Multipair Massive MIMO Full-Duplex Relaying with MRC/MRT Processing

Hien Quoc Ngo*, Himal A. Suraweera†, Michail Matthaiou‡, and Erik G. Larsson*

*Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden (email: {nqhien, egl}@isy.liu.se)

†Department of Electrical & Electronics Engineering, University of Peradeniya, Peradeniya, Sri Lanka (email: himal@ee.pdn.ac.lk)

‡School of Electronics, Electrical Engineering & Computer Science, Queen's University Belfast, Belfast, U.K.

Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden (email: m.matthaiou@qub.ac.uk)

***Abstract*—We consider a multipair relay channel, where multiple sources communicate with multiple destinations with the help of a full-duplex (FD) relay station (RS). All sources and destinations have a single antenna, while the RS is equipped with massive arrays. We assume that the RS estimates the channels by using training sequences transmitted from sources and destinations. Then, it uses maximum-ratio combining/maximum-ratio transmission (MRC/MRT) to process the signals. To significantly reduce the loop interference (LI) effect, we propose two massive MIMO processing techniques: i) using a massive receive antenna array; or ii) using a massive transmit antenna array together with very low transmit power at the RS. We derive an exact achievable rate in closed-form and evaluate the system spectral efficiency. We show that, by doubling the number of antennas at the RS, the transmit power of each source and of the RS can be reduced by 1.5 dB if the pilot power is equal to the signal power and by 3 dB if the pilot power is kept fixed, while maintaining a given quality-of-service. Furthermore, we compare FD and half-duplex (HD) modes and show that FD improves significantly the performance when the LI level is low.**

I. Introduction

Massive MIMO is an emerging technology capable of scaling up conventional MIMO by orders of magnitude [1]–[3]. With simple signal processing techniques, such systems can achieve increased reliability, throughput, and substantial reduction in the total transmitted power. Due to these attractive benefits, massive MIMO combined with cooperative relaying is widely tipped as a key enabler for the development of future energy-efficient mobile networks [4].

On the other hand, FD relaying has attracted significant interest recently, thanks to its ability to eliminate the spectral efficiency penalty of HD relaying. With FD relaying, the relay receives and transmits simultaneously on the same channel [5]. As such, the FD mode utilizes the spectrum resources more efficiently. However, the main limitation in FD operation is the LI due to signal leakage from the relay's output to the input at the reception [6]. The mitigation of LI is an active area of FD research. To this end, progress has been made on both theory and hardware implementation to make FD wireless communication a viable practical solution [7]–[10].

Traditionally, LI suppression is achieved in the antenna/time domain through natural isolation and analog precancellation, which, however, may require sophisticated electronic implementation [7]. In addition, MIMO processing provides an effective means to suppress the LI in the spatial domain. With multiple transmit/receive antennas, precoding schemes can be deployed to mitigate the LI effects. For instance, [6] proposes to direct the LI of a FD decode-and-forward (DF) relay to the least harmful spatial dimensions. In [7], assuming a multiple antenna relay, a range of spatial suppression techniques, including precoding and antenna selection, is analyzed. More recently, [11] analyzed several antenna selection schemes for spatial LI suppression in a MIMO relay channel.

Different from the majority of existing works in the literature, which elaborate on systems that deploy only few antennas, in this paper we consider a massive MIMO FD relay architecture. The large number of spatial dimensions available in a massive MIMO system can be effectively used to suppress the LI in the spatial domain. We assume that a group of K sources communicate with a group of K destinations, with the aid of a massive MIMO FD relay station. More specifically, in this multipair system, we consider MRC/MRT with FD relay operation and investigate the achievable rate and power efficiency. We consider the MRC/MRT processing since it is a very simple signal processing scheme and, most importantly, works particularly well for massive MIMO [1]–[4].

We show that the LI can be significantly reduced, if the RS is equipped with a large receive antenna array or/and is equipped with a large transmit antenna array. At the same time, the inter-pair interference and noise effects disappear. Furthermore, when the number of RS transmit antennas, $N_{\rm tx}$, and the number of RS receive antennas, $N_{\rm rx}$, are large, we can scale down the transmit powers of each source and of the relay proportionally to $1/N_{\rm rx}$ and $1/N_{\rm tx}$, respectively, if the pilot power is kept fixed, and proportionally to $1/\sqrt{N_{\rm rx}}$ and $1/\sqrt{N_{\rm tx}}$, respectively, if the pilot power and the data power are the same. Moreover, we compare FD and HD modes and demonstrate which one yields better performance for different values of the LI level.

Notation: The superscripts $()^*$, $()^T$, and $()^H$ stand for the conjugate, transpose, and conjugate-transpose, respectively. The expectation and variance operations are denoted by $\mathbb{E}\left\{\cdot\right\}$ and $\mathbb{V}{\rm ar}\left(\cdot\right)$, respectively. Finally, we use $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}\right)$ to denote a circularly symmetric complex Gaussian vector $\mathbf{z}$ with zero mean and covariance matrix $\mathbf{\Sigma}$.

The work of H. Q. Ngo and E. G. Larsson was supported in part by the Swedish Research Council (VR), the Swedish Foundation for Strategic Research (SSF), and ELLIIT.

II. System Model

Figure 1 shows the considered multipair DF relaying system where K communication pairs $(\mathtt{S}_k, \mathtt{D}_k)$ share the same time-frequency resource and a common RS. The kth source, $\mathtt{S}_k$, wants to communicate with the kth destination, $\mathtt{D}_k$, via the RS. All source and destination nodes are equipped with a single antenna each, while the RS is equipped with $N_{\rm rx}$ receive antennas and $N_{\rm tx}$ transmit antennas for the FD operation. The total number of antennas at the RS is $N = N_{\rm rx} + N_{\rm tx}$. Without significant loss of generality, we assume that $N_{\rm rx}, N_{\rm tx} \geq K$. Further, the direct links among $\mathtt{S}_k$ and $\mathtt{D}_k$ do not exist due to large path loss and heavy shadowing. Our network configuration is of practical interest, for example, in a cellular setup, where inter-user communication can be realized with the help of a base station equipped with massive arrays.

At time instant i, all K sources $\mathtt{S}_k$, $k = 1, ..., K$, transmit their signals, $\sqrt{p_{\mathtt{S}}} x_k[i]$, to the RS, while the RS broadcasts $\sqrt{p_{\mathtt{R}}} \mathbf{s}[i] \in \mathbb{C}^{N_{\rm tx} \times 1}$ to all K destinations. Here, $p_{\mathtt{S}}$ and $p_{\mathtt{R}}$ are the average transmit powers of each source and of the RS. Since the RS receives and transmits signals at the same frequency, the received signal at the RS is interfered by its own transmitted signal, $\mathbf{s}[i]$. This is called *loop interference*. Denote by $\mathbf{x}[i] \triangleq [x_1[i] \; x_2[i] \; ... \; x_K[i]]^T$. The received signals at the RS and the K destinations are given by [7]

$$\mathbf{y}_{\mathtt{R}}[i] = \sqrt{p_{\mathtt{S}}} \mathbf{G}_{\mathtt{SR}} \mathbf{x}[i] + \sqrt{p_{\mathtt{R}}} \mathbf{G}_{\mathtt{RR}} \mathbf{s}[i] + \mathbf{n}_{\mathtt{R}}[i], \quad (1)$$

$$\mathbf{y}_{\mathtt{D}}[i] = \sqrt{p_{\mathtt{R}}} \mathbf{G}_{\mathtt{RD}}^T \mathbf{s}[i] + \mathbf{n}_{\mathtt{D}}[i], \quad (2)$$

respectively, where $\mathbf{G}_{\mathtt{SR}} \in \mathbb{C}^{N_{\rm rx} \times K}$ and $\mathbf{G}_{\mathtt{RD}}^T \in \mathbb{C}^{K \times N_{\rm tx}}$ are the channel matrices from the K sources to the RS receive antenna array and from the RS transmit antenna array to the K destinations, respectively. The channel matrices account for both small-scale fading and large-scale fading. More precisely, $\mathbf{G}_{\mathtt{SR}}$ and $\mathbf{G}_{\mathtt{RD}}$ can be expressed as $\mathbf{G}_{\mathtt{SR}} = \mathbf{H}_{\mathtt{SR}} \mathbf{D}_{\mathtt{SR}}^{1/2}$ and $\mathbf{G}_{\mathtt{RD}} = \mathbf{H}_{\mathtt{RD}} \mathbf{D}_{\mathtt{RD}}^{1/2}$, where the small-scale fading matrices $\mathbf{H}_{\mathtt{SR}}$ and $\mathbf{H}_{\mathtt{RD}}$ include independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ elements, while $\mathbf{D}_{\mathtt{SR}}$ and $\mathbf{D}_{\mathtt{RD}}$ are the large-scale fading diagonal matrices whose kth diagonal elements are denoted by $\beta_{\mathtt{SR},k}$ and $\beta_{\mathtt{RD},k}$, respectively; $\mathbf{G}_{\mathtt{RR}} \in \mathbb{C}^{N_{\rm rx} \times N_{\rm tx}}$ is the channel matrix between the transmit and receive arrays which represents the LI channel matrix. We model the LI channel via the Rayleigh fading distribution, under the assumptions that the line-of-sight component is efficiently reduced by the antenna isolation and the major effect comes from the scattering components. Note that if hardware LI cancellation (LIC) is applied, $\mathbf{G}_{\mathtt{RR}}$ represents the residual interference due to imperfect LIC [7]. Therefore, the elements of $\mathbf{G}_{\mathtt{RR}}$ can be modeled as i.i.d. $\mathcal{CN}(0, \sigma_{\mathtt{LI}}^2)$ random variables (RVs), where $\sigma_{\mathtt{LI}}^2$ can be understood as the level of LI effect, which depends on the distance between the transmit and receive antenna arrays or/and the capability of the hardware LIC technique. Here, we assume that the distance between the transmit array and the receive array is much larger than the inter-element distance, such that the channels between transmit and receive antennas are i.i.d.;[1] $\mathbf{n}_{\mathtt{R}}[i]$ and $\mathbf{n}_{\mathtt{D}}[i]$ are AWGN vectors at the RS and the K destinations. The elements of $\mathbf{n}_{\mathtt{R}}[i]$ and $\mathbf{n}_{\mathtt{D}}[i]$ are assumed to be i.i.d. $\mathcal{CN}(0,1)$ RVs.

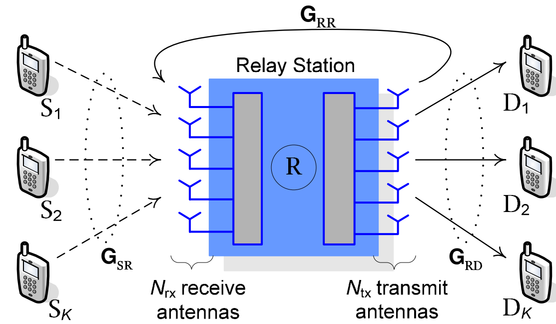

Fig. 1. Multipair full-duplex relaying system.

A. Channel Estimation

A part of coherence interval is used for channel estimation. All sources and destinations simultaneously transmit their pilot sequences of τ symbols to the RS. The received pilot at the RS receive and transmit antenna arrays are given by

$$\mathbf{Y}_{\rm rp} = \sqrt{\tau p_{\rm p}} \mathbf{G}_{\mathtt{SR}} \mathbf{\Phi}_{\mathtt{S}} + \sqrt{\tau p_{\rm p}} \bar{\mathbf{G}}_{\mathtt{RD}} \mathbf{\Phi}_{\mathtt{D}} + \mathbf{N}_{\rm rp}, \quad (3)$$

$$\mathbf{Y}_{\rm tp} = \sqrt{\tau p_{\rm p}} \bar{\mathbf{G}}_{\mathtt{SR}} \mathbf{\Phi}_{\mathtt{S}} + \sqrt{\tau p_{\rm p}} \mathbf{G}_{\mathtt{RD}} \mathbf{\Phi}_{\mathtt{D}} + \mathbf{N}_{\rm tp}, \quad (4)$$

respectively, where $\bar{\mathbf{G}}_{\mathtt{SR}} \in \mathbb{C}^{N_{\rm tx} \times K}$ and $\bar{\mathbf{G}}_{\mathtt{RD}} \in \mathbb{C}^{N_{\rm rx} \times K}$ are the channel matrices from the K sources to the RS transmit antenna array and from the K destinations to the RS receieve antenna array, respectively; $p_{\rm p}$ is the transmit power of each pilot symbol, $\mathbf{N}_{\rm rp}$ and $\mathbf{N}_{\rm tp}$ are the AWGN matrices which include i.i.d. $\mathcal{CN}(0,1)$ elements, while the kth rows of $\mathbf{\Phi}_{\mathtt{S}} \in \mathbb{C}^{K \times \tau}$ and $\mathbf{\Phi}_{\mathtt{D}} \in \mathbb{C}^{K \times \tau}$ are the pilot sequences transmitted from $\mathtt{S}_k$ and $\mathtt{D}_k$, respectively. All pilot sequences are assumed to be pairwisely orthogonal, i.e., $\mathbf{\Phi}_{\mathtt{S}} \mathbf{\Phi}_{\mathtt{S}}^H = \mathbf{I}_K$, $\mathbf{\Phi}_{\mathtt{D}} \mathbf{\Phi}_{\mathtt{D}}^H = \mathbf{I}_K$, and $\mathbf{\Phi}_{\mathtt{S}} \mathbf{\Phi}_{\mathtt{D}}^H = \mathbf{0}_K$. This requires that $\tau \geq 2K$.

The MMSE channel estimates of $\mathbf{G}_{\mathtt{SR}}$ and $\mathbf{G}_{\mathtt{RD}}$ are [13]

$$\hat{\mathbf{G}}_{\mathtt{SR}} = \frac{1}{\sqrt{\tau p_{\rm p}}} \mathbf{Y}_{\rm rp} \mathbf{\Phi}_{\mathtt{S}}^H \tilde{\mathbf{D}}_{\mathtt{SR}} = \mathbf{G}_{\mathtt{SR}} \tilde{\mathbf{D}}_{\mathtt{SR}} + \frac{1}{\sqrt{\tau p_{\rm p}}} \mathbf{N}_{\mathtt{S}} \tilde{\mathbf{D}}_{\mathtt{SR}}, \quad (5)$$

$$\hat{\mathbf{G}}_{\mathtt{RD}} = \frac{1}{\sqrt{\tau p_{\rm p}}} \mathbf{Y}_{\rm tp} \mathbf{\Phi}_{\mathtt{D}}^H \tilde{\mathbf{D}}_{\mathtt{RD}} = \mathbf{G}_{\mathtt{RD}} \tilde{\mathbf{D}}_{\mathtt{RD}} + \frac{1}{\sqrt{\tau p_{\rm p}}} \mathbf{N}_{\mathtt{D}} \tilde{\mathbf{D}}_{\mathtt{RD}}, \quad (6)$$

respectively, where $\tilde{\mathbf{D}}_{\mathtt{SR}} \triangleq \left(\frac{\mathbf{D}_{\mathtt{SR}}^{-1}}{\tau p_{\rm p}} + \mathbf{I}_K \right)^{-1}$, $\tilde{\mathbf{D}}_{\mathtt{RD}} \triangleq \left(\frac{\mathbf{D}_{\mathtt{RD}}^{-1}}{\tau p_{\rm p}} + \mathbf{I}_K \right)^{-1}$, $\mathbf{N}_{\mathtt{S}} \triangleq \mathbf{N}_{\rm rp} \mathbf{\Phi}_{\mathtt{S}}^H$ and $\mathbf{N}_{\mathtt{D}} \triangleq \mathbf{N}_{\rm tp} \mathbf{\Phi}_{\mathtt{D}}^H$. Since the rows of $\mathbf{\Phi}_{\mathtt{S}}$ and $\mathbf{\Phi}_{\mathtt{D}}$ are pairwisely orthogonal, the elements of $\mathbf{N}_{\mathtt{S}}$ and $\mathbf{N}_{\mathtt{D}}$ are i.i.d. $\mathcal{CN}(0,1)$ RVs. Let $\boldsymbol{\mathcal{E}}_{\mathtt{SR}}$ and $\boldsymbol{\mathcal{E}}_{\mathtt{RD}}$ be the estimation error matrices of $\mathbf{G}_{\mathtt{SR}}$ and $\mathbf{G}_{\mathtt{RD}}$, respectively. Then,

$$\mathbf{G}_{\mathtt{SR}} = \hat{\mathbf{G}}_{\mathtt{SR}} + \boldsymbol{\mathcal{E}}_{\mathtt{SR}}, \quad (7)$$

$$\mathbf{G}_{\mathtt{RD}} = \hat{\mathbf{G}}_{\mathtt{RD}} + \boldsymbol{\mathcal{E}}_{\mathtt{RD}}. \quad (8)$$

[1] For example, consider two transmit and receive arrays which are located on the two sides of a building with a distance of 5m. Assume that the system is operating at 2.6GHz. Then, to guarantee uncorrelation between the antennas, the distance between adjacent antennas is about 6cm, which is half a wavelength. Clearly, 5m $\gg$ 6cm. If each array is a cylindrical array with 128 antennas, the physical size of each array is about 28cm$\times$29cm [12] which is still relatively small compared to the distance between two arrays.

From the property of MMSE channel estimation, $\hat{\mathbf{G}}_{\mathtt{SR}}$, $\boldsymbol{\mathcal{E}}_{\mathtt{SR}}$, $\hat{\mathbf{G}}_{\mathtt{RD}}$, and $\boldsymbol{\mathcal{E}}_{\mathtt{RD}}$ are independent. Furthermore, we have that $\hat{\mathbf{G}}_{\mathtt{SR}} \sim \mathcal{CN}\left(\mathbf{0}, \hat{\mathbf{D}}_{\mathtt{SR}}\right)$, $\boldsymbol{\mathcal{E}}_{\mathtt{SR}} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{D}_{\mathtt{SR}} - \hat{\mathbf{D}}_{\mathtt{SR}}\right)$, $\hat{\mathbf{G}}_{\mathtt{RD}} \sim \mathcal{CN}\left(\mathbf{0}, \hat{\mathbf{D}}_{\mathtt{RD}}\right)$, and $\boldsymbol{\mathcal{E}}_{\mathtt{RD}} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{D}_{\mathtt{RD}} - \hat{\mathbf{D}}_{\mathtt{RD}}\right)$, where $\hat{\mathbf{D}}_{\mathtt{SR}}$ and $\hat{\mathbf{D}}_{\mathtt{RD}}$ are diagonal matrices whose kth elements are $\sigma^2_{\mathtt{SR},k} \triangleq \frac{\tau p_{\mathrm{p}} \beta^2_{\mathtt{SR},k}}{\tau p_{\mathrm{p}} \beta_{\mathtt{SR},k}+1}$ and $\sigma^2_{\mathtt{RD},k} \triangleq \frac{\tau p_{\mathrm{p}} \beta^2_{\mathtt{RD},k}}{\tau p_{\mathrm{p}} \beta_{\mathtt{RD},k}+1}$, respectively.

B. Data Transmission

The RS considers the channel estimates as the true channels and employs MRC/MRT processing. More precisely, the RS uses the MRC receiver to decode the signals transmitted from K sources. Simultaneously, it uses MRT precoding scheme to forward the signals to the K destinations.

1) MRC Receiver: With the MRC receiver, the received signal $\mathbf{y}_{\mathtt{R}}[i]$ is separated into K streams by multiplying it with a linear receiver matrix $\hat{\mathbf{G}}^H_{\mathtt{SR}}$ as follows:

$$\begin{aligned}\mathbf{r}[i] &= \hat{\mathbf{G}}^H_{\mathtt{SR}} \mathbf{y}_{\mathtt{R}}[i] \\ &= \sqrt{p_{\mathtt{S}}}\hat{\mathbf{G}}^H_{\mathtt{SR}}\mathbf{G}_{\mathtt{SR}}\mathbf{x}[i] + \sqrt{p_{\mathtt{R}}}\hat{\mathbf{G}}^H_{\mathtt{SR}}\mathbf{G}_{\mathtt{RR}}\mathbf{s}[i] + \hat{\mathbf{G}}^H_{\mathtt{SR}}\mathbf{n}_{\mathtt{R}}[i]. \quad (9)\end{aligned}$$

Then, the kth stream (the kth element of $\mathbf{r}[i]$) is used to decode the signal transmitted from $\mathtt{S}_k$. The kth element of $\mathbf{r}[i]$ can be expressed as

$$\begin{aligned}r_k[i] = &\underbrace{\sqrt{p_{\mathtt{S}}}\hat{\mathbf{g}}^H_{\mathtt{SR},k}\mathbf{g}_{\mathtt{SR},k}x_k[i]}_{\text{desired signal}} + \underbrace{\sqrt{p_{\mathtt{S}}}\sum_{j\neq k}^{K}\hat{\mathbf{g}}^H_{\mathtt{SR},k}\mathbf{g}_{\mathtt{SR},j}x_j[i]}_{\text{interpair interference}} \\ &+ \underbrace{\sqrt{p_{\mathtt{R}}}\hat{\mathbf{g}}^H_{\mathtt{SR},k}\mathbf{G}_{\mathtt{RR}}\mathbf{s}[i]}_{\text{loop interference}} + \underbrace{\hat{\mathbf{g}}^H_{\mathtt{SR},k}\mathbf{n}_{\mathtt{R}}[i]}_{\text{noise}}, \quad (10)\end{aligned}$$

where $\mathbf{g}_{\mathtt{SR},k}$ and $\hat{\mathbf{g}}_{\mathtt{SR},k}$ are the kth columns of $\mathbf{G}_{\mathtt{SR}}$ and $\hat{\mathbf{G}}_{\mathtt{SR}}$, respectively, and $x_k[i]$ is the kth element of $\mathbf{x}[i]$.

2) MRT Precoding: After detecting the signals transmitted from the K sources, the RS will use MRT precoding to process these signals before broadcasting to all K destinations. Owing to the processing delay [7], the transmit vector $\mathbf{s}[i]$ is a precoded version of $\mathbf{x}[i-d]$, where d is the processing delay. More precisely

$$\mathbf{s}[i] = \alpha\hat{\mathbf{G}}^*_{\mathtt{RD}}\mathbf{x}[i-d], \quad (11)$$

where α is the normalization constant, chosen to satisfy a long-term total transmit power constraint at the RS, i.e., $\mathtt{E}\left\{\|\mathbf{s}[i]\|^2\right\} = 1$. Thus, we have

$$\alpha = \sqrt{\frac{1}{N_{\mathrm{tx}}\sum_{k=1}^{K}\sigma^2_{\mathtt{RD},k}}}. \quad (12)$$

From (2) and (11), the received signal at $\mathtt{D}_k$ can be written as

$$\begin{aligned}y_{\mathtt{D},k}[i] &= \alpha\sqrt{p_{\mathtt{R}}}\mathbf{g}^T_{\mathtt{RD},k}\hat{\mathbf{g}}^*_{\mathtt{RD},k}x_k[i-d] \\ &+ \alpha\sqrt{p_{\mathtt{R}}}\sum_{j\neq k}^{K}\mathbf{g}^T_{\mathtt{RD},k}\hat{\mathbf{g}}^*_{\mathtt{RD},j}x_j[i-d] + n_{\mathtt{D},k}[i], \quad (13)\end{aligned}$$

where $\mathbf{g}_{\mathtt{RD},k}$ and $\hat{\mathbf{g}}_{\mathtt{RD},k}$ are the kth columns of $\mathbf{G}_{\mathtt{RD}}$ and $\hat{\mathbf{G}}_{\mathtt{RD}}$, respectively, and $n_{\mathtt{D},k}[i]$ is the kth element of $\mathbf{n}_{\mathtt{D}}[i]$.

III. LI Cancellation with Massive Arrays

In this section, we consider the potential of using massive MIMO technology to cancel the LI due to the FD operation at the RS. Some interesting insights are also presented.

A. Using a Large Receive Antenna Array ($N_{\mathrm{rx}} \to \infty$)

The LI can be canceled out by projecting it onto its orthogonal space. However, this orthogonal projection may harm the desired signal. Yet, when N_{rx} is large, the subspace spanned by the LI is nearly orthogonal to the desired signal's subspace and, hence, the orthogonal projection scheme will work very well. The next question is how to project the LI component? It is interesting to observe that, when N_{rx} grows large, the channel vectors of the desired signal and the LI become nearly orthogonal. Therefore, the MRC receiver can act as an orthogonal projection of the LI. As a result, the LI can be reduced significantly by using large N_{rx} together with the MRC receiver. This main result is summarized in the following proposition.

Proposition 1: Assume that the number of source-destination pairs, K, is fixed. As $N_{\mathrm{rx}} \to \infty$, the received signal at the RS for decoding the signal transmitted from $\mathtt{S}_k$ is given by

$$\frac{r_k[i]}{N_{\mathrm{rx}}} \overset{a.s.}{\to} \sqrt{p_{\mathtt{S}}}\sigma^2_{\mathtt{SR},k}x_k[i], \quad (14)$$

where $\overset{a.s.}{\to}$ denotes almost sure convergence.

Proof 1: See Appendix A.

Note that (14) holds for both cases of N_{tx} being fixed and $N_{\mathrm{tx}} \to \infty$. The aforementioned result implies that, when N_{rx} grows to infinity, the LI effect can be canceled out. Furthermore, the interpair interference and noise effects also disappear. The received signal at the RS after using MRC processing includes only the desired signal and, hence, the capacity of this link grows without bound. As a result, the system performance is limited only by the performance of the communication link $\mathtt{R} \to \mathtt{D}_k$ which does not depend on the LI.

B. Using a Large Transmit Antenna Array and Low Transmit Power ($p_{\mathtt{R}} = E_{\mathsf{R}}/N_{\mathrm{tx}}$, where E_{R} is Fixed, and $N_{\mathrm{tx}} \to \infty$)

Another way to reduce the LI effect is to use low transmit power $p_{\mathtt{R}}$. This will also reduce the quality of the transmission link $\mathtt{R} \to \mathtt{D}_k$ and, hence, the end-to-end (e2e) system performance will be degraded. However, with a large N_{tx}, we can reduce the relay transmit power while maintaining a desired quality-of-service of the transmission link $\mathtt{R} \to \mathtt{D}_k$. Therefore, we propose to use a very large N_{tx} together with low transmit power at the RS. With this method, the LI effect of the transmission link $\mathtt{S}_k \to \mathtt{R}$ becomes negligible, while the quality of the transmission link $\mathtt{R} \to \mathtt{D}_k$ is still fairly good. As a result, we can obtain a good e2e performance.

Proposition 2: Assume that K is fixed, and the transmit power at the RS is $p_{\mathtt{R}} = E_{\mathsf{R}}/N_{\mathrm{tx}}$, where E_{R} is fixed regardless of $N_{\rm tx}$. As $N_{\rm tx} \to \infty$, the received signals at the RS and $\mathtt{D}_k$ converge respectively to

$$r_k[i] \xrightarrow{d} \sqrt{p_{\mathtt{S}}}\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},k} x_k[i] + \sqrt{p_{\mathtt{S}}} \sum_{j\neq k}^{K} \hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},j} x_j[i] + \hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{n}_{\mathtt{R}}[i], \quad (15)$$

$$y_{\mathtt{D},k}[i] \xrightarrow{d} \sqrt{\frac{\sigma_{\mathtt{RD},k}^4 E_{\mathtt{R}}}{\sum_{j=1}^K \sigma_{\mathtt{RD},j}^2}} x_k[i-d] + n_{\mathtt{D},k}[i], \quad (16)$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Proof 2: Using a similar method as in Appendix A.

We can see that, by using a very low transmit power, i.e., the transmit power is scaled proportionally to $1/N_{\rm tx}$, the LI effect at the receive antennas is negligible [see (15)]. Although the transmit power is low, the power level of the desired signal received at each $\mathtt{D}_k$ is good enough thanks to the improved array gain, when $N_{\rm tx}$ grows large. At the same time, interpair interference at each $\mathtt{D}_k$ disappears due to the orthogonality between the channel vectors [see (16)]. As a result, the quality of the second hop $\mathtt{R} \to \mathtt{D}_k$ is still good enough to provide a robust overall e2e performance.

IV. Achievable Rate Analysis

In this section, we derive the e2e achievable rate of the transmission link $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$. The achievable rate is limited by the weakest link, i.e., is equal to the minimum of the achievable rates of the transmissions from $\mathtt{S}_k$ to $\mathtt{R}$ and from $\mathtt{R}$ to $\mathtt{D}_k$ [6]. To obtain this achievable rate, we use a technique from [14]. This technique is widely used in massive MIMO systems since: i) it yields a simplified insightful rate expression; and ii) it does not require instantaneous channel state information (CSI) at the destination [15]–[17]. The e2e achievable rate of the transmission link $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$ is given by

$$R_k = \min\left\{R_{\mathtt{SR},k}, R_{\mathtt{RD},k}\right\}, \quad (17)$$

where $R_{\mathtt{SR},k}$ and $R_{\mathtt{RD},k}$ are the achievable rates of the transmission links $\mathtt{S}_k \to \mathtt{R}$ and $\mathtt{R} \to \mathtt{D}_k$, respectively. We next compute $R_{\mathtt{SR},k}$ and $R_{\mathtt{RD},k}$. To compute $R_{\mathtt{SR},k}$, we consider (10). From (10), the received signal used for detecting $x_k[i]$ at the RS can be written as

$$r_k[i] = \underbrace{\sqrt{p_{\mathtt{S}}}\,\mathtt{E}\left\{\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},k}\right\} x_k[i]}_{\text{desired signal}} + \underbrace{\tilde{n}_{\mathtt{R},k}[i]}_{\text{effective noise}}, \quad (18)$$

where $\tilde{n}_{\mathtt{R},k}[i]$ is considered as the effective noise, given by

$$\tilde{n}_{\mathtt{R},k}[i] = \sqrt{p_{\mathtt{S}}}\left(\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},k} - \mathtt{E}\left\{\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},k}\right\}\right) x_k[i] + \sqrt{p_{\mathtt{S}}} \sum_{j\neq k}^{K} \hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},j} x_j[i] + \sqrt{p_{\mathtt{R}}}\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{G}_{\mathtt{RR}} \mathbf{s}[i] + \hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{n}_{\mathtt{R}}[i].$$

We can see that the "desired signal" and the "effective noise" in (18) are uncorrelated. Therefore, by using the fact that the worst-case uncorrelated additive noise is independent Gaussian noise of the same variance [14], we can obtain an achievable rate as

$$R_{\mathtt{SR},k} = \log_2\left(1 + \frac{p_{\mathtt{S}}\left|\mathtt{E}\left\{\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},k}\right\}\right|^2}{p_{\mathtt{S}}\mathbb{V}\mathrm{ar}\left(\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},k}\right) + \mathtt{MP}_k + \mathtt{LI}_k + \mathtt{AN}_k}\right), \quad (19)$$

where $\mathtt{MP}_k$, $\mathtt{LI}_k$, and $\mathtt{AN}_k$ represent the multipair interference, LI, and additive noise effects, respectively, given by

$$\mathtt{MP}_k \triangleq p_{\mathtt{S}} \sum_{j\neq k}^{K} \mathtt{E}\left\{\left|\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},j}\right|^2\right\}, \quad (20)$$

$$\mathtt{LI}_k \triangleq p_{\mathtt{R}}\alpha^2\,\mathtt{E}\left\{\left\|\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{G}_{\mathtt{RR}} \hat{\mathbf{G}}_{\mathtt{RD}}^*\right\|^2\right\}, \quad (21)$$

$$\mathtt{AN}_k \triangleq \mathtt{E}\left\{\left\|\hat{\mathbf{g}}_{\mathtt{SR},k}\right\|^2\right\}. \quad (22)$$

To compute $R_{\mathtt{RD},k}$, we consider (13). Following a similar methodology as in the derivation of $R_{\mathtt{SR},k}$, we obtain

$$R_{\mathtt{RD},k} = \log_2\left(1 + \frac{\left|\mathtt{E}\left\{\mathbf{g}_{\mathtt{RD},k}^T \hat{\mathbf{g}}_{\mathtt{RD},k}^*\right\}\right|^2}{\mathbb{V}\mathrm{ar}\left(\mathbf{g}_{\mathtt{RD},k}^T \hat{\mathbf{g}}_{\mathtt{RD},k}^*\right) + \sum_{j\neq k}^{K} \mathtt{E}\left\{\left|\mathbf{g}_{\mathtt{RD},k}^T \hat{\mathbf{g}}_{\mathtt{RD},j}^*\right|^2\right\} + \frac{1}{\alpha^2 p_{\mathtt{R}}}}\right). \quad (23)$$

Remark 1: The achievable rates in (19) and (23) are obtained by approximating the effective noise via an additive Gaussian noise. Since the effective noise is a sum of many terms, the central limit theorem guarantees that this is a good approximation, especially for massive MIMO systems.

Remark 2: The achievable rate in (23) is obtained by assuming that the destination, $\mathtt{D}_k$ uses only statistical knowledge of the channel gains $\left(\text{i.e., } \mathtt{E}\left\{\mathbf{g}_{\mathtt{RD},k}^T \hat{\mathbf{g}}_{\mathtt{RD},k}^*\right\}\right)$ to decode the transmitted signals and hence does not need to spend time, frequency, and power resources for CSI acquisition. However, an interesting question is: are our achievable rate expressions fairly good predictors of the system performance? To answer this question, we compare our achievable rate (23) with the achievable rate of the genie receiver, i.e., the RS knows $\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},j}$ and $\mathbf{G}_{\mathtt{RR}}$, and the destination $\mathtt{D}_k$ perfectly knows $\mathbf{g}_{\mathtt{RD},k}^T \hat{\mathbf{g}}_{\mathtt{RD},j}^*$, $j = 1, ..., K$. For this case, the e2e achievable rate of the transmission link $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$ is given by

$$\tilde{R}_k = \min\left\{\tilde{R}_{\mathtt{SR},k}, \tilde{R}_{\mathtt{RD},k}\right\}, \quad (24)$$

where

$$\tilde{R}_{\mathtt{SR},k} = \mathtt{E}\left\{\log_2\left(1 + \frac{p_{\mathtt{S}}\left|\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},k}\right|^2}{p_{\mathtt{S}}\sum_{j\neq k}^{K}\left|\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{g}_{\mathtt{SR},j}\right|^2 + p_{\mathtt{R}}\alpha^2\left\|\hat{\mathbf{g}}_{\mathtt{SR},k}^H \mathbf{G}_{\mathtt{RR}} \hat{\mathbf{G}}_{\mathtt{RD}}^*\right\|^2 + \left\|\hat{\mathbf{g}}_{\mathtt{SR},k}\right\|^2}\right)\right\} \quad (25)$$

and

$$\tilde{R}_{\mathtt{RD},k} = \mathtt{E}\left\{\log_2\left(1+\frac{\alpha^2 p_{\mathtt{R}}\left|\mathbf{g}_{\mathtt{RD},k}^T\hat{\mathbf{g}}_{\mathtt{RD},k}^*\right|^2}{\alpha^2 p_{\mathtt{R}}\sum_{j\neq k}^{K}\left|\mathbf{g}_{\mathtt{RD},k}^T\hat{\mathbf{g}}_{\mathtt{RD},j}^*\right|^2+1}\right)\right\}. \quad (26)$$

In Section V, we will show that the performance gap between the achievable rates given by (17) and (24) are negligible, especially for large $N_{\rm rx}$ and $N_{\rm tx}$. Note that the above achievable rate is obtained under the assumption of perfect CSI for which a lot of resources for the channel estimation are needed.

We next provide a closed-form expression for the exact achievable rate given by (17):

Theorem 1: With MRC/MRT processing, the achievable rate of the transmission link $\mathtt{S}_k \to \mathtt{R} \to \mathtt{D}_k$, for a finite number of RS antennas, is given by

$$R_k = \log_2\left(1+\min\left(\frac{p_{\mathtt{S}} N_{\rm rx}\sigma_{\mathtt{SR},k}^2}{p_{\mathtt{S}}\sum_{j=1}^{K}\beta_{\mathtt{SR},j}+p_{\mathtt{R}}\sigma_{\mathtt{LI}}^2+1},\right.\right.$$
$$\left.\left.\frac{\sigma_{\mathtt{RD},k}^4}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2}\frac{p_{\mathtt{R}}N_{\rm tx}}{p_{\mathtt{R}}\beta_{\mathtt{RD},k}+1}\right)\right). \quad (27)$$

Proof 3: See Appendix B.

A. Discussion of Results

1) We can see from (27) that, when $N_{\rm rx} \to \infty$, the achievable rate of the link $\mathtt{S}_k \to \mathtt{R}$ grows without bound and the e2e performance is limited by the second link $\mathtt{R} \to \mathtt{D}_k$, i.e., as $N_{\rm rx} \to \infty$,

$$R_k \to \log_2\left(1+\frac{\sigma_{\mathtt{RD},k}^4}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2}\frac{p_{\mathtt{R}}N_{\rm tx}}{p_{\mathtt{R}}\beta_{\mathtt{RD},k}+1}\right), \quad (28)$$

which coincides with the insights obtained in Proposition 1.

By using a massive transmit antenna array together with a very low transmit power, $p_{\mathtt{R}} = E_{\mathsf{R}}/N_{\rm tx}$, the loop interference can be canceled out. More precisely, from (27), as $N_{\rm tx} \to \infty$

$$R_k \to \log_2\left(1+\min\left(\frac{p_{\mathtt{S}} N_{\rm rx}\sigma_{\mathtt{SR},k}^2}{p_{\mathtt{S}}\sum_{j=1}^{K}\beta_{\mathtt{SR},j}+1},\frac{\sigma_{\mathtt{RD},k}^4 E_{\mathsf{R}}}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2}\right)\right), \quad (29)$$

which coincides with the insights obtained in Proposition 2.

2) Power Efficiency: We now study the power savings offered by deploying massive arrays at the RS.

Case I: If $p_{\rm p}$ is fixed, $p_{\mathtt{S}} = E_{\mathsf{S}}/N_{\rm rx}$, and $p_{\mathtt{R}} = E_{\mathsf{R}}/N_{\rm tx}$, where E_{S} and E_{R} are fixed regardless of $N_{\rm rx}$ and $N_{\rm tx}$, when $N_{\rm tx}$ and $N_{\rm rx}$ go to infinity with the same speed, the e2e achievable rate can be expressed as

$$R_k \to \log_2\left(1+\min\left(E_{\mathsf{S}}\sigma_{\mathtt{SR},k}^2,\frac{\sigma_{\mathtt{RD},k}^4 E_{\mathsf{R}}}{\sum_{j=1}^{K}\sigma_{\mathtt{RD},j}^2}\right)\right). \quad (30)$$

This implies that, with large antenna arrays, we can reduce the transmitted power of each source and of the RS proportionally to $1/N_{\rm rx}$ and $1/N_{\rm tx}$, respectively, while maintaining a given quality-of-service. At the same time, the sum rate is increased K times (since K communication pairs share the same time-frequency resource).

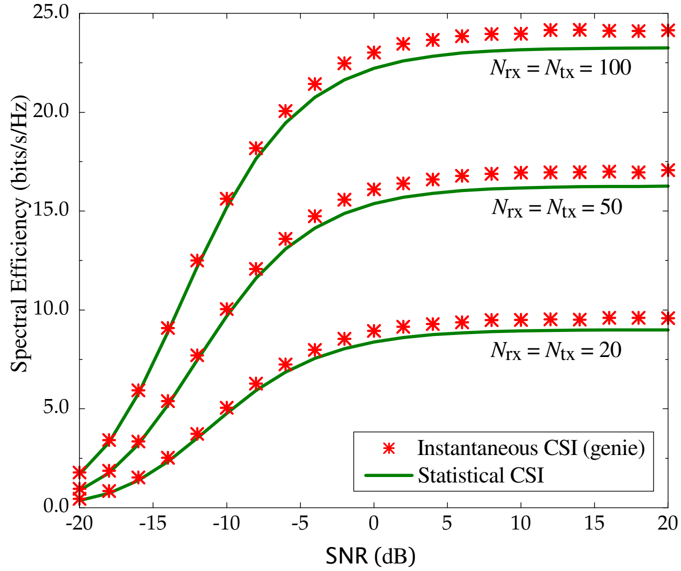

Fig. 2. Spectral efficiency versus SNR ($K = 10$, $\tau = 2K$, and $\sigma_{\mathtt{LI}}^2 = 1$).

Case II: If $p_{\rm p} = p_{\mathtt{S}} = E_{\mathsf{S}}/\sqrt{N_{\rm rx}}$ and $p_{\mathtt{R}} = E_{\mathsf{R}}/\sqrt{N_{\rm tx}}$, where E_{S} and E_{R} are fixed regardless of $N_{\rm rx}$ and $N_{\rm tx}$, when $N_{\rm rx} \to \infty$, $N_{\rm tx} = \kappa N_{\rm rx}$, the spectral efficiency converges to

$$R_k \to \log_2\left(1+\min\left(\tau E_{\mathsf{S}}^2\beta_{\mathtt{SR},k}^2,\frac{\sqrt{\kappa}\tau E_{\mathsf{S}}E_{\mathsf{R}}\beta_{\mathtt{RD},k}^4}{\sum_{j=1}^{K}\beta_{\mathtt{RD},j}^{-2}}\right)\right). \quad (31)$$

We can see that, if the transmit powers of the uplink training and data transmission are the same, (i.e., $p_{\rm p} = p_{\mathtt{S}}$), we cannot reduce the transmit powers of each source and of the RS as aggressively as in *Case I* where the pilot power is kept fixed. Instead, we can scale down the transmit powers of each source and of the RS proportionally to only $1/\sqrt{N_{\rm rx}}$ and $1/\sqrt{N_{\rm tx}}$, respectively. This comes from the fact that, when we cut the transmitted power of each source, both the data signal and the pilot signal suffer from a power reduction, which leads to the so-called "squaring effect" on the spectral efficiency [14].

V. Numerical Results

To evaluate the system performance, we consider the spectral efficiency. The spectral efficiency is defined as the sum-rate (in bits) per channel use:

$$\mathcal{S}_{\rm FD} \triangleq \frac{T-\tau}{T}\sum_{k=1}^{K}R_k, \quad (32)$$

where T is the length of the coherence interval (in symbols). In all illustrative examples, we choose $T = 200$, $K = 10$, $\tau = 2K$, $N_{\rm tx} = N_{\rm rx}$, and $\beta_{\mathtt{SR},k} = \beta_{\mathtt{RD},k} = 1$, $k = 1, ..., K$.

Firstly, we evaluate the validity of our achievable rate given in (27). We choose $\sigma_{\mathtt{LI}}^2 = 1$. We assume that $p_{\rm p} = p_{\mathtt{S}}$, and that the total transmit power of the K sources is equal to the transmit power of the RS, i.e., $p_{\mathtt{R}} = Kp_{\mathtt{S}}$. We define $\mathsf{SNR} \triangleq p_{\mathtt{S}}$. Figure 2 compares the spectral efficiency obtained from (27), where the destination uses the statistical distributions of the channels (i.e., mean of the channel gains) to detect

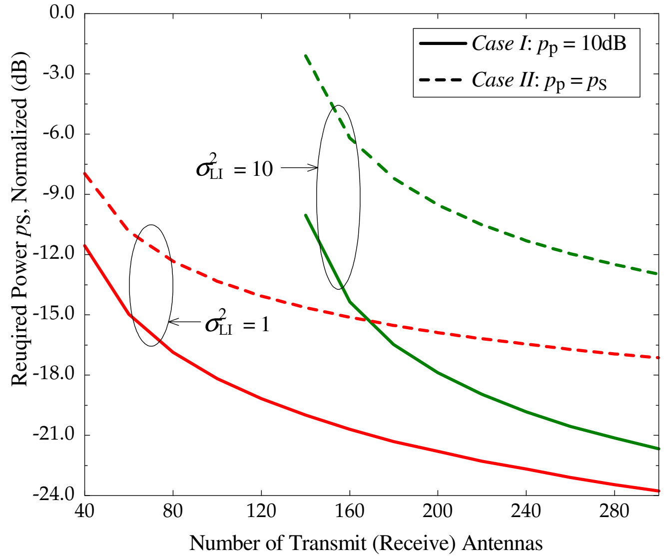

Fig. 3. Transmit power, p_S required to achieve 1 bit/channel use per user ($K = 10$, $\tau = 2K$, and $p_R = Kp_S$).

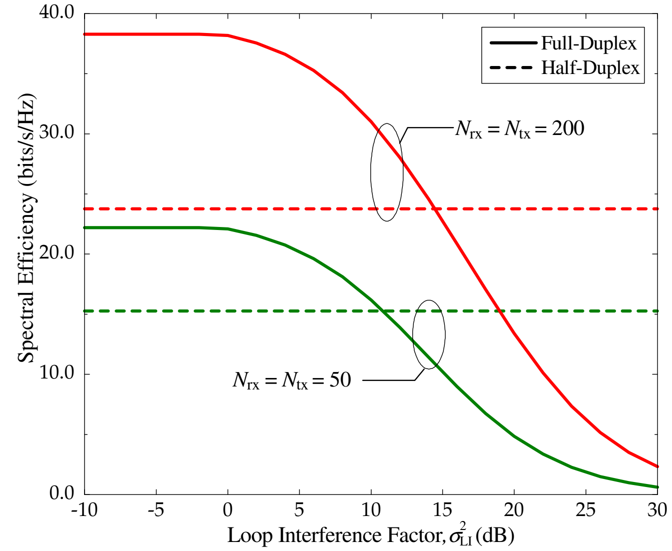

Fig. 4. Spectral efficiency versus the loop interference levels for HD and FD relaying ($K = 10$, $\tau = 2K$, and $p_R = p_P = p_S = 10$ dB).

the transmitted signal, with the one obtained numerically from (24), where we assume that there is a genie receiver (instantaneous CSI) at the destination. We can see that the relative performance gap between the cases with instantaneous and statistical CSI is small, especially at low SNR. This implies that using the mean of the effective channel gain for signal detection is fairly reasonable, and the achievable rate given in (27) is a good predictor of the system performance.

We next examine the power efficiency of using large antenna arrays for two cases: p_P is fixed and $p_P = p_S$. We set $p_R = Kp_S$. Figure 3 shows the required transmit power, p_S, to achieve 1 bits/s/Hz per communication pair. We can see that when the number of antennas increases, the required transmit powers are significantly reduced. As predicted by the analysis, when the number of antennas is large, we can cut back the power by approximately 3 dB and 1.5 dB by doubling the number of antennas for *Case I* and *Case II*, respectively. Regarding the effect of the LI, when σ^2_{LI} increases, we need more transmit power. However, when σ^2_{LI} is high and the number of antennas is small, even if we use infinite transmit power, we cannot achieve a required spectral efficiency. Instead of this, we can add more antennas to reduce the LI effect and achieve the required quality-of-service.

Finally, we compare the performances between the HD and the FD modes. For the HD mode, two orthogonal time slots are allocated for two transmissions: sources to the RS and RS to destinations. The HD mode avoids LI at the cost of imposing a pre-log factor $1/2$ for the spectral efficiency. The spectral efficiency of the HD mode can be obtained directly from (32) by neglecting the LI effect. Note that, for fair comparison, the total energies spent in a coherence interval for both modes are set to be the same. Thus, the transmit power of each source and of the RS used in the HD mode are doubled compared to the power used in the FD mode. In addition, we assume that the number of antennas at the RS used in the HD mode is equal to the total number of transmit and receive antennas used in the FD mode, i.e., is equal to $N_{tx} + N_{rx}$. The adopted comparison is named as the "number of antenna preserved" condition in the literature [10, Section 3].[2]

We choose $p_R = p_P = p_S = 10$ dB. Figure 4 shows the spectral efficiency versus the LI level, σ^2_{LI}, for different number of antennas. As expected, at low σ^2_{LI}, FD relaying outperforms HD relaying. This gain is due to the larger pre-log factor (one) of the FD mode. However, when σ^2_{LI} is high, LI dominates the system performance of the FD mode and, hence, the performance of the HD mode is superior. In this case, by using larger antenna arrays at the RS, we can reduce the effect of the LI and exploit the larger pre-log factor of the FD mode. For example, compared with the case of $N_{rx} = N_{tx} = 50$, the LI can be reduced by 10 dB for $N_{rx} = N_{tx} = 200$, at the spectral efficiency of 15 bits/s/Hz.

VI. Conclusion

In this paper, we proposed and analyzed a multipair FD relaying system where the RS is equipped with massive arrays, while each source and destination have a single antenna. We assume that the RS employs MRC/MRT to process the signals. The analysis takes the channel estimation error, training duration, and coherence interval into account. We show that, by using massive arrays at the RS, the LI can be canceled out. Furthermore, the inter-pair interference and noise disappear.

[2]The comparison between HD and FD modes can also be performed with the "RF chain conserved" condition, where an equal number of total RF chains are assumed. Note that, with the "RF chain conserved" condition, the number of antennas at the RS employed in the HD mode will be less than the total transmit and receive antennas in the FD mode [10, Section III]. Therefore, HD mode with the antenna conserved condition (using a higher number of RF chains) performs much better than HD mode with the RF chain conserved condition. In general, the cost of the required RF chains is significant as opposed to adding an extra antenna.

As a result, massive MIMO can increase the spectral efficiency by $2K$ times compared to the conventional orthogonal HD relaying, and simultaneously reducing the transmit power significantly. In addition, we derived a closed-form expression for the achievable rate and compared the performance of the FD and HD modes.

APPENDIX

A. Proof of Proposition 1

By using the law of large numbers, when the number of receive antennas goes to infinity, the desired signal converges to a deterministic value, while the multipair interference is cancelled out. More specifically, as $N_{\rm rx} \to \infty$,

$$\frac{1}{N_{\rm rx}}\sqrt{p_{\tt S}}\hat{\mathbf{g}}_{{\tt SR},k}^H \mathbf{g}_{{\tt SR},k} x_k\left[i\right] \overset{a.s.}{\to} \sqrt{p_{\tt S}}\sigma_{{\tt SR},k}^2 x_k\left[i\right], \tag{33}$$

$$\frac{1}{N_{\rm rx}}\sqrt{p_{\tt S}}\hat{\mathbf{g}}_{{\tt SR},k}^H \mathbf{g}_{{\tt SR},j} x_k\left[j\right] \overset{a.s.}{\to} 0,\ \forall j \neq k. \tag{34}$$

We next consider the LI. We have

$$\frac{1}{N_{\rm rx}}\sqrt{p_{\tt R}}\hat{\mathbf{G}}_{\tt SR}^H \mathbf{G}_{\tt RR}\mathbf{s}\left[i\right] = \alpha\sqrt{p_{\tt R}}\frac{\hat{\mathbf{G}}_{\tt SR}^H \mathbf{G}_{\tt RR}\hat{\mathbf{G}}_{\tt RD}^*}{N_{\rm rx}}\mathbf{x}\left[i-d\right]. \tag{35}$$

If $N_{\rm tx}$ is fixed, then $\frac{1}{N_{\rm rx}}\sqrt{p_{\tt R}}\hat{\mathbf{G}}_{\tt SR}^H \mathbf{G}_{\tt RR}\mathbf{s}\left[i\right] \overset{a.s.}{\to} 0$, as $N_{\rm rx} \to \infty$. We now consider the case where $N_{\rm tx}$ and $N_{\rm rx}$ tend to infinity with a fixed ratio. The (m,n)th element of the $K \times K$ matrix $\alpha\frac{\hat{\mathbf{G}}_{\tt SR}^H \mathbf{G}_{\tt RR}\hat{\mathbf{G}}_{\tt RD}^*}{N_{\rm rx}}$ is given by

$$\alpha\frac{\hat{\mathbf{g}}_{{\tt SR},m}^H \mathbf{G}_{\tt RR}\hat{\mathbf{g}}_{{\tt RD},n}^*}{N_{\rm rx}} = \sqrt{\frac{1}{\sum_{k=1}^K \sigma_{{\tt RD},k}^2}}\frac{1}{N_{\rm rx}}\hat{\mathbf{g}}_{{\tt SR},m}^H\frac{\mathbf{G}_{\tt RR}\hat{\mathbf{g}}_{{\tt RD},n}^*}{\sqrt{N_{\rm tx}}}. \tag{36}$$

We can see that the vector $\frac{\mathbf{G}_{\tt RR}\hat{\mathbf{g}}_{{\tt RD},n}^*}{\sqrt{N_{\rm tx}}}$ includes i.i.d. zero-mean RVs with variance $\sigma_{{\tt RD},n}^2\sigma_{\tt LI}^2$. This vector is independent of the vector $\hat{\mathbf{g}}_{{\tt SR},m}$. Thus, by using the law of large numbers, we can obtain

$$\alpha\frac{\hat{\mathbf{g}}_{{\tt SR},m}^H \mathbf{G}_{\tt RR}\hat{\mathbf{g}}_{{\tt RD},n}^*}{N_{\rm rx}} \overset{a.s.}{\to} 0, \tag{37}$$

as $N_{\rm rx} \to \infty$, $N_{\rm rx}/N_{\rm tx}$ is fixed. Thus, the LI converges to 0 when $N_{\rm rx}$ grows without bound. Similarly, we can show that

$$\frac{1}{N_{\rm rx}}\hat{\mathbf{G}}_{\tt SR}^H \mathbf{n}_{\tt R}\left[i\right] \overset{a.s.}{\to} 0. \tag{38}$$

Substituting (33), (34), (37), and (38) into (10), we obtain (14).

B. Proof of Theorem 1

We have

$$\hat{\mathbf{g}}_{{\tt SR},k}^H \mathbf{g}_{{\tt SR},k} = \left\|\hat{\mathbf{g}}_{{\tt SR},k}\right\|^2 + \hat{\mathbf{g}}_{{\tt SR},k}^H \boldsymbol{\varepsilon}_{{\tt SR},k}. \tag{39}$$

Therefore,

$$\mathtt{E}\left\{\hat{\mathbf{g}}_{{\tt SR},k}^H \mathbf{g}_{{\tt SR},k}\right\} = \mathtt{E}\left\{\left\|\hat{\mathbf{g}}_{{\tt SR},k}\right\|^2\right\} = \sigma_{{\tt SR},k}^2 N_{\rm rx}. \tag{40}$$

From (39) and (40), the variance of $\hat{\mathbf{g}}_{{\tt SR},k}^H \mathbf{g}_{{\tt SR},k}$ is

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\hat{\mathbf{g}}_{{\tt SR},k}^H \mathbf{g}_{{\tt SR},k}\right) &= \mathtt{E}\left\{\left|\hat{\mathbf{g}}_{{\tt SR},k}^H \mathbf{g}_{{\tt SR},k}\right|^2\right\} - \sigma_{{\tt SR},k}^4 N_{\rm rx}^2 \\ &= \mathtt{E}\left\{\left|\left\|\hat{\mathbf{g}}_{{\tt SR},k}\right\|^2 + \hat{\mathbf{g}}_{{\tt SR},k}^H \boldsymbol{\varepsilon}_{{\tt SR},k}\right|^2\right\} - \sigma_{{\tt SR},k}^4 N_{\rm rx}^2 \\ &= \mathtt{E}\left\{\left\|\hat{\mathbf{g}}_{{\tt SR},k}\right\|^4\right\} + \mathtt{E}\left\{\left|\hat{\mathbf{g}}_{{\tt SR},k}^H \boldsymbol{\varepsilon}_{{\tt SR},k}\right|^2\right\} - \sigma_{{\tt SR},k}^4 N_{\rm rx}^2. \end{aligned} \tag{41}$$

By using [18, Lemma 2.9], we obtain

$$\mathbb{V}\mathrm{ar}\left(\hat{\mathbf{g}}_{{\tt SR},k}^H \mathbf{g}_{{\tt SR},k}\right) = \sigma_{{\tt SR},k}^2 \beta_{{\tt SR},k} N_{\rm rx}. \tag{42}$$

Similarly, we obtain

$$\mathtt{MP}_k = p_{\tt S}\sigma_{{\tt SR},k}^2 N_{\rm rx}\sum_{j\neq k}^K \beta_{{\tt SR},j} \tag{43}$$

$$\mathtt{LI}_k = p_{\tt R}\sigma_{\tt LI}^2\sigma_{{\tt SR},k}^2 N_{\rm rx} \tag{44}$$

$$\mathtt{AN}_k = \sigma_{{\tt SR},k}^2 N_{\rm rx}. \tag{45}$$

Substituting (40), (41)-(44) into (19), we obtain

$$R_{{\tt SR},k} = \log_2\left(1 + \frac{p_{\tt S} N_{\rm rx}\sigma_{{\tt SR},k}^2}{p_{\tt S}\sum_{j=1}^K \beta_{{\tt SR},j} + p_{\tt R}\sigma_{\tt LI}^2 + 1}\right). \tag{46}$$

Similarly, we arrive at (27).

REFERENCES

[1] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[2] 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, no. 1, pp. 40-60, Jan. 2013.

[3] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195 , Feb. 2014.

[4] H. A. Suraweera, H. Q. Ngo, T. Q. Duong, C. Yuen, and E. G. Larsson, “Multi-pair amplify-and-forward relaying with very large antenna arrays,” *IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, June 2013.

[5] T. Riihonen, S. Werner, and R. Wichman, “Hybrid full-duplex/half-duplex relaying with transmit power adaptation,” *IEEE Trans. Wireless Commun.*, vol. 10, pp. 3074–3085, Sept. 2011.

[6] —, “Transmit power optimization for multiantenna decode-and-forward relays with loopback self-interference from full-duplex operation,” in *Proc. Annual Asilomar Conf. Signals, Syst., Comput.*, CA, Nov. 2011.

[7] —, “Mitigation of loopback self-interference in full-duplex MIMO relays,” *IEEE Trans. Signal Process.*, vol. 59, pp. 5983-5993, Dec. 2011.

[8] I. Krikidis, H. A. Suraweera, P. J. Smith, and C. Yuen, “Full-duplex relay selection for amplify-and-forward cooperative networks,” *IEEE Trans. Wireless Commun.*, vol. 11, pp. 4381-4393, Dec. 2012.

[9] M. Duarte, “Full-duplex wireless: Design, implementation and characterization,” Ph.D. dissertation, Rice University, Houston, TX, 2012.

[10] E. Aryafar, M. A. Khojastepour, K. Sundaresan, S. Rangarajan, and M. Chiang, “MIDU: Enabling MIMO full duplex,” in *Proc. ACM Annual Int. Conf. Mobile Comput. Network. (Mobicom)*, Turkey, Aug. 2012.

[11] H. A. Suraweera, I. Krikidis, and C. Yuen, “Antenna selection in the full-duplex multi-antenna relay channel,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, June 2013, pp. 4823-4828.

[12] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, “Linear pre-coding performance in measured very-large MIMO channels,” in *Proc. IEEE Veh. Technol. Conf. (VTC)*, San Francisco, CA, Sept. 2011.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation heory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[14] B. Hassibi and B. M. Hochwald, “How much training is needed in multiple-antenna wireless links?” *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[15] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, “Pilot contamination and precoding in multi-cell TDD systems,” *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2640–2651, Aug. 2011.

[16] H. Yang and T. L. Marzetta, “Performance of conjugate and zero-forcing beamforming in large-scale antenna systems,” *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 172–179, Feb. 2013.

[17] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson , “On the optimality of single-carrier transmission in large-scale antenna systems,” *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 276–279, Aug. 2012.

[18] A. M. Tulino and S. Verdú, “Random matrix theory and wireless communications,” *Foundations and Trends in Communications and Information Theory*, vol. 1, pp. 1–182, June 2004.