular setup toward characterization of the DMT. The authors of [2] have shown that the NAF scheme achieves the best DMT among all possible AF relaying schemes for the Single-Input Single-Output (SISO) single half-duplex relay channel. However, here we show that using independent uniformly random unitary matrices across different time-slots improves the DMT of the NAF scheme for the multi-antenna setup. In order to exploit the potential benefit from random unitary matrix
multiplication, the source transmits in $2B$ consecutive time-slots. In the odd time-slots, the relay listens to the source signal. In the even time slots, the relay multiplies the received signal from the last time-slot with a uniformly random unitary matrix and then amplifies the result with the maximum possible coefficient, which is less than or equal to 1. The destination decodes the transmitted message based on the joint decoding of the signal it receives in the $2B$ time-slots.

**Theorem 5** Consider a single relay channel consisting of a source, a half-duplex relay, and a destination equipped with $m$, $p$, and $n$ antennas, respectively. Let us consider a modified NAF scheme that benefits from the random unitary matrix multiplication at the relay node and the joint decoding at the destination side through $2B$ time-slots. Assuming $B \geq \min^2(p, q) \max(p, q)$ where $q \triangleq \min(m, n)$, the modified NAF scheme achieves the following DMT

$$d_{\text{MNAF}}(r) \geq d_{m \times n}(r) + d_{p \times q}(2r).$$

(58)

**Proof:** The proof is similar to the proof of Theorem 3. Indeed, assuming the source-destination, source-relay, and relay-destination channel matrices are denoted by $F$, $H$, and $G$, respectively, we can show that the end-to-end channel matrix is equal to

$$\mathcal{F} = \begin{pmatrix}
F & 0 & 0 & 0 & \ldots \\
0 & \alpha G \Theta_1 H & 0 & 0 & \ldots \\
0 & 0 & F & 0 & \ldots \\
0 & 0 & 0 & \alpha G \Theta_2 H & F & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots 
\end{pmatrix}.$$

(59)

Here, we observe that the lower-diagonal elements are independent of the diagonal elements. Hence, we can apply Theorem 3.3 in [18]. Accordingly, the DMT corresponding to the end-to-end system is greater than or equal to the summation of the DMT of the forward channel $F$ and the DMT of a two-hop channel utilized in half of the time. In other words, $d_{\text{MNAF}}(r) \geq d_{m \times n}(r) + d_{p \times q}(2r)$. Details of the proof are similar to the proof of Theorem 3.

Figure 4 compares the achievable DMT of the NAF scheme with the achievable DMT of the modified NAF scheme for the single relay channel with $m = n = p = 3$ and $m = n = p = 4$ antennas. Reference [13] has shown that the NAF protocol achieves the DMT $d_{\text{NAF}}(r) \geq d_{m \times n}(r) + d_{GH}(2r)$. As we observe, the modified NAF scheme outperforms the NAF scheme for small values of $r$.

**D. General Full-Duplex Relay Networks**

Here, we generalize the statement of Remark 6 in [1] and Theorem 4.2 in [18] to the multi-antenna case. Indeed, it is shown in [1] that the RS scheme achieves a linear DMT connecting the points $(0, d_{\text{max}})$ and $(r_{\text{max}}, 0)$, where $d_{\text{max}}$ denotes the maximum diversity and $r_{\text{max}}$ denotes the maximum multiplexing-gain (which is 1) for single-antenna full-duplex relay networks whose underlying graph is directed acyclic. However, in the multi-antenna setup, [1] could only show that the RS scheme achieves the maximum diversity gain. To generalize the statement, we have to review the definitions from [1].

11A directed graph is called directed acyclic if it contains no directed cycle [22]
Definition 1 A general multi-relay network with the connectivity graph $G = (V, E)$ is a wireless network with the vertex set $V = \{0, 1, \ldots, K+1\}$ where 0 and $K+1$ represent the source and the destination, respectively, and the other nodes represent the relays. Each pair of vertices that belong to $E$ are connected together through a quasi-static Rayleigh-fading channel and the pairs of vertices that do not lie in $E$ are disconnected from each other. The number of antennas at node $i$ is denoted by $N_i$.

Definition 2 For a network with the connectivity graph $G = (V, E)$, a cut-set on $G$ is defined as a subset $S \subseteq V$ such that $0 \in S, K + 1 \in S^c$. The weight of the cut-set corresponding to $S$, denoted by $w_G(S)$, is defined as

$$w_G(S) = \sum_{a \in S, b \in S^c, (a, b) \in E} N_a N_b.$$  \hfill (60)

Theorem 6 Consider a full-duplex multi-antenna multi-relay network with the graph $G = (V, E)$ where $G$ is directed acyclic. Assume each node has $N$ antennas. The RS scheme achieves the following DMT

$$d_{RS}(r) = \min_{S} w_G(S) \frac{d_{N \times N}(r)}{N^2}. \hfill (61)$$

where $d_{N \times N}(r)$ denotes the DMT of a $N \times N$ multi-antenna channel.

Proof: Using the same path sequence as the one in the proof of Remark 6 in [1] and applying the result of Corollary 1, (61) can be derived. The steps are the same as the steps in the proof of Theorem 3 noting the equivalent point-to-point channel is block lower-triangular and the function $d_{N \times N}(r)$ is convex decreasing.

Remark 1- According to (61), a specific RS scheme with fixed path and timing sequences can simultaneously achieve the maximum diversity gain, which is $\min_{S} w_G(S)$ (refer to [1]), and the maximum multiplexing gain, which is $N$ in a multi-
antenna relay network whose underlying graph is directed acyclic. In other words, the RS scheme is \textit{robust} in the sense that it achieves the corner-points of the optimum DMT with no modification of the scheme parameters.

Figure 5 shows an example of a directed acyclic network. The relays are operating in the full-duplex mode and each node is equipped with two antennas. Here, the weight of the minimum cut-set depicted in the figure is 8. Hence, applying the argument of Theorem 5 the RS scheme achieves \( d_{RS}(r) = 2d_{2\times2}(r) \). However, the DMT upper-bound is equal to \( d_{ub}(r) = d_{2\times4}(r) \), which is obtained from the same cut-set. Although the two DMT’s are equal in the corner-points, they do not coincide in between.

IV. Conclusion

In this paper, we derived new DMT results in various setups of multi-antenna relay network using AF relaying. For this purpose, the application of RS scheme proposed in [1] was studied. It was shown that random unitary matrix multiplication at the relay nodes enables the RS scheme to achieve a better DMT in comparison to the traditional AF relaying. First, the multi-antenna full-duplex single-relay two-hop network was studied for which the RS scheme was shown to achieve the optimum DMT. This result was also generalized to the multi-antenna multi-hop full-duplex relay network with one relay in each hop. Next, applying this result, a new achievable DMT was derived for multi-antenna half-duplex parallel relay networks. Interestingly, it turned out that the DMT of the RS scheme is optimum for the multi-antenna half-duplex parallel two-relay setup with no direct link between the relays. Moreover, using random unitary matrix multiplication was shown to improve the DMT of the NAF scheme of [2] in the setup of the widely used multi-antenna single relay channel. Finally, the general full-duplex multi-antenna relay network was studied and a new lower-bound was obtained on DMT using the RS scheme, assuming that the underlying network graph is directed acyclic.

APPENDIX I

Proof of Lemma I

Assuming a multiplexing gain of \( r \), the diversity of the original system can be written as

\[
\begin{aligned}
d(r) &= \lim_{P \to \infty} - \frac{\log \left( \mathbb{P} \left\{ \log \left| \mathbf{I} + \alpha^2 \mathbf{P} \mathbf{G} \mathbf{H}^H \mathbf{G}^H \left( \mathbf{I} + \alpha^2 \mathbf{G} \mathbf{G}^H \right)^{-1} \right| < r \log(P) \right\} \right)}{\log(P)} \\
&= \frac{\log(2)}{\log(P)}.
\end{aligned}
\]  

(62)
Since $\alpha^2GG^H \succeq 0$, it follows that $\left| I + \alpha^2PHH^HG^H \left( I + \alpha^2GG^H \right)^{-1} \right| < \left| I + \alpha^2PHH^HG^H \right|$. This implies that

$$d(r) \leq d_u(r) \triangleq \lim_{P \to \infty} -\frac{\log \left( \mathbb{P} \left\{ \log \left| I + \alpha^2PHH^HG^H \right| < r \log(P) \right\} \right)}{\log(P)},$$

which is the DMT of the system in (3). Moreover, selecting $\alpha = \min \left( 1, \frac{P}{\mathbb{E}[\|Hx+n\|_2^2]} \right)$ (which guarantees that the output power of the relay remains below $P_{12}$, we have

$$\alpha^2GG^H \preceq \|G\|^2 I.$$  \hspace{1cm} (64)

Defining the event $\mathcal{E} \equiv \{ \|G\|^2 > c \log(P) \}$, and noting that $\|G\|^2$ is a Chi-square random variable with $2pn$ degrees of freedom, we have

$$\mathbb{P}\{\mathcal{E}\} = \sum_{k=0}^{pn-1} \frac{(c \log(P))^k}{k!} e^{-c \log(P)} \sim d(\log(P))^{pn-1} P^{-c},$$

where $d = \frac{c}{e}$. Defining the outage event of the system in (1) as $\mathcal{O}$, we have

$$\mathbb{P}\{\mathcal{O}\} = \mathbb{P}\{\mathcal{O}|\mathcal{E}\} \mathbb{P}\{\mathcal{E}\} + \mathbb{P}\{\mathcal{O}|\mathcal{E}^c\} \mathbb{P}\{\mathcal{E}^c\} \leq \mathbb{P}\{\mathcal{E}\} + \mathbb{P}\{\mathcal{O}|\mathcal{E}^c\} \mathbb{P}\{\mathcal{E}^c\}.$$  \hspace{1cm} (66)

Conditioned on $\mathcal{E}^c$, the probability of the outage event can be upper-bounded as

$$\mathbb{P}\{\mathcal{O}|\mathcal{E}^c\} \leq \mathbb{P}\left\{ \log \left| I + \alpha^2 \frac{P}{\log P + 1} GH^H G^H \right| < r \log(P) \right\},$$

which is equal to the probability of the outage event of the system in (3), denoted as $\mathcal{O}_1$, conditioned on $\mathcal{E}^c$. In other words, $\mathbb{P}\{\mathcal{O}|\mathcal{E}^c\} \leq \mathbb{P}\{\mathcal{O}_1|\mathcal{E}^c\}$. Substituting in (66) yields

$$\mathbb{P}\{\mathcal{O}\} \leq \mathbb{P}\{\mathcal{E}\} + \mathbb{P}\{\mathcal{O}_1|\mathcal{E}^c\} \mathbb{P}\{\mathcal{E}^c\} \leq \mathbb{P}\{\mathcal{E}\} + \mathbb{P}\{\mathcal{O}_1\}.\hspace{1cm} (68)$$

Since the capacity of the two-hop network is equal to the capacity of both the source-relay and the relay-destination links, it follows that the outage event $\mathcal{O}$ includes $\mathcal{O}_{sr}$, the event of outage in the source-relay link, and $\mathcal{O}_{rd}$, the event of outage in the relay-destination link. As a result,

$$\mathbb{P}\{\mathcal{O}\} \geq \max \{ \mathbb{P}\{\mathcal{O}_{sr}\}, \mathbb{P}\{\mathcal{O}_{rd}\} \} \geq \max \{ P^{-mp}, P^{-pn} \} \geq P^{-pn}.$$  \hspace{1cm} (69)

(a) results from the fact that the outage probability corresponding to the multiplexing gain $r$ is greater than or equal to the outage probability corresponding to the multiplexing gain 0. Setting $c = 2pn - 1$, from (65), it is concluded that

$$\frac{\mathbb{P}\{\mathcal{E}\}}{\mathbb{P}\{\mathcal{O}\}} = \Theta \left( \left( \frac{\log(P)}{P} \right)^{pn-1} \right).$$

12Note that as we are looking for a lower-bound for the DMT, it is fine to select any arbitrary value for $\alpha$. 


From the above equation and (68), it follows that

$$P \{ \mathcal{O} \} \leq P \{ \mathcal{O}_1 \}, \quad (71)$$

which incurs that the DMT of the original system is lower-bounded by the DMT of the system in (3). This completes the proof of Lemma 1.

**Appendix II**

**Proof of Lemma 2**

Defining $P' = \frac{P}{c \log(P) + 1}$, the DMT of the system in (3) can be written as

$$d_1(r) = \lim_{P' \to \infty} - \frac{\log \left( \mathbb{P} \left\{ \log \left| I + \alpha^2 P' G^H H G H \right| < r \log(P) \right\} \right)}{\log(P')} \quad (a)$$

$$= \lim_{P' \to \infty} - \frac{\log \left( \mathbb{P} \left\{ \log \left| I + \alpha^2 P' G^H H G H \right| < r' \log(P') \right\} \right)}{\log(P')} \quad (b)$$

$$= d_u(r'), \quad (c)$$

(72)

where $r' \triangleq r \frac{\log(P)}{\log(P')}$ and (a) follows from the fact that as $P' = \frac{P}{c \log(P) + 1}$, we have $P = P' \left( c \log(P') + O(\log \log(P')) \right)$, which implies that $\lim_{P' \to \infty} \frac{\log(P')}{\log(P)} = 1$. In other words, in the first line of the preceding equation, one can substitute $\log(P)$ by $\log(P')$. (b) results from the fact that the second line in the right hand side of the preceding equation is exactly the DMT of the system in (2) at $r'$, which is denoted by $d_u(r')$. Finally, (c) follows from the facts that $\lim_{P' \to \infty} \frac{r'}{r} = 1$ and the continuity of the DMT curve which implies that $d_u(r') = d_u(r)$. This completes the proof of Lemma 3.

**Appendix III**

**Proof of Lemma 3**

First, notice that for values of $i > \min \{ m, p \}$ or $i > \min \{ p, n \}$, $\lambda_i(G)$ or $\lambda_i(H)$ are defined as zero. Hence, the argument of this lemma is obvious in these cases. Now, we prove the argument for $i \leq \min \{ m, n, p \}$. According to the Courant-Fischer-Weyl Theorem [23], we have

$$\lambda_i(G \Theta H) = \max_{S, \dim(S) = i} \min_{x \in S} \frac{\| G \Theta H x \|^2}{\| x \|^2}. \quad (73)$$
Now, let us define $S_0 = (v_1(H), v_2(H), \ldots, v_i(H))$ where $v_j(H)$ denotes the $j$'th column of $V(H)$ and $\langle a_1, a_2, \ldots, a_k \rangle$ denotes the span of $a_1, a_2, \ldots, a_k$. We have

$$\lambda_i(G\Theta H) \geq \min_{x \in S_0} \frac{\|G\Theta H x\|^2}{\|x\|^2} \tag{a}$$

$$= \min_{x' \in C'} \frac{\|G\Theta H V(1,i)(H)x'\|^2}{\|x'\|^2} \tag{b}$$

$$= \min_{x' \in C'} \frac{\|G\Theta U(H)\Lambda^\frac{1}{2}(H) I_{n \times i} x'\|^2}{\|x'\|^2} \tag{c}$$

$$= \min_{x' \in C'} \|G\Theta U(H)\Lambda^\frac{1}{2}(H)x'\|^2 \quad \|x'\|^2 \geq 1 \tag{d}$$

$$\geq \min_{y' \in C'} \lambda_i(H) \|G\Theta U(1,i)(H)y'\|^2 \quad \|y'\|^2 \geq 1 \tag{e}$$

$$\geq \min_{y' \in C'} \lambda_i(H) \|\Lambda^\frac{1}{2}(G)V^H(G)\Theta U(1,i)(H)y'\|^2 \quad \|y'\|^2 \geq 1 \tag{f}$$

$$= \min_{y' \in C'} \lambda_i(G)\lambda_i(H) \|V^H(1,i)(G)\Theta U(1,i)(H)y'\|^2 \quad \|y'\|^2 \geq 1 \tag{g}$$

$$= \lambda_i(G)\lambda_i(H) \lambda_{\min} \left( V^H(1,i)(G)\Theta U(1,i)(H) \right) \tag{h},$$

where $I_{n \times i}$ denotes the diagonal identity $n \times i$ matrix and $\Lambda^\frac{1}{2}(H)$ denotes the square submatrix of $\Lambda^\frac{1}{2}(H)$ consisting of its first $i$ rows and first $i$ columns. Here, (a) follows from the fact that $i) x \in S_0$ is equivalent to $x = V(1,i)x'$ for some $x' \in C'^i$; and ii) $\|x\|^2 = \|x'\|^2$. (b) follows from the fact that for any $x' \in C'$, $\|x'\|^2 \geq 1$, defining $y' = \frac{1}{\sqrt{\lambda_i(H)}} \Lambda^\frac{1}{2}(H)x'$, we have $\|y'\|^2 \geq \|x'\|^2 \geq 1$. (c) follows from the fact that for any unitary matrix $P$, we have $\|PA\|^2 = \|A\|^2$. (d) follows from the fact that defining $z' = \Theta U(1,i)(H)y'$, we have

$$\left\| \Lambda^\frac{1}{2}(G)V^H(G)z' \right\|^2 = \left\| \Lambda^\frac{1}{2}(G)V^H(1,i)(G)z' \right\|^2 + \left\| \Lambda^\frac{1}{2}(G)V^H(i+1,m)(G)z' \right\|^2.$$

(e) follows from the fact that defining $w' = V^H(1,i)(G)\Theta U(1,i)(H)y'$, we have $\left\| \Lambda^\frac{1}{2}(G)w' \right\|^2 \geq \lambda_i(G)\|w'\|^2$. Finally, (f) follows from the Courant-Fischer-Weyl Theorem [23]. (74) completes the proof of Lemma.
We have
\[
\left| \Psi_{t,l} \Psi_{t,l}^H \right| = \lambda_{\min} (\Psi_{t,l}) \prod_{j=1}^{i-1} \lambda_j (\Psi_{t,l})
\]
\[
\leq (a) \lambda_{\min} (\Psi_{t,l}) \left( \frac{\text{Tr} \left\{ \Psi_{t,l} \Psi_{t,l}^H \right\}}{i-1} \right)^{i-1},
\]
(75)
where \((a)\) follows from the Geometric Inequality and the fact that \(\sum_{j=1}^{i-1} \lambda_j (\Psi_{t,l}) \leq \text{Tr} \left\{ \Psi_{t,l} \Psi_{t,l}^H \right\} \). \(\text{Tr} \left\{ \Psi_{t,l} \Psi_{t,l}^H \right\} \) can be upper-bounded as follows:
\[
\text{Tr} \left\{ \Psi_{t,l} \Psi_{t,l}^H \right\} = \left\| V_{(1,i)} (G) \Theta_i U_{(1,i)} (H) \right\|^2
\]
\[
\leq (a) \left\| V_{(1,i)} (G) \right\|^2 \left\| \Theta_i U_{(1,i)} (H) \right\|^2
\]
\[
\leq (b) \left\| V_{(1,i)} (G) \right\|^2 \left\| U_{(1,i)} (H) \right\|^2
\]
\[
\leq (c) \epsilon^2.
\]
(76)
In the preceding equation, \((a)\) follows from the fact that \(\|AB\|^2 \leq \|A\|^2 \|B\|^2\) for any two matrices \(A\) and \(B\). \((b)\) results from the fact that \(\|\Theta A\| = \|A\|\), for any matrix \(A\) and any unitary matrix \(\Theta\). Finally, \((c)\) follows from the fact that as \(U(H)\) and \(V(G)\) are unitary matrices, each of their columns has unit norm. Combining (75) and (76) yields
\[
\left| \Psi_{t,l} \Psi_{t,l}^H \right| \leq \lambda_{\min} (\Psi_{t,l}) \left( \frac{\epsilon^2}{i-1} \right)^{i-1},
\]
(77)
which implies that
\[
\lambda_{\min} (\Psi_{t,l}) \geq c \left| \Psi_{t,l} \Psi_{t,l}^H \right|,
\]
(78)
for some constant \(c\). Denoting the \(j\)th column of \(\Psi_{t,l}\) as \(\Psi_{t,l}^{(j)}\), we write the determinant of \(\Psi_{t,l} \Psi_{t,l}^H\) as
\[
\left| \Psi_{t,l} \Psi_{t,l}^H \right| = \prod_{j=1}^{i} \beta_j,
\]
(79)
where \(\beta_j\) denotes the square norm of the projection of \(\Psi_{t,l}^{(j)}\) over the null-space of the subspace spanned by \(\left\{ \Psi_{t,l}^{(s)} \right\}_{s=1}^{j-1}\). Combining (78) and (79) yields
\[
\mathbb{P} \left\{ \lambda_{\min} (\Psi_{t,l}) \leq \epsilon \right\} \leq \mathbb{P} \left\{ \prod_{j=1}^{i} \beta_j \leq \kappa \epsilon \right\}
\]
\[
\leq \mathbb{P} \left\{ \bigcup_{j=1}^{i} \left\{ \beta_j \leq \sqrt{\kappa \epsilon} \right\} \right\}
\]
\[
\leq (a) \sum_{j=1}^{i} \mathbb{P} \left\{ \beta_j \leq \sqrt{\kappa \epsilon} \right\},
\]
(80)
where \(\kappa = \frac{1}{c}\) and \((a)\) follow from the union bound on the probability. \(\Psi_{t,l}^{(s)}\) can be considered as the projection of the \(s\)th column of the matrix \(R_l \triangleq \Theta_l U_{(1,i)} (H)\), denoted by \(R_{l,s}\), over the \(i\)-dimensional subspace spanned by \(\left\{ V_i (G) \right\}_{i=1}^{25}\), which
is denoted by $\mathcal{P}$. Now, consider a random unit vector $w$ in $\mathcal{P}$, which is orthogonal to the first $j-1$ columns of the matrix $R_l$. Since $\{\Psi_{i,l}^{(s)}\}_{i=1}^{j-1}$ are the projections of the first $j-1$ columns of $R_l$ over $\mathcal{P}$, it follows that $w$ is also orthogonal to $\{\Psi_{i,l}^{(s)}\}_{s=1}^{j-1}$. In other words, $w$ belongs to $\mathcal{Q}^\perp$, the null-space of the subspace spanned by $\{\Psi_{i,l}^{(s)}\}_{s=1}^{j-1}$. Hence, we have

$$\beta_j = \left\| \Psi_{i,l}^{(j)H} w \right\|^2 \geq \left\| \Psi_{i,l}^{(j)H} \right\|^2,$$

where $a^H \ast \mathcal{T}$ denotes the projection of the vector $a$ over the subspace $\mathcal{T}$. The second line in the preceding equation follows from the fact that the norm of the projection of a vector over a subspace is greater than the norm of the projection of that vector over any arbitrary unit vector in that subspace. Since $\Psi_{i,l}^{(j)}$ is the projection of $R_l^{(j)}$ over $\mathcal{P}$, $R_l^{(j)}$ can be written as

$$R_l^{(j)} = \Psi_{i,l}^{(j)} + \Psi_{i,l}^{(j)\perp},$$

where $\Psi_{i,l}^{(j)\perp}$ denotes the projection of $R_l^{(j)}$ over $\mathcal{P}^\perp$, the null-space of $\mathcal{P}$. Since $w \in \mathcal{P}$, it follows that $w^H \Psi_{i,l}^{(j)\perp} = 0$. This means that $\Psi_{i,l}^{(j)H} w = R_l^{(j)H} w$. Combining the above with (81) yields

$$\beta_j \geq \left\| R_l^{(j)H} w \right\|^2.$$

Since $\Theta_l$ and $U(H)$ are unitary matrices and $U(H)$ is isotropically distributed [20], it follows that $\Theta_l U(H)$ is an isotropic unitary matrix. This means i) $\{R_l^{(s)}\}_{s=1}^{j-1}$ is an orthonormal set, which implies that $R_l^{(j)}$ is orthogonal to $R_l^{(s)}$, $s = 1, \ldots, j-1$, and ii) $R_l^{(j)}$ is an isotropic unit vector. As a result, $R_l^{(j)}$ is an isotropic unit vector in the $(p-j+1)$-dimensional subspace perpendicular to $\{R_l^{(s)}\}_{s=1}^{j-1}$. Noting that $w$ is also in this subspace and $R_l^{(j)}$ and $w$ are independent of each other, from [24], Lemma 3, the CDF of $Z_{j} \triangleq \left\| R_l^{(j)H} w \right\|^2$ can be computed as

$$F_{Z_j}(z) = 1 - (1 - z)^{p-j}. \quad (84)$$

Combining (80), (83), and (84), it follows that

$$\mathbb{P} \{\lambda_{\min}(\Psi_{i,l}) \leq \varepsilon\} \leq \sum_{j=1}^{i} F_{Z_j}(\sqrt{\eta \varepsilon}) \leq \sum_{j=1}^{i} \left[ 1 - (1 - \sqrt{\eta \varepsilon})^{p-j} \right] \leq \sum_{j=1}^{i} (p-j) \sqrt{\eta \varepsilon} \leq \eta \sqrt{\varepsilon},$$

where $\eta = i(p - \frac{i+1}{2}) \sqrt{\kappa}$. In the above equation, (a) follows from the fact that $(1-x)^n \geq 1 - nx, \forall n \geq 0, 0 \leq x \leq 1$. This completes the proof of Lemma 4.

13 Note that as $j \leq i$, $\mathcal{Q}^\perp$ has at least dimension of 1.
APPENDIX V

PROOF OF LEMMA \[5\]

We can assume that \( f(x) \) is defined such that \( x \leq y \Rightarrow f(x) \leq f(y) \); otherwise, we can redefine \( f(x) \) as \( f(x) = \inf_{y \geq x} f(y) \). Let us define the set \( S(\epsilon) \) for every \( \epsilon > 0 \) as \( S(\epsilon) = \{ x \in [0, \infty)^n \mid d(x, R) \leq \sqrt{n} \epsilon, x \in \mathbb{Z}^n \} \) where \( d(x, R) = \inf_{y \in R} \|x - y\| \). Also, let us define the partial order \( \leq \) for two sets \( A, B \subseteq [0, \infty)^n \) as follows: \( A \leq B \) iff for every \( x \in B \) there exists a \( y \in A \) such that \( y \leq x \). One can easily verify that \( S(\epsilon) \leq \mathcal{R} \). Now, let us define the set \( L(\epsilon) \) as \( L(\epsilon) = \{ x \in S(\epsilon) \mid \exists y \in S(\epsilon), y < x \} \), i.e., \( L(\epsilon) \) consists of the minimal members of \( S(\epsilon) \). Notice that the elements of \( S(\epsilon) \) can be mapped to the elements of \( \mathbb{Z}^n_+ \) such that the partial order \( < \) between the real vectors is kept between the corresponding integer vectors.\[14\] For every subset \( A \subseteq \mathbb{Z}^n_+ \) and every \( x \in A \), there exists a minimal member \( y \in A \) such that \( y \leq x \). Accordingly, we have such a property for \( S(\epsilon) \). This means \( L(\epsilon) \leq S(\epsilon) \). Noticing \( S(\epsilon) \leq \mathcal{R} \), we conclude that \( L(\epsilon) \leq \mathcal{R} \).

On the other hand, it is easy to check that \( L(\epsilon) \) is a finite set, i.e. \( |L(\epsilon)| < \infty \). Hence, we have

\[
\mathbb{P}(\mathcal{R}) \leq \mathbb{P} \left( \bigcup_{x' \in L(\epsilon)} (x \geq x') \right) \leq |L(\epsilon)| \max_{x' \in L(\epsilon)} \mathbb{P} \{ x \geq x' \} \leq P^\min_{x \in L(\epsilon)} f(x),
\]

where (a) follows from \( L(\epsilon) \leq \mathcal{R} \) and (b) follows from the fact that \( |L(\epsilon)| < \infty \). Now, let us define \( h(\epsilon) = \min_{x \in L(\epsilon)} f(x) \).

For every \( \epsilon \), we have \( h(\epsilon) = \inf_{x \in \mathcal{R}} f(x) \) (otherwise, according to \[86\], the statement of the Lemma is proved). Now, we prove that \( \lim_{\epsilon \to 0} h(\epsilon) \) exists and in fact, we have \( \lim_{\epsilon \to 0} h(\epsilon) = \inf_{x \in \mathcal{R}} f(x) \). As \( f(x) \) is uniformly continuous, there exists a positive function \( g(\epsilon) \) \( (g(\epsilon) > 0) \) such that for all \( x, y, \|x - y\| \leq g(\epsilon) \), we have \( |f(x) - f(y)| \leq \epsilon \). Consider any positive constant \( \delta > 0 \) and any \( \epsilon \leq g(\delta) \). According to the definition, for any \( x \in L(\epsilon) \), there exists a \( y \in \mathcal{R} \) such that \( \|x - y\| \leq \sqrt{n} \epsilon \leq g(\delta) \).

Accordingly, \( |f(x) - f(y)| \leq \epsilon \). Hence, we have \( h(\epsilon) = \min_{x \in L(\epsilon)} f(x) \geq \inf_{x \in \mathcal{R}} f(x) - \epsilon \). On the other hand, we know \( h(\epsilon) \leq \inf_{x \in \mathcal{R}} f(x) \). This proves \( \lim_{\epsilon \to 0} h(\epsilon) = \inf_{x \in \mathcal{R}} f(x) \). Noticing \( \lim_{\epsilon \to 0} h(\epsilon) = \inf_{x \in \mathcal{R}} f(x) \) and applying \[86\] proves the Lemma.

REFERENCES

[1] S. Oveis Gharan, A. Bayesteh, and A. K. Khandani, “On the diversity-multiplexing tradeoff in multiple-relay network,” Information Theory, IEEE Transactions on, vol. 55, pp. 5423–5444, December 2009.
[2] K. Azarian, H. El Gamal, and Ph. Schniter, “On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels,” Information Theory, IEEE Transactions on, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.
[3] Sh. Oveis Gharan, A. Bayesteh, and A. K. Khandani, “On diversity-multiplexing tradeoff in multiple-antennas multiple-relay networks,” University of Waterloo, Tech. Rep., July 2008, available online at http://cst.uwaterloo.ca/.
[4] ——, “Diversity-multiplexing tradeoff in multiple-relay network-part II: multiple-antenna networks,” in 46th Allerton Conference on Communication, Control, and Computing, September 2008.
[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: efficient protocols and outage behavior,” Information Theory, IEEE Transactions on, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
[6] M. Yukel and E. Erkip, “Cooperative wireless systems: a diversity-multiplexing tradeoff perspective,” Information Theory, IEEE Transactions on, Aug. 2006, under Review.

\[14\] It should be noted that \( \leq \) has the properties that i) \( A \leq B \) and \( B \leq C \), results in \( A \leq C \); and ii) \( A \leq A \).

\[15\] Such a mapping could be as follows: \( m(x) = \frac{x}{\epsilon} \).
[7] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, “A simple cooperative diversity method based on network path selection,” IEEE J. Select. Areas Commun., vol. 24, no. 3, pp. 659–672, March 2006.

[8] Sh. Yang and J.-C. Belfiore, “Towards the optimal amplify-and-forward cooperative diversity scheme,” Information Theory, IEEE Transactions on, vol. 53, pp. 3114–3126, Sept. 2007.

[9] Sh. Oveis Gharan, A. Bayesteh, and A. K. Khandani, “Asymptotic analysis of amplify and forward relaying in a parallel MIMO relay network,” Information Theory, IEEE Transactions on, 2006, revised, available online at http://cst.uwaterloo.ca/.

[10] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” Information Theory, IEEE Transactions on, vol. 51, pp. 3037–3063, Sept. 2005.

[11] T. M. Cover and A. El Gamal, “Capacity theorems for the relay channel,” Information Theory, IEEE Transactions on, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[12] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Approximate capacity of Gaussian relay networks,” in IEEE Int. Symp. Information Theory, 2008.

[13] Sh. Yang and J.-C. Belfiore, “Optimal space-time codes for the MIMO amplify-and-forward cooperative channel,” Information Theory, IEEE Transactions on, vol. 53, no. 2, pp. 647–663, 2007.

[14] M. Gastpar and M. Vetterli, “On the capacity of large Gaussian relay networks,” Information Theory, IEEE Transactions on, vol. 51, pp. 765–779, March 2005.

[15] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, “Capacity scaling laws in MIMO relay networks,” IEEE Trans. on Wireless Communications, vol. 5, no. 6, pp. 1433–1444, June 2006.

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

[17] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, “Fading relay channels: performance limits and space-time signal design,” IEEE J. Select. Areas Commun., vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[18] K. Sreeram, S. Birenjith and P. Vijay Kumar, “Multi-hop cooperative wireless networks: diversity multiplexing tradeoff and optimal code design,” February 2008, Available online at http://arxiv.org/pdf/0802.1888.

[19] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Wireless network information flow: a deterministic approach,” Information Theory, IEEE Transactions on, June 2009, submitted for Publication.

[20] A. M. Tulino and S. Verdu, Random Matrix Theory and Wireless Communications. Now Publishers Inc, 2004.

[21] T. M. Cover and J. A. Thomas, Elements of Information Theory. New york: Wiley, 1991.

[22] R. Diestel, Graph Theory. Springer, 2006.

[23] R. A. Horn and C. R. Johnson, Matrix Analysis. Cambridge University Press, 1985.

[24] A. Bayesteh and A. K. Khandani, “On the user selection for MIMO broadcast channels,” Information Theory, IEEE Transactions on, vol. 54, no. 3, pp. 1086–1107, 2008.