Multi-Pair Amplify-and-Forward Relaying with Very Large Antenna Arrays

Himal A. Suraweera, Hien Quoc Ngo, Trung Q. Duong, Chau Yuen and Erik G. Larsson

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Multi-Pair Amplify-and-Forward Relaying with Very Large Antenna Arrays

Himal A. Suraweera*, Hien Quoc Ngo†, Trung Q. Duong‡, Chau Yuen* and Erik G. Larsson†

*Singapore University of Technology and Design, Singapore (e-mail: {himalsuraweera, yuenchau}@sutd.edu.sg)
†Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden (e-mail: {nqhien, egl}@isy.liu.se)
‡Blekinge Institute of Technology, Karlskrona, Sweden (e-mail: quang.trung.duong@bth.se)

Abstract—We consider a multi-pair relay channel where multiple sources simultaneously communicate with destinations using a relay. Each source or destination has only a single antenna, while the relay is equipped with a very large antenna array. We investigate the power efficiency of this system when maximum ratio combining/maximal ratio transmission (MRC/MRT) or zero-forcing (ZF) processing is used at the relay. Using a very large array, the transmit power of each source or relay (or both) can be made inversely proportional to the number of relay antennas while maintaining a given quality-of-service. At the same time, the achievable sum rate can be increased by a factor of the number of source-destination pairs. We show that when the number of antennas grows to infinity, the asymptotic achievable rates of MRC/MRT and ZF are the same if we scale the power at the sources. Depending on the large scale fading effect, MRC/MRT can outperform ZF or vice versa if we scale the power at the relay.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) technology has now become an integral feature of many advanced communication systems. A cellular architecture with MIMO that has gained significant research interest is multi-user MIMO (MU-MIMO) in which an antenna array simultaneously serves a multiplicity of autonomous co-channel users/nodes [1]. While the current systems have limited number of antennas (e.g. the LTE standard allows for up to 8 antenna ports), MU systems having a large number of antennas at the base station (*very large MIMO*) have been advocated recently in [2]–[4]. Very large MU-MIMO systems can substantially reduce the interference with simple signal processing techniques and achieve increased reliability and throughput, and significant reduction in total transmitted power [5].

On the other hand, relaying has been extensively explored to provide expanded coverage and high throughput, especially at the cell edge [6]. However, inter-user interference can cause major performance degradation in MU relay systems [7]. As a result, a large body of performance analysis work on MU relay systems, e.g., [8]–[10] has mainly avoided interference slots by adopting spectrally inefficient policies such as orthogonalization of time/frequency. Another line of work on MU relay systems has considered deploying complicated precoder/decoder designs; e.g., in [11] and advanced joint

network coding and signal alignment techniques for the multi-pair two-way relay channel [12].

In this paper, we analyze the performance of a multi-pair relaying scenario where a group of K sources and K destinations communicate using a single relay equipped with N antennas, where $N \gg K$. To the best of our knowledge, there is no prior work that analyzes the effects of large antenna arrays on the performance of the considered relay system. In the first time slot, all K sources simultaneously transmit their signals to the relay. In the second time slot, a linearly transformed version of the received signal at the relay is forwarded to the K destinations. For this multi-pair relay channel, we study the achievable rate vs. power efficiency performance with (1) maximum ratio combining/maximal ratio transmission (MRC/MRT) and (2) zero-forcing (ZF) at the relay.

We show that when N is large, we can cut the transmit power at each source or/and the relay proportionally to $1/N$ with no performance degradation. The asymptotic achievable rates of MRC/MRT and ZF for $N \to \infty$ are derived for cases when the transmit power of each source or/and the relay is made inversely proportional to N. The results show that when the transmit power of each source scales as $1/N$ while keeping a fixed transmit power at the relay, the fast fading, interference from other sources, and noise at the destination disappear and hence, the system performance does not depend on the quality of the channel in the second hop. In contrast, for the case when the transmit power of each source is fixed and the transmit power of the relay is scaled down as $1/N$, the system performance does not depend on the channel quality of the first hop. We further show that when the transmit power at the relay is cut proportionally to $1/N$, with very large N, depending on the large-scale fading effect, MRC/MRT performs better than ZF or vice versa.

Notation: †, $|| \cdot ||$ and $\mathsf{Tr}\,(\cdot)$ denote the conjugate transpose operation, Euclidean norm and the trace of a matrix respectively. $\mathbb{E}\{x\}$ stands for the expectation of a random variable x, and $\boldsymbol{I}_M$ is the identity matrix of size M. $\overset{a.s.}{\to}$ and $\overset{d}{\to}$ denote the almost sure convergence and the convergence in distribution, respectively.

II. SYSTEM MODEL

Consider a scenario with a group of K sources, S_k, K destinations, D_k, for $k = 1, \ldots, K$, and a single relay, R.
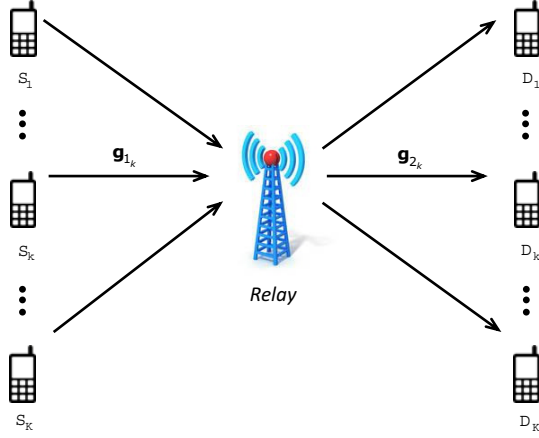
Fig. 1. System model. The channel strengths of the source-relay, $(S_k - R)$ and relay-destination, $(R - D_k)$ links for $k = 1, \ldots, K$ are η_{1k} and η_{2k}, respectively.

Each source/destination is equipped with a single antenna while R is equipped with N antennas as shown in Fig. 1. The source S_k wants to communicate with the destination D_k. Communication in this network occurs via R since there are no direct links among S_k and D_k, for any k due to heavy shadowing and path loss phenomenon.

During the first phase all sources simultaneously transmit their symbols to R, and the received $N \times 1$ signal vector can be written as

$$\boldsymbol{y}_R = \sqrt{P_{\mathsf{t}}} \boldsymbol{G}_1 \boldsymbol{x} + \boldsymbol{n}_R, \tag{1}$$

where $\sqrt{P_{\mathsf{t}}} \boldsymbol{x} = \sqrt{P_{\mathsf{t}}} [x_1, x_2, \ldots, x_K]^T$ are transmitted symbols with $\mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^\dagger\} = \boldsymbol{I}_K$ (the average transmitted power of each source is P_{t}) and $\boldsymbol{n}_R$ is an $N \times 1$ additive white Gaussian noise (AWGN) vector at the relay node with $\mathbb{E}\{\boldsymbol{n}_R \boldsymbol{n}_R^\dagger\} = \sigma_n^2 \boldsymbol{I}_N$. The $N \times K$ channel matrix between the K sources and R is expressed as $\boldsymbol{G}_1 = \boldsymbol{H}_1 \boldsymbol{D}_1^{1/2}$ where $\boldsymbol{H}_1$ contains independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ entries and $\boldsymbol{D}_1$ is a $K \times K$ diagonal matrix, where $[\boldsymbol{D}_1]_{kk} = \eta_{1k}$. Moreover, we model the $N \times K$ channel matrix between the K destinations and R as $\boldsymbol{G}_2 = \boldsymbol{H}_2 \boldsymbol{D}_2^{1/2}$ where $\boldsymbol{H}_2$ contains i.i.d. $\mathcal{CN}(0,1)$ entries and $\boldsymbol{D}_2$ is a $K \times K$ diagonal matrix, where $[\boldsymbol{D}_2]_{kk} = \eta_{2k}$. Note that $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ represent independent fast fading, while $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ represent path-loss attenuation, and log-normal shadow fading. The assumption of independent fast fading is sufficiently realistic for systems where the antennas are spaced sufficiently far apart [5].

During the second phase, R re-transmits a transformation of the received signal given by $\widetilde{\boldsymbol{y}}_R = \boldsymbol{W} \boldsymbol{y}_R$. The signal at D_k can be expressed as

$$y_{D_k} = \sqrt{P_{\mathsf{t}}} \boldsymbol{g}_{2_k}^\dagger \boldsymbol{W} \boldsymbol{g}_{1_k} x_k + \sqrt{P_{\mathsf{t}}} \sum_{i=1, i \neq k}^{K} \boldsymbol{g}_{2_k}^\dagger \boldsymbol{W} \boldsymbol{g}_{1_i} x_i$$
$$+ \boldsymbol{g}_{2_k}^\dagger \boldsymbol{W} \boldsymbol{n}_R + n_{D_k}, \tag{2}$$

where $\boldsymbol{g}_{1_i}$ is the i-th column of $\boldsymbol{G}_1$, $\boldsymbol{g}_{2_k}$ is the k-th column of $\boldsymbol{G}_2$, $\boldsymbol{W}$ is the $N \times N$ transformation matrix normal-

ized to satisfy a total power constraint, P_{r}, at the relay as $\mathsf{Tr}\left(\mathsf{E}\left\{ \widetilde{\boldsymbol{y}}_R \widetilde{\boldsymbol{y}}_R^\dagger \right\} \right) = P_{\mathsf{r}}$, and n_{D_k} is the AWGN at D_k with $\mathsf{E}\{n_{D_k} n_{D_k}^\dagger\} = \sigma_n^2$.

As a result, the instantaneous end-to-end (e2e) signal-to-interference-noise ratio (SINR) at D_k can be written as

$$\gamma_k = \frac{P_{\mathsf{t}} |\boldsymbol{g}_{2_k}^\dagger \boldsymbol{W} \boldsymbol{g}_{1_k}|^2}{P_{\mathsf{t}} \sum_{i=1, i \neq k}^{K} |\boldsymbol{g}_{2_k}^\dagger \boldsymbol{W} \boldsymbol{g}_{1_i}|^2 + \|\boldsymbol{g}_{2_k}^\dagger \boldsymbol{W}\|^2 \sigma_n^2 + \sigma_n^2}. \tag{3}$$

A. MRC/MRT at the Relay

When CSI is available at R, it is natural to apply a transformation based on the MRC/MRT principle[1]. In the first phase, the relay uses MRC to combine the signals transmitted from K sources and then in the second phase, it uses MRT precoding to forward data to K destinations. Hence the relay transformation matrix is given by $\boldsymbol{W} = a_{\mathsf{mrc}} \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger$. In this case, to meet the power constraint at the relay, we have

$$a_{\mathsf{mrc}} = \sqrt{\frac{P_{\mathsf{r}}}{\mathsf{Tr}\left(P_{\mathsf{t}} \left(\boldsymbol{G}_1^\dagger \boldsymbol{G}_1 \right)^2 \boldsymbol{G}_2^\dagger \boldsymbol{G}_2 + \sigma_n^2 \boldsymbol{G}_1^\dagger \boldsymbol{G}_1 \boldsymbol{G}_2^\dagger \boldsymbol{G}_2 \right)}}. \tag{4}$$

From (2), the received signal at D_k for MRC/MRT at the relay is given by

$$y_{D_k} = a_{\mathsf{mrc}} \sqrt{P_{\mathsf{t}}} \boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_k} x_k + a_{\mathsf{mrc}} \sqrt{P_{\mathsf{t}}} \sum_{i=1, i \neq k}^{K} \boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_i} x_i$$
$$+ a_{\mathsf{mrc}} \boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{n}_R + n_{D_k}. \tag{5}$$

Hence, the e2e SINR can be expressed as

$$\gamma_k^{\mathsf{mrc}} = \frac{P_{\mathsf{t}} \frac{|\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_k}|^2}{\|\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger\|^2}}{P_{\mathsf{t}} \sum_{i=1, i \neq k}^{K} \frac{|\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_i}|^2}{\|\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger\|^2} + \sigma_n^2 + \frac{\sigma_n^2}{a_{\mathsf{mrc}}^2 \|\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger\|^2}}. \tag{6}$$

B. ZF at the Relay

We now consider the use of ZF receivers and precoders at the relay. With ZF processing, the transformation matrix can be expressed as $\boldsymbol{W} = a_{\mathsf{zf}} \boldsymbol{G}_2 \left(\boldsymbol{G}_2^\dagger \boldsymbol{G}_2 \right)^{-1} \left(\boldsymbol{G}_1^\dagger \boldsymbol{G}_1 \right)^{-1} \boldsymbol{G}_1^\dagger$, where a_{zf} is chosen to satisfy the power constraint at the relay[2], i.e.,

$$a_{\mathsf{zf}} = \sqrt{\frac{P_{\mathsf{r}}}{\mathsf{Tr}\left(P_{\mathsf{t}} \left(\boldsymbol{G}_2^\dagger \boldsymbol{G}_2 \right)^{-1} + \sigma_n^2 \left(\boldsymbol{G}_2^\dagger \boldsymbol{G}_2 \right)^{-1} \left(\boldsymbol{G}_1^\dagger \boldsymbol{G}_1 \right)^{-1} \right)}}. \tag{7}$$

---

[1]Note that the choice of $\boldsymbol{W}$ based on the MRC/MRT principle is not optimal for maximizing the SINR. However, finding the optimal $\boldsymbol{W}$ in an analytical form seems impossible due to the non-convex nature of the problem. Furthermore, with very large antennas arrays, the channel vectors are nearly orthogonal, and hence MRC/MRT is nearly optimal [2].

[2]Reference [13, Sec. IV] also considers ZF at the relay and employs a *fixed gain* for long term power normalization. In contrast, we consider a variable gain in (7). Since for receive filtering/transmit ZF precoding at the relay, *instantaneous* channel information must be used and a fixed gain can result in high peak-to-average power ratio signals, the gain in (7) is a good practical choice when implementing ZF at the relay.

For this case, we have

$$\boldsymbol{g}_{2_k}^\dagger \boldsymbol{W} \boldsymbol{g}_{1_i} = a_{\mathsf{zf}} \delta_{ki}, \tag{8}$$

where $\delta_{ki} = 1$ when $k = i$ and 0 otherwise. Therefore, from (2), we can write the received signal at D_k as

$$y_{D_k} = a_{\mathsf{zf}} \sqrt{P_{\mathsf{t}}} x_k + a_{\mathsf{zf}} \left[\left(\boldsymbol{G}_1^\dagger \boldsymbol{G}_1 \right)^{-1} \boldsymbol{G}_1^\dagger \right]_k \boldsymbol{n}_R + n_{D_k}, \tag{9}$$

where $[\boldsymbol{A}]_k$ is the k-th row of the matrix $\boldsymbol{A}$.

Now we can express the e2e signal-to-noise ratio (SNR) as

$$\gamma_k^{\mathsf{zf}} = \frac{a_{\mathsf{zf}}^2 P_{\mathsf{t}}}{a_{\mathsf{zf}}^2 \left[\left(\boldsymbol{G}_1^\dagger \boldsymbol{G}_1 \right)^{-1} \right]_{kk} \sigma_n^2 + \sigma_n^2}. \tag{10}$$

C. Orthogonal Scheme

For comparison with MRC/MRT and ZF, we also consider a "*naive scheme*" that employs orthogonal channel access. Specifically, to completely avoid the inter-user interference, each $S_k - D_k$ pair for $k = 1, \ldots, K$ uses $\frac{1}{2K}$ channel resources for communication. At the relay, MRC/MRT is employed as it maximizes the e2e SNR[3]. Therefore, the e2e SNR at D_k can be expressed as

$$\gamma_k^{\mathsf{ns}} = \frac{\frac{P_{\mathsf{t}} \|\boldsymbol{g}_{1_k}^\dagger\|^2}{\sigma_n^2} \frac{P_{\mathsf{r}} \|\boldsymbol{g}_{2_k}\|^2}{\sigma_n^2}}{\frac{P_{\mathsf{t}} \|\boldsymbol{g}_{1_k}^\dagger\|^2}{\sigma_n^2} + \frac{P_{\mathsf{r}} \|\boldsymbol{g}_{2_k}^\dagger\|^2}{\sigma_n^2} + 1}. \tag{11}$$

III. LARGE N ANALYSIS

In this section, we further simplify the e2e SINR expressions (6) and (10) in the very large N regime. These new expressions illuminate several aspects of the achievable rate vs. power efficiency performance in the considered network. Here we assume that $N \gg K$. We further assume that when N is large, the elements of the channel matrices $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are still independent. Note that even with very large N, the physical size of the antenna array can be small. For example, at 2.6 GHz, a cylindrical array with 128 antennas and $\lambda/2$ antenna spacing occupies only a physical size of 28 cm × 29 cm and even with this array, the antennas experience nearly independent fading [3], [14].

The achievable ergodic sum rate of the system is given by[4]

$$C_{\mathsf{sum}}^\star = \mathbb{E} \left\{ \sum_{k=1}^K \frac{1}{2\alpha_{\mathsf{f}}} \log_2 \left(1 + \gamma_k^\star \right) \right\}, \tag{12}$$

where $\star = \{\mathsf{mrc}, \mathsf{zf}, \mathsf{ns}\}$ refers to MRC/MRT, ZF and naive schemes and the pre-log factor $\frac{1}{2}$ is due to the half-duplex relaying. For MRC/MRT and ZF: $\alpha_{\mathsf{f}} = 1$ and for the naive scheme $\alpha_{\mathsf{f}} = K$.

---

[3]In the remainder of the paper, "MRC/MRT" is used to refer to the scheme (cf. Section II-A) where $K > 1$ destinations are simultaneously served.

[4]The *exact* analysis of $C_{\mathsf{sum}}^{\mathsf{mrc}}$ and $C_{\mathsf{sum}}^{\mathsf{zf}}$ for *arbitrary* N is not a mathematically tractable problem since the required probability density functions (p.d.f.s) of γ_k^{mrc} and γ_k^{zf} do not readily permit mathematical manipulation. A closed-form expression $C_{\mathsf{sum}}^{\mathsf{ns}}$ can be derived, but not reported since our main focus in this paper is to analyze the impact of the very large array, where $N \gg K$.

In the following analysis, we will consider three cases: namely, *Case I) Fixed* $NP_{\mathsf{t}}, N \to \infty$; *Case II) Fixed* $NP_{\mathsf{r}}, N \to \infty$; *Case III) Fixed* NP_{t}, *Fixed* $NP_{\mathsf{r}}, N \to \infty$.

A. MRC/MRT at the Relay

Case I): If $P_{\mathsf{t}} = \frac{E_{\mathsf{t}}}{N}$ where E_{t} is fixed, then from (5) we have

$$\frac{y_{D_k}}{\sqrt{N}} = \frac{a_{\mathsf{mrc}} \sqrt{E_{\mathsf{t}}} \boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_k} x_k}{N} + \sum_{i=1, i \neq k}^K \frac{a_{\mathsf{mrc}} \sqrt{E_{\mathsf{t}}} \boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_i} x_i}{N}$$
$$+ \frac{a_{\mathsf{mrc}} \boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{n}_R}{\sqrt{N}} + \frac{n_{D_k}}{\sqrt{N}}. \tag{13}$$

In the very large N regime, we apply the law of large numbers given by [15]

$$\frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{g}_{2_i}}{N} \underset{N \to \infty}{\overset{a.s.}{\longrightarrow}} \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases} \tag{14}$$

and note that

$$Na_{\mathsf{mrc}} = \sqrt{\frac{P_{\mathsf{r}}}{\mathsf{Tr}\left(\frac{E_{\mathsf{t}}}{N^3} \left(\boldsymbol{G}_1^\dagger \boldsymbol{G}_1 \right)^2 \boldsymbol{G}_2^\dagger \boldsymbol{G}_2 + \sigma_n^2 \frac{1}{N^2} \boldsymbol{G}_1^\dagger \boldsymbol{G}_1 \boldsymbol{G}_2^\dagger \boldsymbol{G}_2 \right)}}$$

$$\underset{N \to \infty}{\overset{a.s.}{\longrightarrow}} \sqrt{\frac{P_{\mathsf{r}}}{\mathsf{Tr}\left(E_{\mathsf{t}} \boldsymbol{D}_1^2 \boldsymbol{D}_2 + \sigma_n^2 \boldsymbol{D}_1 \boldsymbol{D}_2 \right)}}. \tag{15}$$

Now re-expressing the first term in (13) as

$$a_{\mathsf{mrc}} \sqrt{E_{\mathsf{t}}} \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_k}}{N} = Na_{\mathsf{mrc}} \sqrt{E_{\mathsf{t}}} \sum_{i=1}^K \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{g}_{2_k} \boldsymbol{g}_{1_k}^\dagger \boldsymbol{g}_{1_k}}{N^2}.$$

Therefore, when N tends to infinity, we have

$$a_{\mathsf{mrc}} \sqrt{E_{\mathsf{t}}} \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_k}}{N} \overset{a.s.}{\longrightarrow} \sqrt{\frac{P_{\mathsf{r}} E_{\mathsf{t}}}{\mathsf{Tr}\left(E_{\mathsf{t}} \boldsymbol{D}_1^2 \boldsymbol{D}_2 + \sigma_n^2 \boldsymbol{D}_1 \boldsymbol{D}_2 \right)}} \eta_{1k} \eta_{2k}. \tag{16}$$

Similarly, for $i \neq k$, re-expressing the second term in (13) as

$$a_{\mathsf{mrc}} \sqrt{E_{\mathsf{t}}} \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_i}}{N} = Na_{\mathsf{mrc}} \sqrt{E_{\mathsf{t}}} \sum_{j=1}^K \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{g}_{2_j} \boldsymbol{g}_{1_j}^\dagger \boldsymbol{g}_{1_i}}{N^2},$$

we obtain

$$a\sqrt{E_{\mathsf{t}}} \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_i}}{N^2} \to 0. \tag{17}$$

Note that the third term in (13) can be written as

$$a_{\mathsf{mrc}} \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{n}_R}{\sqrt{N}} = Na_{\mathsf{mrc}} \sum_{i=1}^K \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{g}_{2_i} \boldsymbol{g}_{1_i}^\dagger \boldsymbol{n}_R}{N\sqrt{N}}.$$

We apply the law of large numbers and the Lindeberg-Lévy central limit theorem and obtain[5]

$$a_{\mathsf{mrc}} \frac{\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_i}}{\sqrt{N}} \overset{d}{\to} \sqrt{\frac{P_{\mathsf{r}}}{\mathsf{Tr}\left(E_{\mathsf{t}} \boldsymbol{D}_1^2 \boldsymbol{D}_2 + \sigma_n^2 \boldsymbol{D}_1 \boldsymbol{D}_2 \right)}} \eta_{2k} \widetilde{n}_R, \tag{18}$$

---

[5]Lindeberg-Lévy central limit theorem: Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be $n \times 1$ vectors whose elements are i.i.d. random variables with zero mean and variances of σ_p^2 and σ_q^2, respectively. Then $\frac{1}{\sqrt{n}} \boldsymbol{p}^H \boldsymbol{q} \overset{d}{\to} \mathcal{CN}\left(0, \sigma_p^2 \sigma_q^2 \right)$, as $n \to \infty$.

where $\widetilde{n}_R \sim \mathcal{CN}\left(0, \eta_{1k}\sigma_n^2\right)$. Substituting (16), (17) and (18) into (13) and since $\frac{1}{\sqrt{N}}n_{D_k} \to 0$ we have

$$\frac{y_{D_k}}{\sqrt{N}} \to \sqrt{\frac{P_r E_t}{\mathsf{Tr}\left(E_t \boldsymbol{D}_1^2 \boldsymbol{D}_2 + \sigma_n^2 \boldsymbol{D}_1 \boldsymbol{D}_2\right)}}\eta_{1k}\eta_{2k}x_k$$
$$+ \sqrt{\frac{P_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_1^2 \boldsymbol{D}_2 + \sigma_n^2 \boldsymbol{D}_1 \boldsymbol{D}_2\right)}}\eta_{2k}\widetilde{n}_R. \quad (19)$$

Now from (19) we obtain

$$\gamma_k^{\mathsf{mrc}} \to \frac{E_t \eta_{1k}}{\sigma_n^2}, \text{ as } N \to \infty. \quad (20)$$

For $K = 1$, it is clear that we can reduce the transmit power by a factor of $1/N$ with no reduction in performance due to the array gain. But here we consider multiple sources, and the above result implies that by using a large number of relay antennas, we can still obtain the same array gain as in the case of single source. Interestingly, from (20) we can see that when N grows large and NP_t is fixed, the e2e SNR does not depend on the transmit power at the relay and the large-scale fading of the second hop. This is due to the fact that, with MRT precoding at the relay in the second phase and P_r is fixed, as N goes to infinity, the effect of inter-user interference and noise at D_k disappears. Finally, the sum rate follows directly by substituting (20) into (12).

Case II): If $P_r = \frac{E_r}{N}$ where E_r is fixed, we observe that

$$N^2 a_{\mathsf{mrc}} \xrightarrow[N\to\infty]{a.s.} \sqrt{\frac{E_r}{\mathsf{Tr}\left(P_t \boldsymbol{D}_1^2 \boldsymbol{D}_2\right)}}. \quad (21)$$

Therefore when N grows without bound, the first term in (5) tends to

$$a_{\mathsf{mrc}}\sqrt{P_t}\boldsymbol{g}_{2_k}^\dagger \boldsymbol{G}_2 \boldsymbol{G}_1^\dagger \boldsymbol{g}_{1_k} \xrightarrow{a.s.} \sqrt{\frac{E_r}{\mathsf{Tr}\left(\boldsymbol{D}_1^2 \boldsymbol{D}_2\right)}}\eta_{1k}\eta_{2k}. \quad (22)$$

Similarly, when N goes to infinity, the second and third terms in (5) converge almost sure to 0. Therefore,

$$y_{D_k} \xrightarrow{a.s.} \sqrt{\frac{E_r}{\mathsf{Tr}\left(\boldsymbol{D}_1^2 \boldsymbol{D}_2\right)}}\eta_{1k}\eta_{2k}x_k + n_{D_k}, \text{ as } N \to \infty. \quad (23)$$

From (23), as $N \to \infty$, we have

$$\gamma_k^{\mathsf{mrc}} \to \frac{E_r \eta_{1k}^2 \eta_{2k}^2}{\mathsf{Tr}\left(\boldsymbol{D}_1^2 \boldsymbol{D}_2\right)\sigma_n^2}. \quad (24)$$

The above result shows that the transmit power at relay can be made inversely proportional to the number of relay antennas without compromising the quality-of-service. Note that in the special case where $\eta_{1k} = \eta_1$ and $\eta_{2k} = \eta_2$ for $k = 1, \ldots, K$, we have $\gamma_k^{\mathsf{mrc}} \to \frac{1}{K}\frac{E_r \eta_2}{\sigma_n^2}$ which does not depend on the transmit power of each source and the channel quality of the first hop.

Case III): If $P_t = \frac{E_t}{N}$ and $P_r = \frac{E_r}{N}$ where E_t and E_r are fixed, using a similar approach as above we can show that

$$y_{D_k} \xrightarrow[n\to\infty]{d} \sqrt{\frac{E_t E_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_1^2 \boldsymbol{D}_2 + \sigma_n^2 \boldsymbol{D}_1 \boldsymbol{D}_2\right)}}\eta_{1k}\eta_{2k}x_k$$
$$+ \sqrt{\frac{E_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_1^2 \boldsymbol{D}_2 + \sigma_n^2 \boldsymbol{D}_1 \boldsymbol{D}_2\right)}}\eta_{2k}\widetilde{n}_R + n_{D_k}. \quad (25)$$

Therefore,

$$\gamma_k^{\mathsf{mrc}} \to \frac{\frac{E_t \eta_{1k}}{\sigma_n^2}}{1 + \frac{\mathsf{Tr}\left(E_t \boldsymbol{D}_1^2 \boldsymbol{D}_2 + \sigma_n^2 \boldsymbol{D}_1 \boldsymbol{D}_2\right)}{E_r \eta_{1k}\eta_{2k}^2}}, \text{ as } N \to \infty. \quad (26)$$

Interestingly, by using large antenna arrays, we can scale down both transmit powers of source and relay nodes by a factor of $1/N$ with no reduction in performance. Furthermore, we can see that when $E_t \to \infty$, the above SINR coincides with the result for the case P_r is fixed, and when $E_r \to \infty$, the above SINR coincides with the result for the case P_t is fixed.

In the special case where $\eta_{1k} = \eta_1$ and $\eta_{2k} = \eta_2$ for $k = 1, \ldots, K$; we have $\gamma_k^{\mathsf{mrc}} \to \frac{\frac{E_t \eta_1}{\sigma_n^2}}{1 + \frac{K}{\frac{E_r \eta_2}{\sigma_n^2}}\left(1 + \frac{E_t \eta_1}{\sigma_n^2}\right)}$.

B. ZF at the Relay

By following a similar derivation as in the case of MRC/MRT, we can obtain the same power scaling law as follows.

Case I): In the very large N regime we first note that

$$\frac{a_{\mathsf{zf}}}{N} \xrightarrow{a.s.} \sqrt{\frac{P_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_2^{-1} + \sigma_n^2 \left(\boldsymbol{D}_1 \boldsymbol{D}_2\right)^{-1}\right)}}. \quad (27)$$

Then from (9) we have

$$\frac{y_{D_k}}{\sqrt{N}} \xrightarrow[N\to\infty]{d} \sqrt{\frac{P_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_2^{-1} + \sigma_n^2 \left(\boldsymbol{D}_1 \boldsymbol{D}_2\right)^{-1}\right)}}x_k$$
$$+ \sqrt{\frac{P_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_2^{-1} + \sigma_n^2 \left(\boldsymbol{D}_1 \boldsymbol{D}_2\right)^{-1}\right)\eta_{1k}^2}}\widetilde{n}_R + \frac{n_{D_k}}{\sqrt{N}}, \quad (28)$$

where $\widetilde{n}_R \sim \mathcal{CN}\left(0, \eta_{1k}\sigma_n^2\right)$. Hence, when N grows without bound, we obtain

$$\gamma_k^{\mathsf{zf}} \xrightarrow{a.s.} \frac{E_t \eta_{1k}}{\sigma_n^2}. \quad (29)$$

Case II): In this case for very large N, a_{zf} tends to

$$a_{\mathsf{zf}} \xrightarrow{a.s.} \sqrt{\frac{E_r}{\mathsf{Tr}\left(P_t \boldsymbol{D}_2^{-1}\right)}}. \quad (30)$$

Therefore, we can write (9) as

$$y_{D_k} \to \sqrt{\frac{E_r}{\mathsf{Tr}\left(P_t \boldsymbol{D}_2^{-1}\right)}}\sqrt{P_t}x_k + n_{D_k}, \text{ as } N \to \infty. \quad (31)$$

which leads to

$$\gamma_k^{\mathsf{zf}} \to \frac{E_r}{\mathsf{Tr}\left(\boldsymbol{D}_2^{-1}\right)\sigma_n^2}, \quad \text{as } N \to \infty. \tag{32}$$

In the special case when $\eta_{2k} = \eta_2$ for $k = 1, \ldots, K$, we have $\gamma_k \to \frac{1}{K} \frac{E_r \eta_2}{\sigma_n^2}$.

Case III): If $P_t = \frac{E_t}{N}$ and $P_r = \frac{E_r}{N}$ where E_t and E_r are fixed, when N tends to infinity, we obtain

$$\frac{a_{\mathsf{zf}}}{\sqrt{N}} \xrightarrow{a.s.} \sqrt{\frac{E_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_2^{-1} + \sigma_n^2 \left(\boldsymbol{D}_1 \boldsymbol{D}_2\right)^{-1}\right)}}. \tag{33}$$

Therefore,

$$y_{D_k} \xrightarrow{d} \sqrt{\frac{E_t E_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_2^{-1} + \sigma_n^2 \left(\boldsymbol{D}_1 \boldsymbol{D}_2\right)^{-1}\right)}} x_k \tag{34}$$

$$+ \sqrt{\frac{E_r}{\mathsf{Tr}\left(E_t \boldsymbol{D}_2^{-1} + \sigma_n^2 \left(\boldsymbol{D}_1 \boldsymbol{D}_2\right)^{-1}\right) \eta_{1k}^2}} \widetilde{n}_R + n_{D_k},$$

where $\widetilde{n}_R \sim \mathcal{CN}\left(0, \eta_{1k}\sigma_n^2\right)$. Hence, when N goes to infinity, we obtain

$$\gamma_k^{\mathsf{zf}} \to \frac{\frac{E_t \eta_{1k}}{\sigma_n^2}}{1 + \frac{\mathsf{Tr}\left(E_t \boldsymbol{D}_2^{-1} + \sigma_n^2 \left(\boldsymbol{D}_1 \boldsymbol{D}_2\right)^{-1}\right)\eta_{1k}}{E_r}}. \tag{35}$$

In the special case where $\eta_{1k} = \eta_1$ and $\eta_{2k} = \eta_2$ for $k = 1, \ldots, K$, we have $\gamma_k^{\mathsf{zf}} \to \frac{\frac{E_t \eta_1}{\sigma_n^2}}{1 + \frac{K}{\frac{E_r \eta_2}{\sigma_n^2}}\left(1 + \frac{E_t \eta_1}{\sigma_n^2}\right)}$.

Remark 1: In *Case I*) the asymptotic rate ($N \to \infty$) of D_k with MRC/MRT and ZF processing is same and given by $C_k^{\mathsf{mrc}} = C_k^{\mathsf{zf}} = \frac{1}{2}\log_2\left(1 + \frac{E_t \eta_{1k}}{\sigma_n^2}\right)$.

Remark 2: In *Case II*) the asymptotic rates at D_k with MRC/MRT and ZF, $C_k^{\mathsf{mrc}} \lessgtr C_k^{\mathsf{zf}}$, is determined by

$$\frac{1}{\eta_{1k}^2 \eta_{2k}^2} \sum_{i \neq k}^{K} \eta_{1i}^2 \eta_{2i} \gtrless \sum_{i \neq k}^{K} \frac{1}{\eta_{2i}}. \tag{36}$$

Remark 3: In *Case III*) the asymptotic rates at the kth destination with MRC/MRT and ZF, $C_k^{\mathsf{mrc}} \lessgtr C_k^{\mathsf{zf}}$, is determined by

$$\frac{1}{\eta_{1k}^2 \eta_{2k}^2} \sum_{i \neq k}^{K} \eta_{1i} \eta_{2i} \left(1 + \frac{E_t \eta_{1i}}{\sigma_n^2}\right) \gtrless \sum_{i \neq k}^{K} \frac{1}{\eta_{1i} \eta_{2i}} \left(1 + \frac{E_t \eta_{1i}}{\sigma_n^2}\right). \tag{37}$$

C. Orthogonal Scheme

We now consider the asymptotic sum rate of the orthogonal scheme. In *Case I*): It is easy to show that $\gamma_k^{\mathsf{ns}} \to \frac{E_t \eta_{1k}}{\sigma_n^2}$ and $C_{\mathsf{sum}}^{\mathsf{ns}} = \frac{1}{2K}\sum_{i=1}^{K}\log_2\left(1 + \frac{E_t \eta_{1k}}{\sigma_n^2}\right)$. In *Case II*): $\gamma_k^{\mathsf{ns}} \to \frac{E_r \eta_{2k}}{\sigma_n^2}$ and $C_{\mathsf{sum}}^{\mathsf{ns}} = \frac{1}{2K}\sum_{i=1}^{K}\log_2\left(1 + \frac{E_r \eta_{2k}}{\sigma_n^2}\right)$ and in *Case III*): We have $\gamma_k^{\mathsf{ns}} \to \frac{\frac{E_t \eta_{1k}}{\sigma_n^2} \frac{E_r \eta_{2k}}{\sigma_n^2}}{\frac{E_t \eta_{1k}}{\sigma_n^2} + \frac{E_r \eta_{2k}}{\sigma_n^2} + 1}$ and
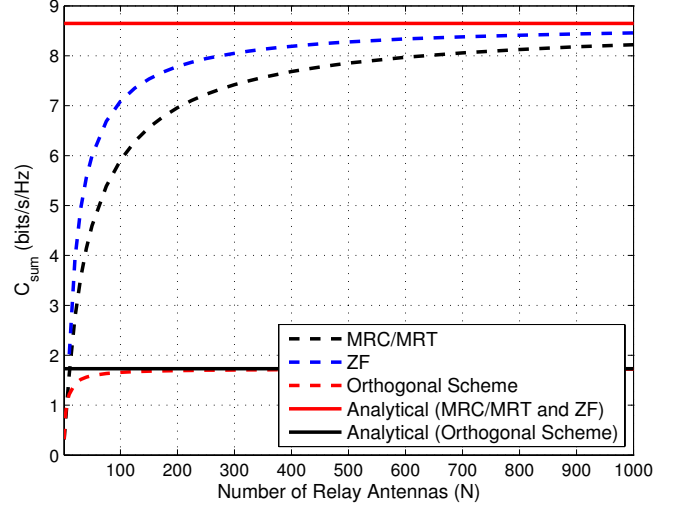
$$C_{\mathsf{sum}}^{\mathsf{ns}} = \frac{1}{2K} \sum_{i=1}^{K} \log_2\left(1 + \frac{\frac{E_t \eta_{1k}}{\sigma_n^2} \frac{E_r \eta_{2k}}{\sigma_n^2}}{\frac{E_t \eta_{1k}}{\sigma_n^2} + \frac{E_r \eta_{2k}}{\sigma_n^2} + 1}\right). \tag{38}$$

Fig. 2. *Case I*): Sum rate vs. the number of relay antennas. $E_t = 10$ dB, $P_r = 1$ and $K = 5$ users are served. $\boldsymbol{D}_1 = \boldsymbol{D}_2 = \boldsymbol{I}_5$.
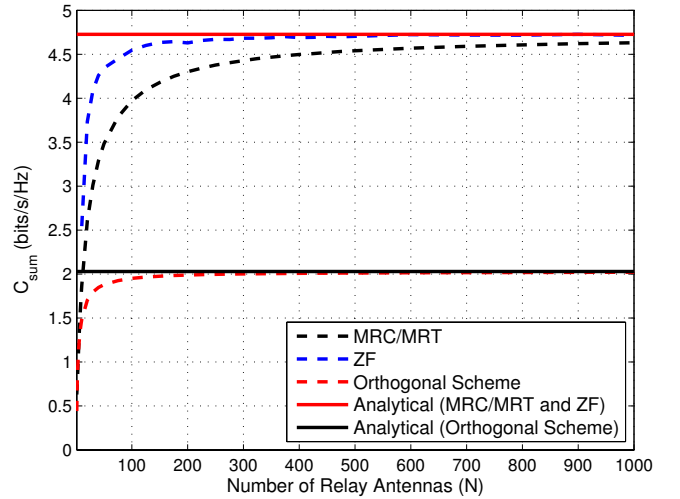

Fig. 3. *Case II*): Sum rate vs. the number of relay antennas. $P_t = 1$, $E_r = 10$ dB and $K = 5$ users are served. $\boldsymbol{D}_1 = \boldsymbol{D}_2 = \boldsymbol{I}_5$.

IV. NUMERICAL RESULTS

The sum rates achieved by the multi-pair relay system are evaluated through simulations and compared with our asymptotic analytical results. Without loss of generality, $\sigma_n^2 = 1$ is assumed.

Fig. 2 shows the simulated sum rate vs. the number of relay antennas and the presented analytical asymptotic results for *Case I*). Clearly, as the number of antennas increases, the sum rates of MRC/MRC, ZF and the naive schemes approaches the corresponding constant values predicted by our analysis. Interestingly, the sum rate curve of ZF has a sharper knee than the MRC/MRT counterpart on the way to the same asymptotic constant and the achieved sum rate is 8.65 bits/s/Hz. Though it is not explicitly seen in Fig. 2, for small number of antennas ($N \leq 6$), the naive scheme exhibits a better sum rate than MRC/MRT as expected. However, as N increases, the sum

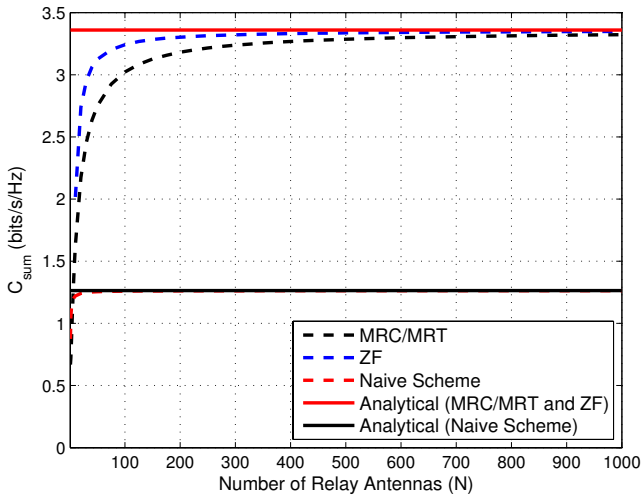Fig. 4. *Case III)*: Sum rate vs. the number of relay antennas. $P_t = 10$ dB, $E_r = 10$ dB and $K = 5$ users are served. $\boldsymbol{D}_1 = \boldsymbol{D}_2 = \boldsymbol{I}_5$.
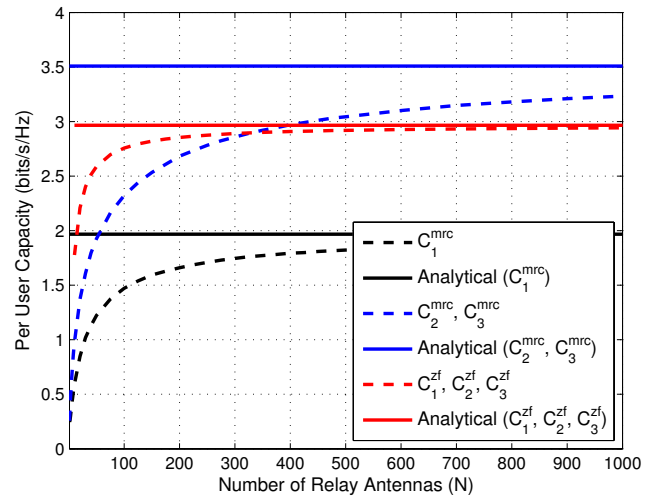

Fig. 5. *Case II)*: Per user rate vs. the number of relay antennas. $K = 3$ are served.

rate offered by the naive scheme rapidly saturates while the sum rates of the MRC/MRT and ZF schemes show a rapid improvement. This is because with a limited number of antennas, interference cannot be significantly reduced and thus lowers the sum rate of MRC/MRT. But when N grows large, the random channel vectors between sources/destinations and relay become pairwise orthogonal and hence, the interference is canceled out. At the same time, we gain from simultaneously serving K source-destination pairs in the same time-frequency resource.

Figs. 3 and 4 show results for the second and third power scaling laws; *Case II)* and *Case III)*. Both MRC/MRT and ZF achieves the same sum rate of 4.73 and 3.36 bits/s/Hz in *Case II)* and *Case III)*, respectively. Moreover, similar trends in results as in Fig. 2 can be observed.

The rates achieved by individual destinations are illustrated in Fig. 5 for slow-fading coefficients; $\eta_{11} = 2, \eta_{12} = 2, \eta_{13} = 2$ and $\eta_{21} = 1, \eta_{22} = 3, \eta_{23} = 3$ and *Case II)*. Results in Fig. 5 confirm that the sum rate of MRC/MRT can be higher than the sum rate of ZF depending on the slow fading parameters. $C_{sum}^{mrc} = 8.98$ and $C_{sum}^{zf} = 8.90$. Recall that $C_{sum}^{mrc} < C_{sum}^{zf}$ in the example of Fig. 3. Interestingly, when $N \leq 350$, all three users in the ZF system achieve a higher rate than the users in the MRC system. However, when N is very large, two users of the MRC achieve a higher rate than the ZF users.

V. CONCLUSION

We have shown that relay systems can benefit significantly from the use of very large antenna arrays. The offered sum rates of a multi-pair relay system was investigated for three different power scaling laws. At the relay, MRC/MRT and ZF processing was considered. We derived asymptotic sum rate results and confirmed their accuracy using computer simulations. Several insights were extracted using the analysis to illuminate the comparative performances between MRC/MRT and ZF. For example, the asymptotic achievable rates of MRC/MRT and ZF are the same if we scale the power at the sources.

REFERENCES

[1] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, pp. 36–46, Sept. 2007.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590-3600, Nov. 2010.

[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, pp. 40-60, Jan. 2013.

[4] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, Y. R. Yang, and L. Zhong, "Argos: Practical base stations with large-scale multi-user beamforming," in *Proc. ACM MobiCom 2012*, Istanbul, Turkey, Aug. 2012, pp. 53-64.

[5] H. Q. Ngo, E. G. Larsson and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, (accepted). [Online]. Available: http://arxiv.org/pdf/1112.3810v2.pdf

[6] M. Dohler and Y. Li, *Cooperative Communications: Hardware, Channel & PHY*. Wiley & Sons, 2010.

[7] A. Agustin and J. Vidal, "Amplify-and-forward cooperation under interference-limited spatial reuse of the relay slot," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1952-1962, May 2008.

[8] N. Yang, M. Elkashlan and J. Yuan, "Outage probability of multiuser relay networks in Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 2120-2132, June 2010.

[9] H. Ding, J. Ge, D. B. da Costa, and Z. Jiang, "A new efficient low-complexity scheme for multi-source multi-relay cooperative networks," *IEEE Trans. Veh. Technol.*, vol. 60, pp. 716-722, Feb. 2011.

[10] J. Kim, D. S. Michalopoulos, and R. Schober, "Diversity analysis of multi-user multi-relay networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2380-2389, July 2011.

[11] J. Cao, Z. Zhong and F. Wang, "Regenerative multi-way relaying: Relay precoding and ordered MMSE-SIC receiver," in *Proc. IEEE VTC Spring 2012*, Yokohama, Japan, May 2012, pp. 1-5.

[12] Z. Zhao, Z. Ding, M. Peng, W. Wang and K. K. Leung, "A special case of multi-way relay channel: When beamforming is not applicable," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2046-2051, July 2011.

[13] R. H. Y. Louie, Y. Li and B. Vucetic, "Zero forcing in general two-hop relay networks," in *IEEE Trans. Veh. Technol.*, vol. 59, pp. 191-202, Jan. 2010.

[14] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE VTC Fall 2011*, San Francisco, CA, Sept. 2011, pp. 1-5.

[15] H. Cramér, *Random Variables and Probability Distributions*, 3rd ed. Cambridge, UK: Cambridge University Press, 1970.