University of
Hertfordshire UH

Rate-Splitting Robustness in Multi-Pair Massive MIMO Relay Systems

Anastasios Papazafeiropoulos and Tharmalingam Ratnarajah

Abstract—Relay systems improve both coverage and system capacity. Towards this direction, full-duplex (FD) technology, being able to boost the spectral efficiency by transmitting and receiving simultaneously on the same frequency and time resources, is envisaged to play a key role in future networks. However, its benefits come at the expense of self-interference (SI) from their own transmit signal. At the same time, massive multiple-input massive multiple-output (MIMO) systems, bringing unconventionally many antennas, emerge as a promising technology with huge degrees-of-freedom (DoF). To this end, this paper considers a multi-pair decode-and-forward FD relay channel, where the relay station is deployed with a large number of antennas. Moreover, the rate-splitting (RS) transmission has recently been shown to provide significant performance benefits in various multi-user scenarios with imperfect channel state information at the transmitter (CSIT). Engaging the RS approach, we employ the deterministic equivalent (DE) analysis to derive the corresponding sum-rates in the presence of interferences. Initially, numerical results demonstrate the robustness of RS in half-duplex (HD) systems, since the achievable sum-rate increases without bound, i.e., it does not saturate at high signal-to-noise ratio (SNR). Next, we tackle the detrimental effect of SI in FD. In particular, and most importantly, not only FD outperforms HD, but also RS enables increasing the range of SI over which FD outperforms HD. Furthermore, increasing the number of relay station antennas, RS appears to be more efficacious due to imperfect CSIT, since SI decreases. Interestingly, increasing the number of users, the efficiency of RS worsens and its implementation becomes less favorable under these conditions. Finally, we verify that the proposed DEs, being accurate for a large number of relay station antennas, are tight approximations even for realistic system dimensions.

Index Terms—Rate-splitting, massive MIMO systems, half-duplex relaying, full-duplex relaying, deterministic equivalent analysis.

I. INTRODUCTION

Massive multiple-input massive multiple-output (MIMO) technology is a key enabler for the fifth generation (5G) wireless communication systems achieving energy-efficient transmission and high spectral efficiency [2]–[4]. According to its characteristic topology, a large number of service antennas per unit area performs coherent linear processing, and offers an unprecedented number of degrees-of-freedom (DoF). Among

its benefits, we emphasize the substantial reduction of both intra-cell and inter-cell interference, which ultimately, lead to high performance efficiency, both spectral and energy.

In a parallel direction, in-band full-duplex (FD) is a novel technology that doubles the throughput induced by standard half-duplex relaying by means of simultaneous transmission and reception at the same frequency and time during a wireless communication [5], [6]. Moreover, its theoretical and experimental progress towards its practical implementation [7]–[9] is notable. Actually, the theoretical progress can lead to a practical achievement with new opportunities. However, this is quite demanding because the FD transmission is accompanied by an inherent obstacle. Specifically, this obstacle is the so-called self-interference (SI) due to the leakage from the relay's output to its input [7]. It is worthwhile to mention that the main difference between SI and general interference is that SI is known at the receiver, which could be sufficient for SI suppression. There are several challenges for the mitigation of SI, being crucial for FD operation. For example, the received signal and the SI may exhibit a large amplitude difference going to exceed the dynamic range of the analog-to-digital converter at the receiver side [9]. Although the SI cancellers try their best to maximize the cancellation performance, residual interference remains and rate saturation at high signal-to-noise ratio (SNR) appears. Hence, the circumvention of the harmful consequences of the SI takes a prominent position in the research area of FD systems. Among the suppression methods for SI, MIMO processing, specialized in the spatial domain, provides an exceptionally effective means [7], [10], [11]. As a result, driving to massive MIMO is a reasonable approach for next-generation systems.

To grasp the benefits of massive MIMO the accurate knowledge of channel state information at the transmitter (CSIT) is required. In fact, accurate CSIT becomes even more challenging as the number of antennas increases [12], [13]. In such case, the Time Division Duplex (TDD) design has proved to be a more feasible solution against Frequency Division Duplex (FDD) schemes because the latter are accompanied with further channel estimation and feedback challenges [3], [4], [14]–[16]. The DoF decrease as the CSIT inaccuracy increases. Especially, in realistic scenarios, where CSIT is imperfect, linear precoding techniques lead to a rate ceiling at high SNR, if the error variance is fixed.

In order to enhance the sum DoF, the rate splitting (RS) strategy has been proposed [17]. The RS outperforms conventional broadcasting at high SNR because it does not experience

any ceiling effect [18]–[21][1]. According to this strategy, the message intended for one user is split into a private part and a common part by using a fraction of the total power. The private part is transmitted by means of zero-forcing (ZF) beamforming, while the common part is superimposed on top of the precoded private part by means of the remaining power. The common message is drawn from a public codebook and decoded by all users. At the receiver side, the decoding procedure involves first the decoding of the common message by means of successive interference cancellation, and then, the decoding of the private message of each user follows.

Although the relaying in previous cellular generations was mostly used for coverage enhancement, in today's cellular networks, it is shown that it can improve both coverage and system capacity [23]. In this regard, relaying has been already considered as one of the salient features in 3GPP Long Term Evolution (LTE) advanced [24]. Especially, the importance of relaying in massive MIMO systems has been already demonstrated in several studies [25] .

In the area of massive MIMO relaying, both half-duplex (HD) and FD have been studied [26]–[32]. In particular, in the case of HD, the spectral efficiency has been investigated for a very large number of relay station antennas [26]–[30], [33]. On the other hand, e.g., FD relaying with a large number of antennas and linear processing as well as the scaling behavior with the number of relay antennas of the self-interference were analyzed in [31] in terms of the end-to-end achievable rate. Towards this direction, the asymptotic performance of amplify-and-forward massive MIMO relay systems with additive hardware impairments was determined in [32].

A. Motivation-Contributions

Following the research trends and needs in massive MIMO and FD systems, we consider a collection of K sources communicating with another collection of K destinations through an intermediate massive MIMO FD relay station, and we focus on the application of RS. In particular, in our architecture scenario, two sources, leading to rate saturation, are faced. The first includes the multi-user interference with imperfect CSIT in the second link, and the second concerns the SI emerging from the FD transmission. This work tackles the challenge of mitigating the rate saturation by leveraging the RS approach. In particular, we investigate the robustness of the RS method in realistic massive MIMO FD settings suffering from both pilot contamination and SI. The motivation of this work started by the observation that in FD systems the CSIT is altered due to the presence of SI. Furthermore, it is known that RS is applicable in multi-user settings with imperfect CSIT. Hence, these observations suggest that RS will be effective in the mitigation of the SI and the consecutive circumvention of the rate saturation due to the overall imperfect CSIT. Note that our system setup is quite general, since it can model cellular networks with some users transmitting simultaneously signals

to several other users via an infrastructure-based relay station serving several roles such as a low power base station [34]. Moreover, having a MIMO relaying in the scene, we test RS in the basic scenario of just HD transmission. The main contributions are summarized as follows:

- Contrary to existing works such as [5]–[11], which have studied FD MIMO systems, we focus on massive MIMO systems, and examine the impact of SI, when RS transmission is applied at the second link. For the sake of comparison, we also present the results corresponding to an HD relay system. It is shown that RS is robust in both multipair HD and FD settings.
- We derive the deterministic SINRs of NoRS and RS in multipair FD systems with imperfect CSIT and use them to investigate the performance benefits of RS over NoRS in the presence of SI. Actually, first, we obtain the estimated channels of both links by means of MMSE estimation. Next, we apply RS in the second link by designing the precoder of the private and common messages, and we consider suitable power allocation. Although the basic implementation of the RS strategy assumes just ZF precoding for the transmission of the private messages except [17], we consider regularized ZF (RZF) precoding because it is another low-complexity linear processing technique applicable in massive MIMO systems. However, RZF provides better performance than ZF. Finally, we provide the DEs of the SINRs of the private and common messages. Note that these deterministic expressions allow avoiding any Monte Carlo simulations with very high precision.
- Above this, RS is robust in HD and FD scenarios because it can mitigate the multi-user interference taking place in the second link of both HD and FD cases. In fact, we elaborate on the impact of the severity of SI. Actually, RS is able to mitigate the saturation due to the SI in spite of the knowledge of perfect or imperfect CSIT. Furthermore, in the case of lower SI, RS behaves better. The same observation is made as the number of relay station antennas is increased, since then, SI becomes lower.
- We show that an increase of the number of user elements (UEs) in a multipair FD system results in a reduction of the performance gain of RS over NoRS because the common message has to be decoded by more UEs. Moreover, we quantify this decrease exhibited due to a less mitigated SI.

The remainder of this paper is structured as follows. Section II presents the system and signal models for both links of the multi-pair FD relay system. Section III presents the data transmission phase, while in Section III-B, we provide the estimated channels obtained during the uplink training phase of the two links. Next, we present the RS approach. In Section IV, we present the end-to-end transmission by obtaining the SINR of each link. Section V exposes the DE analysis, which enables the design of the precoder of the common message, and mainly, the derivation of the achievable rates in the presence of SI. The numerical results are placed in Section VI, while Section VII summarizes the paper.

---

[1]Interestingly, a further gain of RS over no RS (NoRS) can be achieved by optimizing the precoders [22], where for the sake of exposition and comparison, henceforth, we denote by NoRS all the conventional techniques to contrast with the RS techniques.

TABLE I
NOTATIONS SUMMARY

Notation	Description
K	Communication pairs
M, N	Numbers of transmit and receive antennas
S_k, D_k	The kth source and destination
T_c, B_c	Coherence time, bandwidth
σ_{SI}^2	The variance of the elements of the self-interference matrix
τ	Duration of the training phase
p_{S}, p_{tr}	Average transmit power per source and transmit power per pilot symbol
$\mathbf{G}_{\mathrm{SR}}$, $\mathbf{G}_{\mathrm{rD}}$, $\mathbf{G}_{\mathrm{RR}}$	Channel matrices of the first link, second link, and self-interference
$\mathbf{H}_{\mathrm{SR}}$, $\mathbf{H}_{\mathrm{RD}}$	Small-scale fading matrices of the first and second links
$\mathbf{D}_{\mathrm{SR}}$, $\mathbf{D}_{\mathrm{RD}}$	Large-scale fading matrices of the first and second links
$\mathbf{f}_c$, $\mathbf{f}_k$	Precoding vectors of the common and private messages corresponding to UE k
ρ_c, ρ_k	Powers allocated to the common and private messages corresponding to UE k
λ	Normalization of the precoded message
γ_{SR}, R_{SR}	SINR and achievable rate of the first link
$\gamma_{\mathrm{RD},k}^{\mathrm{c}}$, $\gamma_{\mathrm{RD},k}^{\mathrm{p}}$	SINR of the common and private messages of the second link
$R_{\mathrm{RD}}^{\mathrm{c}}$, $R_{\mathrm{RD},k}^{\mathrm{p}}$	Achievable rates of the common and private messages of the second link

Notation: Vectors and matrices are denoted by boldface lower and upper case symbols. $(\cdot)^{\mathsf{T}}$, $(\cdot)^*$, $(\cdot)^{\mathsf{H}}$, and $\mathrm{tr}(\cdot)$ represent the transpose, conjugate, Hermitian transpose, and trace operators, respectively. The expectation operator is denoted by $\mathbb{E}[\cdot]$. The $\mathrm{diag}\{\cdot\}$ operator generates a diagonal matrix from a given vector, and the symbol $\triangleq$ declares definition. The notations $\mathbb{C}^{M \times 1}$ and $\mathbb{C}^{M \times N}$ refer to complex M-dimensional vectors and $M \times N$ matrices, respectively. Finally, $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ denotes a circularly symmetric complex Gaussian variable with zero-mean and covariance matrix $\boldsymbol{\Sigma}$.

II. SYSTEM MODEL

The concept of our model involves a multipair FD relaying system with a common relay station R and K communication pairs $(\mathrm{S}_k, \mathrm{D}_k)$, $k = 1, \ldots, K$ sharing the same time-frequency resources. Specifically, we consider K user pairs, where the kth source S_k exchanges information through a relay operating in decode-and-forward protocol with the kth UE destination D_k. Moreover, the system suffers from SI due to the simultaneous transmission and reception, since it operates under an FD mode. Note that there is no direct link between the source S_k and the corresponding destination D_k because of heavy shadowing and large path-loss. The source and the destination pairs are equipped with a single antenna, while the FD relay station is deployed with N receive antennas and M transmit antennas, i.e., it includes $V = M + N$ antennas in total[2].

A. Signal Model

We consider frequency-flat channels between the source user k and the relay as well as between the relay and destination UE k, modeled as Rayleigh block fading. The channels are assumed static across a coherence block of T channel uses with the channel realizations between blocks being independent. The size of the block is defined by the product between the coherence time T_c and the coherence bandwidth B_c.

Specifically, the frequency-flat channel matrices between the K sources and the relay station's receive antenna array as well as between the relay station's transmit antenna array and the K destinations, modeled as Rayleigh block fading, are denoted by $\mathbf{G}_{\mathrm{SR}} \in \mathbb{C}^{N \times K}$ as well as $\mathbf{G}_{\mathrm{RD}} \in \mathbb{C}^{M \times K}$, respectively. We express each channel realization as[3]

$$\mathbf{G}_{\mathrm{SR}} \triangleq \mathbf{H}_{\mathrm{SR}} \mathbf{D}_{\mathrm{SR}}^{1/2} \qquad (1)$$

$$\mathbf{G}_{\mathrm{RD}} \triangleq \mathbf{H}_{\mathrm{RD}} \mathbf{D}_{\mathrm{RD}}^{1/2}. \qquad (2)$$

These channel matrices account for both small-scale and large-scale fadings. Specifically, the matrices $\mathbf{H}_{\mathrm{SR}}$ and $\mathbf{H}_{\mathrm{RD}}$, having independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ elements, describe small-scale fading, while the matrices $\mathbf{D}_{\mathrm{SR}}$ and $\mathbf{D}_{\mathrm{RD}}$ are diagonal and express the large-scale fading in terms of the kth diagonal elements, which are denoted by $\beta_{\mathrm{SR},k}$ and $\beta_{\mathrm{RD},k}$, respectively. Furthermore, assuming that there is no line-of-sight component, the SI channel is modeled by means of the Rayleigh fading distribution. Mathematically, it is described by the $\mathbf{G}_{\mathrm{RR}} \in \mathbb{C}^{M \times N}$ channel matrix between the relay's transmit and receive arrays. In other words, the elements of the SI channel matrix $\mathbf{G}_{\mathrm{RR}}$ can be modeled as i.i.d. complex Gaussian random variables with zero mean and variance σ_{SI}^2, i.e., $\mathcal{CN}(0, \sigma_{\mathrm{SI}}^2)$. The physical meaning of σ_{SI}^2 can be seen as the level of SI that is dependent on the distance between the transmit and receive antenna arrays. Also, the assumption that the channels between the transmit and receive antennas are i.i.d. considers that the distance between the transmit and the receive arrays is much larger than the distance between the antenna elements.

III. END-TO-END TRANSMISSION

This section presents the data transmission and the uplink estimation phases of the multipair decode-and-forward FD model as well as the RS approach.

---

[2]This network configuration is of high practical interest. For example, it can describe a cellular setup, where the communication between two users is performed by means of a massive antenna low power base station.

[3]According to the favorable propagation assumption which has been validated in practice [9], we consider that the channels from the relay station to different sources and destinations are independent [31].

A. Data Transmission

At time instant n, the K user sources S_k $(k = 1, \ldots, K)$ transmit simultaneously their signals to the relay, which, in turn, broadcasts the signal to all K destinations. Actually, we denote $\sqrt{p_S}u_k[n]$ the kth user transmit signal at time n with p_S being the average transmit power of each source since $\mathbb{E}\left\{|u_k[n]|^2\right\} = 1$, while at the relay station the received signal is interfered with its transmit signal.

Herein, we present the conventional input-output signal model (NoRS) as a measure of comparison. More precisely, the signal received by the receive antenna array of the relay from all the sources is given by [31]

$$\mathbf{y}_R[n] = \sqrt{p_S}\mathbf{G}_{SR}\mathbf{u}[n] + \mathbf{G}_{RR}\mathbf{s}[n] + \mathbf{z}_R[n], \qquad (3)$$

while the signal received by the K destinations from the transmit antenna array of the relay station is written as

$$\mathbf{y}_D[n] = \mathbf{G}_{RD}^H\mathbf{s}[n] + \mathbf{z}_D[n], \qquad (4)$$

where $\mathbf{z}_R \sim \mathcal{CN}(0, \mathbf{I}_N)$ and $\mathbf{z}_R \sim \mathcal{CN}(0, \mathbf{I}_K)$ are the additive white Gaussian noises (AWGNs) at the relay station and the K destinations, respectively. Note that $u[n]$ is a vector whose k-th element is $u_k[n]$, and the vector $s[n]$ expresses the transmitted signal from relay to destinations. For the sake of complexity, we assume that the relay station applies linear processing, i.e., the relay station achieves the decoding of the transmitted signals from the K sources by employing a linear receiver, and at the same time, the relay forwards the signals to the K destinations by using linear precoding. In the general case, the linear decoder and precoder are given by $\mathbf{W}^H$ and $\mathbf{F}$, respectively. Specifically, the received signal is seperated into K streams after multiplication with the linear receiver $\mathbf{W}^H$ according to

$$\begin{aligned}\mathbf{r}[n] &= \mathbf{W}^H\mathbf{y}_R[n] \\ &= \sqrt{p_S}\mathbf{W}^H\mathbf{G}_{SR}\mathbf{u}[n] + \mathbf{W}^H\mathbf{G}_{RR}\mathbf{s}[n] + \mathbf{W}^H\mathbf{z}_R[n].\end{aligned} \quad (5)$$

The kth element of $\mathbf{r}[n]$, or equivalently, the kth stream enables the decoding of the signal transmitted from the kth source S_k. More precisely, we have

$$\begin{aligned}\mathbf{r}_k[n] &= \sqrt{p_S}\mathbf{w}_k^H\mathbf{g}_{SR,k}u_k[n] + \sqrt{p_S}\sum_{j\neq k}^K \mathbf{w}_k^H\mathbf{g}_{SR,j}u_j[n] \\ &+ \mathbf{w}_k^H\mathbf{G}_{RR}\mathbf{s}[n] + \mathbf{w}_k^H\mathbf{z}_R[n],\end{aligned} \quad (6)$$

where the first and second terms represent the desired signal and the interpair interference, while the third and last term express the SI and the post-processed noise. Note that $\mathbf{g}_{SR,k}$ and $\mathbf{w}_k$ are the kth columns of $\mathbf{G}_{SR}$ and $\mathbf{W}$, respectively.

Having detected the signals transmitted from the K sources, the relay station employs linear precoding to process them. Then, the relay station broadcasts the signals to all K destinations. If we assume that the processing delay is equal to $d \geq 1$, we have[4]

$$\mathbf{s}[n] = \mathbf{u}[n-d], \qquad (7)$$

---

[4]This common assumption in the existing literature for FD systems, enables us to assume that at a given time instant, the receive and transmit signals at the relay station are uncorrelated. Also, we assume that the relay can obtain the source signals without any error. Otherwise, $s[n]$ in (7) would include a noise term.

where $\mathbf{u}[n-d]$ includes the linear precoding matrix. By substituting of (7) into (4), we obtain the received signal at D_k as

$$y_{D,k}[n] = \mathbf{g}_{RD,k}^H\mathbf{u}[n-d] + z_{D,k}[n] \qquad (8)$$

with $\mathbf{g}_{RD,k}$ being the kth column of $\mathbf{G}_{RD}$, while $z_{D,k}$ is the kth element of $\mathbf{z}_D$.

Choosing MMSE/RZF processing, i.e., employing MMSE for the decoder and RZF for the precoder, we achieve to maximize the received SNR by not taking into account the interpair interference [16]. In other words, MMSE and RZF behave quite well. Hereafter, we omit the time index from our analysis for the sake of simplicity.

B. Pilot Training Phase

In practical systems, the relay station has to estimate both the channels $\mathbf{G}_{SR}$ and $\mathbf{G}_{RD}$. A good transmission protocol to implement the current design is TDD, which is the most favorable scheme for massive MIMO. According to TDD, the protocol consists of coherence blocks having duration of T channel uses. In turn, each block is split into $\tau \geq 2K$ training pilot symbols to guarantee that the source and the destination user elements (UEs) are spatially separable by the relay station and the remaining channel uses are allocated for the data transmission symbols[5]. Note that during the data transmission phase, the channel is known due to the property of the channel reciprocity. After sending the pilots, the received signal matrices at the receive and transmit antennas of the relay are given by

$$\mathbf{Y}_r^{tr} = \sqrt{\tau p_{tr}}\left(\mathbf{G}_{SR}\mathbf{\Phi}_S + \bar{\mathbf{G}}_{RD}\mathbf{\Phi}_D\right) + \mathbf{Z}_r^{tr}, \qquad (9)$$

$$\mathbf{Y}_t^{tr} = \sqrt{\tau p_{tr}}\left(\bar{\mathbf{G}}_{SR}\mathbf{\Phi}_S + \mathbf{G}_{RD}\mathbf{\Phi}_D\right) + \mathbf{Z}_t^{tr}, \qquad (10)$$

where the channel matrices from the K sources to the transmit antenna array of the relay station and from the K destinations to the receive antenna array of the relay station are given by $\bar{\mathbf{G}}_{SR} \in \mathbb{C}^{M\times K}$ and $\bar{\mathbf{G}}_{RD} \in \mathbb{C}^{N\times K}$, respectively. Similarly, $\mathbf{Z}_r^{tr}$ and $\mathbf{Z}_t^{tr}$ denote AWGN matrices having i.i.d. $\mathcal{CN}(0,1)$ elements. Also, the kth rows of $\mathbf{\Phi}_S \in \mathbb{C}^{K\times\tau}$ and $\mathbf{\Phi}_D \in \mathbb{C}^{K\times\tau}$ are the pilot sequences transmitted from the corresponding source and destination users, i.e., $\mathbf{S}_k$ and $\mathbf{D}_k$. Actually, we assume that all the pilot sequences are pairwisely orthogonal, which requires that $\tau \geq 2K$, since $\mathbf{\Phi}_S\mathbf{\Phi}_S^H = \mathbf{I}_K$, $\mathbf{\Phi}_D\mathbf{\Phi}_D^H = \mathbf{I}_K$, and $\mathbf{\Phi}_S\mathbf{\Phi}_D^H = \mathbf{0}$. Note that p_{tr} denotes the transmit power of each pilot symbol.

Under the assumption that the relay station applies minimum mean square-error (MMSE) estimation to estimate the channels $\mathbf{G}_{SR}$ and $\mathbf{G}_{RD}$, the estimated channels can be written by following the corresponding procedure in [31] as

$$\begin{aligned}\hat{\mathbf{G}}_{SR} &= \frac{1}{\sqrt{\tau p_{tr}}}\mathbf{Y}_r^{tr}\mathbf{\Phi}_S^H\tilde{\mathbf{D}}_{SR} \\ &= \mathbf{G}_{SR}\tilde{\mathbf{D}}_{SR} + \frac{1}{\sqrt{\tau p_{tr}}}\mathbf{N}_S\tilde{\mathbf{D}}_{SR}\end{aligned} \quad (11)$$

---

[5]The pilot sequences of τ symbols are transmitted simultaneously by all the sources and destinations.

and

$$\hat{\mathbf{G}}_{\mathrm{RD}} = \frac{1}{\sqrt{\tau p_{\mathrm{tr}}}} \mathbf{Y}_{\mathrm{t}}^{\mathrm{tr}} \boldsymbol{\Phi}_{\mathrm{D}}^{\mathsf{H}} \tilde{\mathbf{D}}_{\mathrm{RD}}$$

$$= \mathbf{G}_{\mathrm{RD}} \tilde{\mathbf{D}}_{\mathrm{RD}} + \frac{1}{\sqrt{\tau p_{\mathrm{tr}}}} \mathbf{N}_{\mathrm{D}} \tilde{\mathbf{D}}_{\mathrm{RD}}, \qquad (12)$$

where $\mathbf{N}_{\mathrm{S}} = \mathbf{Z}_{\mathrm{t}}^{\mathrm{tr}} \boldsymbol{\Phi}_{\mathrm{D}}^{\mathsf{H}}$ and $\mathbf{N}_{\mathrm{D}} = \mathbf{Z}_{\mathrm{r}}^{\mathrm{tr}} \boldsymbol{\Phi}_{\mathrm{S}}^{\mathsf{H}}$. In addition, we have $\tilde{\mathbf{D}}_{\mathrm{SR}} = \left(\tilde{\mathbf{D}}_{\mathrm{SR}}^{-1} / (\tau p_{\mathrm{tr}}) + \mathbf{I}_K \right)$ and $\tilde{\mathbf{D}}_{\mathrm{RD}} = \left(\tilde{\mathbf{D}}_{\mathrm{RD}}^{-1} / (\tau p_{\mathrm{tr}}) + \mathbf{I}_K \right)$. Given that the rows of $\boldsymbol{\Phi}_{\mathrm{S}}$ and $\boldsymbol{\Phi}_{\mathrm{D}}$ are pairwisely orthogonal, the elements of $\mathbf{N}_{\mathrm{S}}$ and $\mathbf{N}_{\mathrm{D}}$ are i.i.d. obeying to the $\mathcal{CN}(0,1)$ distribution.

Taking into account the property of orthogonality of MMSE estimation, we decompose the current channels in terms of the estimated channels as [35]

$$\mathbf{G}_{\mathrm{SR}} = \hat{\mathbf{G}}_{\mathrm{SR}} + \mathbf{E}_{\mathrm{SR}} \qquad (13)$$

$$\mathbf{G}_{\mathrm{RD}} = \hat{\mathbf{G}}_{\mathrm{RD}} + \mathbf{E}_{\mathrm{RD}}, \qquad (14)$$

where $\mathbf{E}_{\mathrm{SR}}$ and $\mathbf{E}_{\mathrm{RD}}$ are the estimation error matrices of $\mathbf{G}_{\mathrm{SR}}$ and $\mathbf{G}_{\mathrm{RD}}$. Actually, the rows of $\hat{\mathbf{G}}_{\mathrm{SR}}$, $\mathbf{E}_{\mathrm{SR}}$, $\hat{\mathbf{G}}_{\mathrm{RD}}$, and $\mathbf{E}_{\mathrm{RD}}$ are mutually independent and distributed as $\mathcal{CN}\left(\mathbf{0}, \hat{\mathbf{D}}_{\mathrm{SR}}\right)$, $\mathcal{CN}\left(\mathbf{0}, \mathbf{D}_{\mathrm{SR}} - \hat{\mathbf{D}}_{\mathrm{SR}}\right)$, $\mathcal{CN}\left(\mathbf{0}, \hat{\mathbf{D}}_{\mathrm{RD}}\right)$, and $\mathcal{CN}\left(\mathbf{0}, \mathbf{D}_{\mathrm{RD}} - \hat{\mathbf{D}}_{\mathrm{RD}}\right)$. Note that $\hat{\mathbf{D}}_{\mathrm{SR}}$ and $\hat{\mathbf{D}}_{\mathrm{RD}}$ are diagonal matrices with $\left[\hat{\mathbf{D}}_{\mathrm{SR}}\right]_{kk} = \sigma_{\mathrm{SR},k}^2$ and $\left[\hat{\mathbf{D}}_{\mathrm{RD}}\right]_{kk} = \sigma_{\mathrm{RD},k}^2$ being the diagonal elements of $\hat{\mathbf{D}}_{\mathrm{SR}}$, and $\hat{\mathbf{D}}_{\mathrm{RD}}$, which are equal to $\sigma_{\mathrm{SR},k}^2 = \tau p_{\mathrm{tr}} \beta_{\mathrm{SR},k}^2 / (\tau p_{\mathrm{tr}} \beta_{\mathrm{SR},k} + 1)$ and $\sigma_{\mathrm{RD},k}^2 = \tau p_{\mathrm{tr}} \beta_{\mathrm{RD},k}^2 / (\tau p_{\mathrm{tr}} \beta_{\mathrm{RD},k} + 1)$, respectively.

C. RS Approach

After having described the conventional multipair with relay transmission (NoRS) in Section III, we focus on the application of the promising RS transmission method that is going to be applied in the second link between the relay station and the destination users. Below, we provide shortly its presentation.

The main benefit of the RS transmission, taking place in multi-user scenarios, is the achievement of unsaturated sum-rate with increasing SNR despite the presence of imperfect CSIT as was shown in [18], [19], [22], [36]. The NoRS strategy treats as noise every multi-user interference originating from the imperfect CSIT. On the other hand, the RS strategy is able to bridge treating interference as noise and perform interference decoding through the presence of a common message. Thus, the key to boost the sum-rate performance is the ability to decode part of the interference[6]. This observation motivates us to investigate the potential benefits of RS in the presence of the SI, since the SI has the effect of altering the CSI between the estimation stage and the transmission stage.

According to the RS method, the message, intended for destination UE k, is split into two parts, namely, the common and private parts. Regarding the common part, it is drawn from

---

[6]At the time of demodulation, a user needs to know the precoded channel to perform coherent demodulation. Actually, the user does not need to know the channel itself and the precoder itself, but just the inner product of the two, i.e. the precoded channel. The same action takes place in conventional MU-MIMO. In practice, this is achieved through the use of demodulation reference signals, called DMRS in LTE-A [37].

a public codebook and it has to be decoded by all UEs with zero error probability. As far as the private part is concerned, it has to be decoded only by destination UE k. It is worthwhile to mention that the messages, intended for the other UEs, consist of a private part only. In mathematical terms, the transmit signal is written as

$$\mathbf{u} = \underbrace{\sqrt{\rho_{\mathrm{c}}} \mathbf{f}_c u_c}_{\text{common part}} + \underbrace{\sum_{k=1}^{K} \sqrt{\rho_{\mathrm{k}}} \mathbf{f}_k u_k}_{\text{private part}}, \qquad (15)$$

where u_c and u_k are the common and the private messages for UE k, while $\mathbf{f}_c$ denotes the precoding vector of the common message with unit norm and $\mathbf{f}_k$ is the linear precoder corresponding to UE k. More concretely, the private message $u_k \ \forall k$ is superimposed over the common message u_c and sent with linear precoding. In addition, ρ_{c} is the power allocated to the common message. Regarding the decoding procedure, the first step is the decoding of the common message by each UE, while all private messages are treated as noise. The next step includes the subtraction of the contribution of the common message in the received signal by each UE, and thus, each UE is able to decode its own private message. Herein, we focus on the application of the RZF precoder for the private messages, as mentioned before.

Remark 1 (Conventional Transmission (NoRS Approach)): According to the conventional approach, there is no common message transmission. Thus, since no common part exists, (15) degenerates to

$$\mathbf{u} = \sum_{k=1}^{K} \sqrt{\rho_{\mathrm{k}}} \mathbf{f}_k u_k, \qquad (16)$$

where λ is a normalization parameter inside $\mathbf{f}_k$ given by $\lambda = \frac{K}{\mathbb{E}[\mathrm{tr}\,\mathbf{F}^{\mathsf{H}}\mathbf{F}]}$.

IV. END-TO-END ACHIEVABLE RATE

This section considers the presentation of the transmission between the kth source user and the corresponding destination user through the multiple antennas relay station, i.e., $(\mathrm{S}_k \rightarrow \mathrm{R} \rightarrow \mathrm{D}_k)$. Reasonably, this rate depends on the weakest link between the two hops, or else, this rate is limited by the minimum of the achievable rates of the two links [10]. More concretely, the achievable user rate from end-to-end is given by

$$R_k = \min\{R_{\mathrm{SR},k}, R_{\mathrm{RD},k}\}, \qquad (17)$$

where $R_{\mathrm{SR},k}$ and $R_{\mathrm{RD},k}$ denote the achievable rates of the corresponding links.

In the first transmission link, a conventional MAC is considered with an MMSE decoder at the relay station, while, in the second hop, we employ the RS scheme with an RZF precoder for the transmission of the private messages.

A. $\mathrm{S}_k \rightarrow \mathrm{R}$ *(Conventional Transmission)*

During the first hop, we set $p_{\mathrm{S}} = \rho$, where ρ refers to the SNR, since the AWGN is assumed to have unit variance. Thus,

the SINR of the source UE k is expressed by means of (6) as

$$\gamma_{\mathrm{SR},k} = \frac{\rho|\mathbf{w}_k^{\mathsf{H}}\mathbf{g}_{\mathrm{SR},k}|^2}{\sum_{j\neq k}^K \rho|\mathbf{w}_k^{\mathsf{H}}\mathbf{g}_{\mathrm{SR},j}|^2 + \|\mathbf{w}_k^{\mathsf{H}}\mathbf{G}_{\mathrm{RR}}\|^2 + \|\mathbf{w}_k^{\mathsf{H}}\|^2}. \quad (18)$$

Note that we have relied on the worst-case assumption by treating the multi-user interference and distortion noises as independent Gaussian noises [38]. According to this SINR, we obtain the achievable sum-rate, being a lower bound of the mutual information between the received signal and the transmitted symbols, as

$$\mathrm{R}_{\mathrm{SR}} = \sum_{k=1}^K \mathrm{R}_{\mathrm{SR},k}, \quad (19)$$

where $\mathrm{R}_{\mathrm{SR},k} = \frac{T-\tau}{T}\log_2(1+\gamma_{\mathrm{SR},k})$.

B. $\mathrm{R} \to \mathrm{D}_k$

During the second link, we employ the RS transmission scheme, in order to mitigate the saturation of the system at high SNR. Specifically, we apply uniform power allocation for the private messages, however, the power allocated to the common part is different. The allocation scheme assumes $\rho_{\mathrm{c}} = \rho(1-t)$ to the common message and $\rho_k = \rho t/K$ to the private message of each UE, where $t \in (0,1]$. The t parameter is used to adjust the fraction of the total power spent on the transmission of the private messages.

Following the RS principles, we have to evaluate the SINRs of both common and private messages. Assuming that perfect CSI is available at the receivers and given that the transmit signal is given by (15), the corresponding SINRs are given by

$$\gamma_{\mathrm{RD},k}^{\mathrm{c}} = \frac{\rho_{\mathrm{c}}|\mathbf{g}_{\mathrm{RD},k}^{\mathsf{H}}\mathbf{f}_{\mathrm{c}}|^2}{\sum_{j=1}^K \frac{\rho t}{K}|\mathbf{g}_{\mathrm{RD},k}^{\mathsf{H}}\mathbf{f}_j|^2 + 1} \quad (20)$$

$$\gamma_{\mathrm{RD}}^{\mathrm{c}} = \min_k\left(\gamma_{\mathrm{RD},k}^{\mathrm{c}}\right) \quad (21)$$

$$\gamma_{\mathrm{RD},k}^{\mathrm{p}} = \frac{\frac{\rho t}{K}|\mathbf{g}_{\mathrm{RD},k}^{\mathsf{H}}\mathbf{f}_k|^2}{\sum_{j\neq k}^K \frac{\rho t}{K}|\mathbf{g}_{\mathrm{RD},k}^{\mathsf{H}}\mathbf{f}_j|^2 + 1}. \quad (22)$$

Note that $\gamma_{\mathrm{RD}}^{\mathrm{c}} = \min_k\left(\gamma_{\mathrm{RD},k}^{\mathrm{c}}\right)$ and $\gamma_{\mathrm{RD},k}^{\mathrm{p}}$ correspond to the SINRs of the common and private messages, respectively. In this case, the achievable sum-rate is written as

$$\mathrm{R}_{\mathrm{RD}} = \mathrm{R}_{\mathrm{RD}}^{\mathrm{c}} + \sum_{j=1}^K \mathrm{R}_{\mathrm{RD},j}^{\mathrm{p}}, \quad (23)$$

where, similar to (19), we have $\mathrm{R}_{\mathrm{RD}}^{\mathrm{c}} = \frac{T-\tau}{T}\log_2(1+\gamma_{\mathrm{RD}}^{\mathrm{c}})$ and $\mathrm{R}_{\mathrm{RD},k}^{\mathrm{p}} = \frac{T-\tau}{T}\log_2\left(1+\gamma_{\mathrm{RD},k}^{\mathrm{p}}\right)$ corresponding to the achievable rates of the common and private messages, respectively.

V. DETERMINISTIC EQUIVALENT PERFORMANCE ANALYSIS

The DEs of the SINRs for both links are such that for each link it holds that $\gamma_k - \bar\gamma_k \xrightarrow[M\to\infty]{\text{a.s.}} 0$[7], where γ_k is the SINR of the kth user and $\bar\gamma_k$ is the corresponding DE. In this direction,

[7]Note that $\xrightarrow[M\to\infty]{\text{a.s.}}$ denotes almost sure convergence, and $a_n \asymp b_n$ expresses the equivalence relation $a_n - b_n \xrightarrow[M\to\infty]{\text{a.s.}} 0$ with a_n and b_n being two infinite sequences.

the corresponding deterministic rate of UE k is obtained by the dominated convergence [39] and the continuous mapping theorem [40] by means of (19), (23) for both links as

$$R_{\mathrm{i},k} - \bar{R}_{\mathrm{i},k} \xrightarrow[M\to\infty]{\text{a.s.}} 0, \quad \mathrm{i} = \mathrm{SR}, \mathrm{RD} \quad (24)$$

where $\bar{R}_{\mathrm{i},k}$ is the DE $R_{\mathrm{i},k}$.

A. DE of the Achievable Rate of the First Hop ($\mathrm{S}_k \to \mathrm{R}$)

The design of the first hop, being basically a MAC, follows a standard uplink transmission scheme. We choose the MMSE linear decoder, in order to keep the implementation complexity to a reasonable level and at the same time achieve a high rate.

The MMSE decoder is designed by means of the channel estimate $\hat{\mathbf{G}}_{\mathrm{SR}}$, as [16]

$$\mathbf{W}_{\mathrm{SR}} = \left(\hat{\mathbf{W}}_{\mathrm{SR}} + \mathbf{Z}_{\mathrm{SR}} + N\alpha_{\mathrm{RSR}}\mathbf{I}_M\right)^{-1}\hat{\mathbf{G}}_{\mathrm{SR}}, \quad (25)$$

where we define

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{SR}} \triangleq \left(\hat{\mathbf{W}}_{\mathrm{SR}} + \mathbf{Z}_{\mathrm{SR}} + N\alpha_{\mathrm{SR}}\mathbf{I}_M\right)^{-1} \quad (26)$$

with $\hat{\mathbf{W}}_{\mathrm{SR}} \triangleq \hat{\mathbf{G}}_{\mathrm{SR}}\hat{\mathbf{G}}_{\mathrm{SR}}^{\mathsf{H}}$. The matrix $\mathbf{Z}_{\mathrm{SR}} \in \mathbb{C}^{N\times N}$ is an arbitrary Hermitian nonnegative definite matrix and α_{SR} is a regularization parameter scaled by N, in order to converge to a constant, as N, $K \to \infty$. Although α_{RD} and $\mathbf{Z}_{\mathrm{RD}}$ can be optimized, this is outside the scope of this paper and we leave it for future work.

The data transmission during this hop has a duration of $T - \tau$ time slots. The DE of the k user rate, when K, N go to infinity with a given ration $\beta = N/K$, is provided by the following theorem.

Theorem 1: The DE of the SINR of UE k for the first link of a multipair FD system with MMSE decoding and imperfect CSIT is given by (27), where

$$\mu_{jk} \asymp \frac{\delta_j'}{N} + \frac{\left|\delta_k''\right|^2 \delta_k'}{N(1+\delta_k)^2} - 2\mathrm{Re}\left\{\frac{\delta_k''\delta_k'}{N(1+\delta_k)}\right\}. \quad (28)$$

Also, we have $\delta_{\mathrm{SR},k} = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},k}$, $\delta_{\mathrm{SR},k}' = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},k}'$, $\delta_{\mathrm{SR},k}'' = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},k}''$, $\delta_{\mathrm{SR},jk} = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},jk}$, $\delta_{\mathrm{SR},jk}' = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},j}\mathbf{T}_{\mathrm{SR},jk}'$, $\delta_{jk}'' = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},jk}\mathbf{T}_{\mathrm{SR},jk}$, $\mathbf{S} = \mathbf{Z}_{\mathrm{SR}}/N$, and $\tilde{a} = \alpha_{\mathrm{SR}}$ where

* $\mathbf{T}_{\mathrm{SR},k} = \mathbf{T}_{\mathrm{SR},k}(\tilde{a})$ and $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_K]^{\mathsf{T}} = \boldsymbol{\delta}(\tilde{a}) = e(\tilde{a})$ are given by [41, Thm. 1] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D} = \hat{\mathbf{D}}_{\mathrm{SR},k}$, $\mathbf{L} = \hat{\mathbf{D}}_{\mathrm{SR},k}$ $\forall k \in \mathcal{K}$,
* $\mathbf{T}_{\mathrm{SR},k}' = \mathbf{T}_{\mathrm{SR},k}'(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D} = \hat{\mathbf{D}}_{\mathrm{SR},k}$, $\mathbf{K} = \mathbf{D}_{\mathrm{SR},k} - \hat{\mathbf{D}}_{\mathrm{SR},k}, \forall k \in \mathcal{K}$,
* $\mathbf{T}_{\mathrm{SR},k}'' = \mathbf{T}_{\mathrm{SR},k}''(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D} = \hat{\mathbf{D}}_{\mathrm{SR},k}$, $\mathbf{K} = \mathbf{I}_N, \forall k \in \mathcal{K}$,
* $\mathbf{T}_{\mathrm{RR}} = \mathbf{T}_{\mathrm{RR}}(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D} = \mathbf{I}_N$, $\mathbf{K} = \sigma_{\mathrm{SI}}^2\mathbf{I}_N, \forall k \in \mathcal{K}$,
* $\mathbf{T}_{\mathrm{SR},jk} = \mathbf{T}_{\mathrm{SR},jk}(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D} = \hat{\mathbf{D}}_{\mathrm{SR},k}$, $\mathbf{K} = \hat{\mathbf{D}}_{\mathrm{SR},k}, \forall k \in \mathcal{K}$,
* $\mathbf{T}_{\mathrm{SR},jk} = \mathbf{T}_{\mathrm{SR},jk}'(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D} = \hat{\mathbf{D}}_{\mathrm{SR},k}$, $\mathbf{K} = \hat{\mathbf{D}}_{\mathrm{SR},k}$, $\mathbf{L} = \hat{\mathbf{D}}_{\mathrm{SR},j} \forall k \in \mathcal{K}$.
* $\mathbf{T}_{\mathrm{SR},jk} = \mathbf{T}_{\mathrm{SR},jk}''(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D} = \hat{\mathbf{D}}_{\mathrm{SR},k}$, $\mathbf{K} = \hat{\mathbf{D}}_{\mathrm{SR},k}$, $\mathbf{L} = \hat{\mathbf{D}}_{\mathrm{SR},k} \forall k \in \mathcal{K}$.

Proof: The proof of Theorem 1 is given in Appendix A. ∎

$$\bar{\gamma}_{\text{SR},k} = \frac{\rho \delta_{\text{SR},k}^2}{\rho \frac{1}{N} \delta'_{\text{SR},k} + \frac{1}{N} \delta''_{\text{SR},k} + \frac{1}{N} \operatorname{tr} \mathbf{T}_{\text{RR}} + \rho \sum_{j \neq k}^{K} \frac{\mu_{\text{SR},jk}}{N}} \tag{27}$$

B. DE of the Achievable Rate of the Second Hop with RS ($\text{R} \to \text{D}_k$)

This section presents the DE of the user rate during the data transmission with RS in the second link, which takes place for $T - \tau$ time slots. In fact, we derive the DE of the kth UE in the asymptotic limit of K, M for fixed ratio $\zeta = K/M$.

In addition, we provide the precoder design for the common message, implemented to be used under the RS approach. Moreover, among the main results, we present the DEs of the SINRs characterizing the transmissions of the common and the private messages of UE k.

C. Precoder Design

For the sake of simplicity, we employ linear precoding during the application of the RS method. Actually, the RS method includes two different types of precoders for the transmission of the private and common messages, respectively. In the case of a MISO broadcast channel (BC) with imperfect CSI, the optimal precoder has to be optimized numerically [22]. However, we consider that the transmission of the private message takes place by using RZF due to the prohibitive complexity, as mentioned in a previous section[8]. Further elaboration follows.

1) Precoding of the Private Messages: Given that the complexity increases in large MIMO systems as $M \to \infty$, the choice of RZF for the transmission of the private messages is the prevailing solution. In such case, the relay station implements its RZF precoder, constructed by means of the channel estimate $\hat{\mathbf{G}}_{\text{RD}}$, as [16]

$$\begin{aligned} \mathbf{F}_{\text{RD}} &= \sqrt{\lambda} \left(\hat{\mathbf{W}}_{\text{RD}} + \mathbf{Z}_{\text{RD}} + M\alpha_{\text{RD}} \mathbf{I}_M \right)^{-1} \hat{\mathbf{G}}_{\text{RD}} \\ &= \sqrt{\lambda} \hat{\mathbf{\Sigma}}_{\text{RD}} \hat{\mathbf{G}}_{\text{RD}}, \end{aligned} \tag{29}$$

where we define

$$\hat{\mathbf{\Sigma}}_{\text{RD}} \triangleq \left(\hat{\mathbf{W}}_{\text{RD}} + \mathbf{Z}_{\text{RD}} + M\alpha_{\text{RD}} \mathbf{I}_M \right)^{-1} \tag{30}$$

with $\hat{\mathbf{W}}_{\text{RD}} \triangleq \hat{\mathbf{G}}_{\text{RD}} \hat{\mathbf{G}}_{\text{RD}}^{\mathsf{H}}$ and λ being a normalization parameter that satisfies $\lambda = \frac{K}{\mathbb{E}\left[\operatorname{tr} \mathbf{F}_{\text{RD}}^{\mathsf{H}} \mathbf{F}_{\text{RD}} \right]}$, which is a long-term total transmit power constraint at the relay. Similar to the definition of the MMSE decoder, $\mathbf{Z}_{\text{RD}} \in \mathbb{C}^{M \times M}$ is an arbitrary Hermitian nonnegative definite matrix and α_{RD} is a regularization parameter scaled by M, in order to converge to a constant, as $M, K \to \infty$. In addition, α_{RD} and $\mathbf{Z}_{\text{RD}}$ can be optimized as well, but this is outside the scope of this paper and we leave it for future work.

[8]Note that an extra gain of RS over NoRS can be achieved by jointly optimizing the power allocation as well as the precoders of the common and private messages [22]. However it is not really practical to resort to this type of optimization for large-scale systems such as massive MIMO, where the use of deterministic equivalent analysis is commonly used in order to get some further insight into the system behaviour in terms of different aspects such as the impact of hardware impairments [20].

2) Precoding of the Common Message: Herein, we provide the design of the precoder $\mathbf{f}_c$ of the common message by following a similar procedure to [19]. In particular, taking into account that in the large number of antennas regime the different channel estimates tend to be orthogonal, we express $\mathbf{f}_c$ as a linear sum of these channel estimates in the subspace of $\hat{\mathbf{G}}_{\text{RD}}$, $\mathcal{S} = \operatorname{Span}\left(\hat{\mathbf{G}}_{\text{RD}} \right)$. In other words, $\mathbf{f}_c$ is designed as a weighted matched beamforming. Mathematically, we have

$$\mathbf{f}_c = \sum_k \alpha_k \hat{\mathbf{g}}_{\text{RD},k}. \tag{31}$$

The objective is to maximize the achievable rate of the common message $\text{R}_{\text{RD},k}^c$. This optimization problem is formed as

$$\begin{aligned} \mathcal{P}_1 \ : \ & \max_{\mathbf{f}_c \in \mathcal{S}} \min_k q_{\text{RD},k} |\mathbf{g}_{\text{RD},k}^{\mathsf{H}} \mathbf{f}_c|^2, \\ & \text{s.t.} \quad \|\mathbf{f}_c\|^2 = 1 \end{aligned} \tag{32}$$

where $q_{\text{RD},k} = \frac{\rho_c \lambda}{\lambda \sum_{j=1}^{K} \frac{\rho}{K} |\mathbf{g}_{\text{RD},k}^{\mathsf{H}} \mathbf{f}_j|^2 + 1}$. The optimal $\{\alpha_k^*\}$ is yielded by the following proposition.

Proposition 1: In the large system limit, the optimal solution of the practical problem set by $\mathcal{P}_1$ is given by

$$\alpha_k^* = \frac{1}{\sqrt{M \sum_{j=1}^{K} \frac{q_k \frac{1}{M^2} \operatorname{tr}^2 \hat{\mathbf{D}}_{\text{RD},k}}{q_j \frac{1}{M^2} \operatorname{tr}^2 \hat{\mathbf{D}}_{\text{RD},j}}}}, \ \forall k. \tag{33}$$

Proof: We achieve to result in an optimization problem with deterministic variables by obtaining the DEs of the equation and the constraint of $\mathcal{P}_1$. Indeed, applying Lemma [42, Lem. B.26] to (32), we have

$$\begin{aligned} \mathcal{P}_2 \ : \ & \max_{\alpha_k} \min_k q_{\text{RD},k} \frac{1}{M^2} |\alpha_k \operatorname{tr} \hat{\mathbf{D}}_{\text{RD},k}|^2, \\ & \text{s.t.} \quad \sum_k \alpha_k^2 = \frac{1}{M}. \end{aligned} \tag{34}$$

Use of Lemma 2 in [43] indicates that the optimal solution, satisfying $\mathcal{P}_2$, results, if all terms are equal. Specifically, the optimal solution is found when $q_{\text{RD},k} \alpha_k^2 \frac{1}{M^2} \operatorname{tr}^2 \hat{\mathbf{D}}_{\text{RD},k} = q_{\text{RD},j} \alpha_j^2 \frac{1}{M^2} \operatorname{tr}^2 \hat{\mathbf{D}}_{\text{RD},j}, \forall k \neq j$, and the proof is concluded. $\blacksquare$

Theorem 2: The DEs of the SINRs of UE k for the second link of a multipair FD system, corresponding to the private and common messages with RZF precoding and imperfect CSIT, are given by

$$\bar{\gamma}_{\text{RD},k}^{\text{p}} = \frac{\bar{\lambda} \frac{\rho t}{K} \delta_{\text{RD},k}^2}{\bar{\lambda} \frac{\rho t}{K} \sum_{j \neq k}^{K} \frac{Q_{\text{RD},jk}}{M} + (1 + \delta_{\text{RD},k})^2} \tag{35}$$

$$\bar{\gamma}_{\text{RD},k}^c = \frac{\rho_c \bar{\lambda} \alpha_k^2 \delta_{\text{RD},k}^2}{\bar{\lambda} \frac{\rho t}{K} \sum_{j=1}^{K} \frac{Q_{\text{RD},jk}}{M} + (1 + \delta_{\text{RD},k})^2}, \tag{36}$$

where

$$\bar{\lambda} = K \left(\frac{1}{M} \sum_{k=1}^{K} \frac{\delta'_{\text{RD},k}}{\left(1 + \delta_{\text{RD},k}\right)^2} \right)^{-1},$$

and

$$Q_{\text{RD},jk} \asymp \frac{\delta''_{\text{RD},jk}}{M} + \frac{\left| \delta'''_{\text{RD},jk} \right|^2 \delta''_{\text{RD},jk}}{M \left(1 + \delta_{\text{RD},k}\right)^2}$$
$$- 2\text{Re} \left\{ \frac{\delta'''_{\text{RD},jk} \delta''_{\text{RD},jk}}{M \left(1 + \delta_{\text{RD},k}\right)} \right\}. \qquad (37)$$

Also, we have $\delta_{\text{RD},k} = \frac{1}{M} \text{tr} \, \hat{\mathbf{D}}_{\text{RD},k} \mathbf{T}_{\text{SR},k}$, $\delta'_{\text{RD},jk} = \frac{1}{M} \text{tr} \, \hat{\mathbf{D}}_{\text{RD},k} \mathbf{T}'_{\text{SR},k}$, $\delta'_{\text{RD},jk} = \frac{1}{M} \text{tr} \, \hat{\mathbf{D}}_{\text{RD},j} \mathbf{T}'_{\text{SR},k}$, $\delta''_{\text{RD},jk} = \frac{1}{M} \text{tr} \, \hat{\mathbf{D}}_{\text{RD},jk} \mathbf{T}''_{\text{SR},k}$, $\delta'''_{\text{RD},jk} = \frac{1}{M} \text{tr} \, \hat{\mathbf{D}}_{\text{RD},jk} \mathbf{T}'''_{\text{SR},k}$, $\mathbf{S} = \mathbf{Z}_{\text{RD}}/M$, and $\tilde{a} = \alpha_{\text{RD}}$ where

* $\mathbf{T}_{\text{SR},k} = \mathbf{T}_{\text{SR},k}(\tilde{a})$ and $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_K]^{\mathsf{T}} = \boldsymbol{\delta}(\tilde{a}) = \mathbf{e}(\tilde{a})$ are given by [41, Thm. 1] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D}_{\text{SR},k} = \hat{\mathbf{D}}_{\text{SR},k} \, \forall k \in \mathcal{K}$,
* $\mathbf{T}_{\text{SR},jk} = \mathbf{T}_{\text{SR},jk}(\tilde{a})$ is given by [41, Thm. 1] for $\mathbf{S} = \mathbf{S}$, $\mathbf{D}_{\text{SR},k} = \hat{\mathbf{D}}_{\text{SR},k} \, \forall k \in \mathcal{K}$,
* $\mathbf{T}_{\text{SR},jk} = \mathbf{T}'_{\text{SR},jk}(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{L}_{\text{SR},k} = \mathbf{I}_M$, $\mathbf{D}_{\text{SR},k} = \hat{\mathbf{D}}_{\text{SR},k} \, \forall k \in \mathcal{K}$,
* $\mathbf{T}_{\text{SR},jk} = \mathbf{T}''_{\text{SR},jk}(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{K} = \hat{\mathbf{D}}_{\text{SR},j}$, $\mathbf{K} = \hat{\mathbf{D}}_{\text{SR},j}$, $\mathbf{D}_k = \hat{\mathbf{D}}_{\text{SR},k}, \forall k \in \mathcal{K}$,
* $\mathbf{T}_{\text{SR},jk} = \mathbf{T}'''_{\text{SR},jk}(\tilde{a})$ is given by [16, Thm. 2] for $\mathbf{S} = \mathbf{S}$, $\mathbf{L} = \hat{\mathbf{D}}_{\text{SR},k}$ $\mathbf{K} = \hat{\mathbf{D}}_{\text{SR},k}$, $\mathbf{D}_k = \hat{\mathbf{D}}_{\text{SR},k}, \forall k \in \mathcal{K}$.

Proof: The proof of Theorem 2 is given in Appendix B. ∎

The following remark will enable us to shed light on the interesting properties of multipair FD systems with a large number of relay station antennas during the presentation and investigation of the numerical results.

Remark 2 (Impact of increasing transmit and receive relay station antennas): According to [31], the impact of SI cancels out, when the SI is projected onto its orthogonal complement. Unfortunately, following this direction, the orthogonal projection may harm the desired signal, unless the receive antenna array grows large (tending to infinity). In such case, the channel vectors of the desired signal and the loop interference become nearly orthogonal, and actually, the impact of SI is reduced.

In a parallel path, if the size of the transmit antenna array is increased, the relay station will be able to focus its emitted energy into the proper destination users. Moreover, the transmission towards the receive antennas of the relay station is avoided. Hence, the SI reduces almost to zero.

Remark 3 (Reduction to HD transmission): Changing the system model, describing the FD transmission, to HD transmission by neglecting the SI term and changing the prelog factor in the achievable rate, we reduce to the expressions providing the DE rates of the private and common messages corresponding to (35) and (36).

D. Power Allocation

The normal method to obtain the optimal power splitting ratio t, maximizing (23), includes the derivation of the first-order derivative. However, the complicated form of the solution,

led us to follow a suboptimal power allocation method similar to [19], where RS outperforms the conventional broadcasting schemes. Interestingly, the solution allows us to extract useful observations. According to the main idea, the allocation of the fraction t results by setting the total transmit power of the private messages of RS, in order to achieve approximately the same sum rate as the conventional multi-user BC with full power. The remaining power is allocated for the transmission of the common message of RS, which boosts the sum rate. The gain in the sum-rate of the second link, achieved by the RS strategy with comparison to the NoRS transmission, is given by the difference

$$\Delta R_{\text{RD}} = \text{R}^{\text{c}}_{\text{RD}} + \sum_{k=1}^{K} \left(\text{R}^{\text{p}}_{\text{RD},k} - \text{R}^{\text{NoRS}}_{\text{RD},k} \right). \qquad (38)$$

Proposition 2: The necessary condition, described by $\text{R}^{\text{p}}_{\text{RD},k} \leq \text{R}^{\text{NoRS}}_{\text{RD},k}$, becomes equality, when the power splitting ratio t is given by

$$t = \min \left\{ \frac{K}{\rho \bar{Y}}, 1 \right\}, \qquad (39)$$

where $\bar{Y} = \dfrac{\bar{\lambda} \frac{1}{K} (\delta_{\text{RD},k})^2}{\bar{\lambda} \frac{1}{K} \sum_{j \neq k}^{K} \dfrac{Q_{\text{RD},jk}}{M} + \left(1 + \delta_{\text{RD},k}\right)^2}$. In such case, the sum-rate gain $\Delta R_{\text{RD},k}$ becomes

$$\Delta R_{\text{RD}} \geq \text{R}^{\text{c}}_{\text{RD}} - \log_2 e. \qquad (40)$$

Proof: See Appendix C. ∎

VI. NUMERICAL RESULTS

This section presents the verification of the accuracy of the derived DE expressions (analytical results) by means of comparison with the Monte Carlo simulation results. Moreover, the numerical illustrations allow to gain insights on the system performance of the considered model, and mostly on the impact of SI. In particular, the bullets represent the simulation results.

A. Simulation Setup

We consider a Rayleigh block-fading channel, where the coherence time and the coherence bandwidth are $T_c = 5 \, \text{ms}$ and $B_c = 100 \, \text{KHz}$, respectively. As a result, the coherence block consists of $T = 500$ channel uses. The simulation topology assumes $K = 10$ communication pairs, located randomly inside a disk with a diameter of $1000 \, \text{m}$. The pilot length is $\tau = 20$. In each block, we assume fast fading by means of $\mathbf{h}_{\text{SR},k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\mathbf{h}_{\text{RD},k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$. Also, we account for path-loss and shadowing, where $\mathbf{D}_{\text{RS}}$ is a $K \times K$ diagonal matrix with elements across the diagonal modeled as $\beta^m_{\text{SR},k} = \frac{10^{s_{\text{SR},k} - 1.53}}{d^{3.76}_{\text{SR},k}}$ with $d_{\text{SR},k}$ being the distance in meters between the receive antenna m at the relay station and source UE k, and $s_{\text{SR},k} \sim \mathcal{N}(0, 3.16)$ representing the shadowing effect [44]. Without loss of generality, we assume the same large-scale conditions for the second link. In addition, the power of the uplink training symbols for both links is $p_{\text{tr}} = 2 \, \text{dB}$ [16]. The number of transmit and receive antennas is $M = N = 100$. Note that these parameters hold throughout this section, unless otherwise stated.

B. Robustness of RS in FD systems?-Comparisons

The metric, we employ, to shed light on this meaningful question is the theoretical DE sum-rate and the corresponding Monte Carlo simulation. Actually, the theoretical curves are obtained by means of Theorems 1 and 2 as well as (19) and (23) by means of (24). On the other hand, the simulated curves are obtained by averaging the corresponding rate over 10^3 random channel instances. The choice of t took place by means of Proposition 2. Although the DEs are derived for K, M, and $N \to \infty$ with given ratios, they coincide with the simulated curves even for finite values K, M, and N[9].
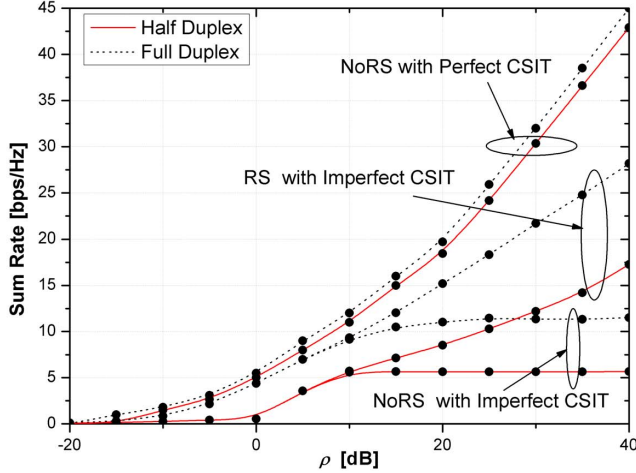

Fig. 1. Sum-rate versus ρ for different transmission techniques and comparison between HD and FD ($M = 100$, $K = 10$, $T = 500$, $p_{\mathrm{tr}} = 2$ dB, $\sigma_{\mathrm{SI}}^2 = 0$ dB).

1) Comparison between FD and HD: Fig. 1 provides the comparison between FD and HD strategies in different transmission settings, when $\sigma_{\mathrm{SI}}^2 = 0$ dB. Specifically, the dashed black lines denote the FD method, while the red solid lines depict the HD method. In all cases, FD outperforms HD as expected. In addition, we show that RS with perfect CSIT for both FD and HD increase with ρ without bound. The practical scenario, where CSIT is imperfect, is depicted by means of rate saturation in the case of NoRS, while RS provides an unsaturated rate. Although the gap between FD and HD is kept constant at high SNR in the case of NoRS, RS appears to be even more preferable at the same SNR regime, as expected. Hence, RS proves to be robust since the rate does not saturate, and it is even more preferable at high SNR. Overall, RS is beneficial for both FD and HD, but FD outperforms HD due to the prelog factor and the mitigation of SI, especially at high SNR (increasing gap with increasing ρ).

2) Varying the severity of SI: Fig. 2 presents the sum-rate versus the SNR for varying σ_{SI}^2. Actually, the solid lines correspond to RS, while the dashed lines represent the implementation of NoRS. The highest line corresponds to the

---

[9]It is shown that the simulations coincide with the DEs for $M = N = 20$ number of antennas. Therefore, the DEs provide reliable results even for low system dimensions. It should be noted that this is not a new observation. Similar observations have been made in the literature even for an 8×8 system [16], [40]–[42].

---

smallest value of σ_{SI}^2, i.e., $\sigma_{\mathrm{SI}}^2 = -1$ dB, because in FD, the least the σ_{SI}^2 is, the largest the rate becomes. When $\sigma_{\mathrm{SI}}^2 = 10$ dB, the slope is less than the other cases showing that the higher the σ_{SI}^2 is, the less capable RS is of mitigating the SI.
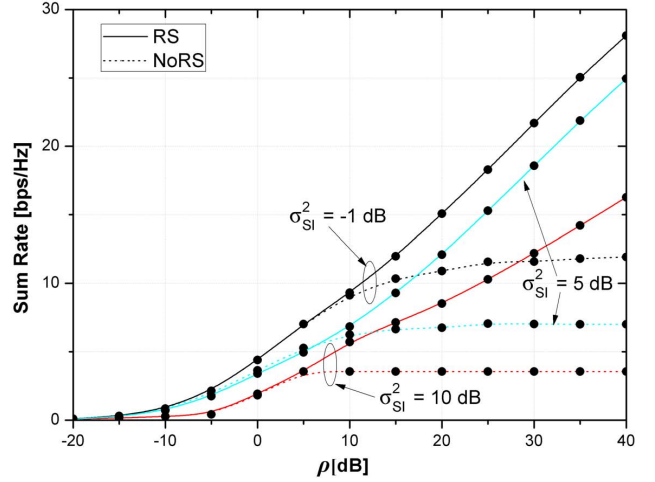

Fig. 2. Sum-rate versus ρ for varying σ_{SI}^2 ($M = 100$, $K = 10$, $T = 500$, $p_{\mathrm{tr}} = 2$ dB).

Fig. 3 illustrates the sum-rate versus the variance of the SI for $\rho = 20$ dB, when the number of transmit antennas of the relay station M increases. At the same figure, we have plotted the sum-rate corresponding to the HD case, which does not depend on σ_{SI}^2, i.e., it is constant with regard to M (parallel to the horizontal axis). Furthermore, it is exposed that, as the number M increases, the impact of SI becomes less severe, and hence, the sum-rate is higher. Until a specific value of σ_{SI}^2, the sum-rate does not change, which means that the impact of SI is negligible. After this value, the sum-rate starts decreasing. Moreover, in the case that $M = 100$, we can observe that HD appears with higher sum-rate, if the impact from σ_{SI}^2 is large enough, i.e., $\sigma_{\mathrm{SI}}^2 > 18$ dB. For the sake of comparison, we have included the plots corresponding to the NoRS transmission. It is revealed that RS provides an increase in the range of SI over which FD outperforms HD.

3) Varying the Number of Relay Station Transmit Antennas M: Fig. 4 presents the sum-rate with RS versus the SNR for increasing M. Clearly, the higher the number of transmit antennas M, the less severe the SI becomes. In such case, the sum-rate becomes larger. Moreover, we have added a line in the case that $M = 200$ corresponding to the optimum solution, in order to compare the results from the proposed sub-optimum power allocation method and the optimum solution which is obtained numerically. As can be seen, the solution provided by the sub-optimum method appears performance, which is very close to the performance obtained by means of exhaustive search. Especially, in the high-SNR regime, the two lines coincide, which shows that the sub-optimum method behaves as optimum in this region.

4) Varying the Number of Destination Users K: Fig. 5 aims at the verification that the achievable rate due to common message degrades with the number of destination users K, and also quantifies the sum-rate in the case of FD. Obviously,

increasing the number K from 5 to 8, the achievable rate degrades because the common message has to be decoded by more destination users. A solution for this, known as hierarchical rate-splitting (HRS), retains the benefits of RS, has been presented in [19]. However, the study of HRS in the case of FD is left for future work.
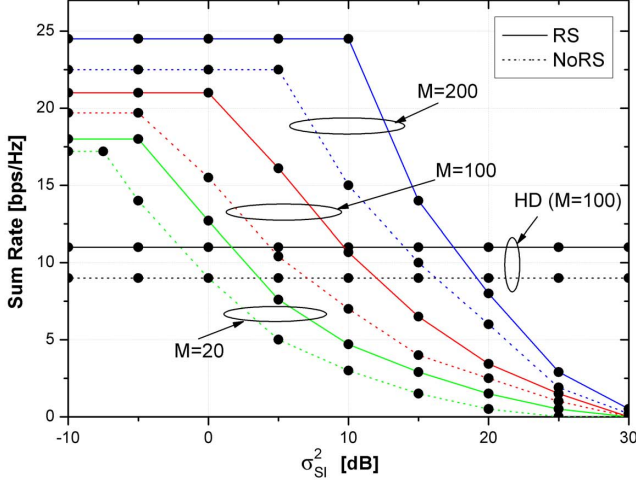

Fig. 3. Sum-rate versus σ_{SI}^2 for varying M ($\sigma_{\mathrm{SI}}^2 = 0$ dB, $K = 10$, $T = 500$, $p_{\mathrm{tr}} = 2$ dB).
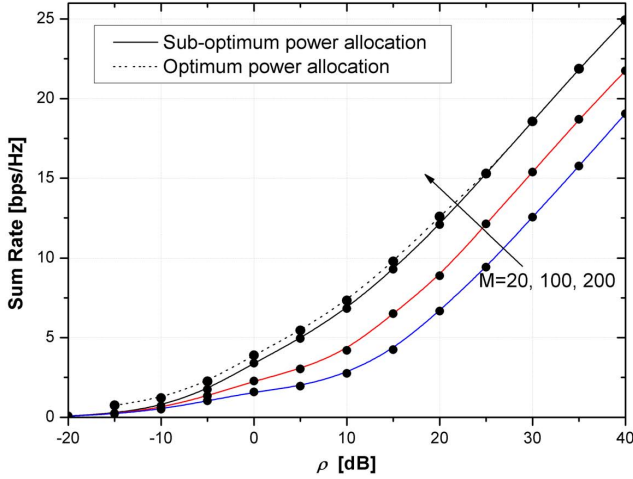
Fig. 4. Sum-rate versus ρ for varying M ($K = 10$, $T = 500$, $p_{\mathrm{tr}} = 2$ dB).

Fig. 5. Sum-rate versus ρ for varying K ($M = 100$, $T = 500$, $p_{\mathrm{tr}} = 2$ dB).

Initially, we presented the RS method and the FD approach. Next, by considering realistic channels with imperfect CSIT, we obtained the estimated channels of the first and the second links. Having applied a conventional multi-user uplink design for the first link, we assumed the implementation of RS in the second link. Actually, we provided the DE analysis of the achievable rate for the first link, and then, for the second one, where the precoders for the common and private messages were designed based on RS. Notably, the validation of the analytical results was demonstrated by means of simulations that depicted that the asymptotic results can be applicable even for systems of finite dimensions. Remarkably, RS proved to be robust in both cases of HD and FD relaying. Furthermore, among the interesting outcomes of this paper, it was extracted that by increasing the number of relay antennas or by decreasing the severity of SI, RS appears to be more robust. After a certain value of the SI, this property of RS degrades. Furthermore, we showed that in the case of a dense environment with increasing number of users, the ability of RS to tackle SI and multi-user saturation worsens because the common message has to be decoded by more users. As a future work, we plan to focus on the robustness of RS in different system models with altered CSIT. For example, we plan to implement the RS strategy in millimeter wave systems with imperfect CSIT that consider hybrid precoding.

VII. CONCLUSIONS

RS achieves to mitigate the degradation emerged in multi-user systems with imperfect CSIT. Motivated by this observation, we proposed the RS strategy to tackle the saturation occured in multi-pair MIMO relay systems with imperfect CSIT. Interestingly, we considered relay stations employing a large number of antennas (massive MIMO), in order to address the perfomance issues of the forthcoming 5G networks. Specifically, the objective of this work was to examine the potential robustness of the RS transmission method in multi-pair massive MIMO relay systems obeying mostly to the FD design.
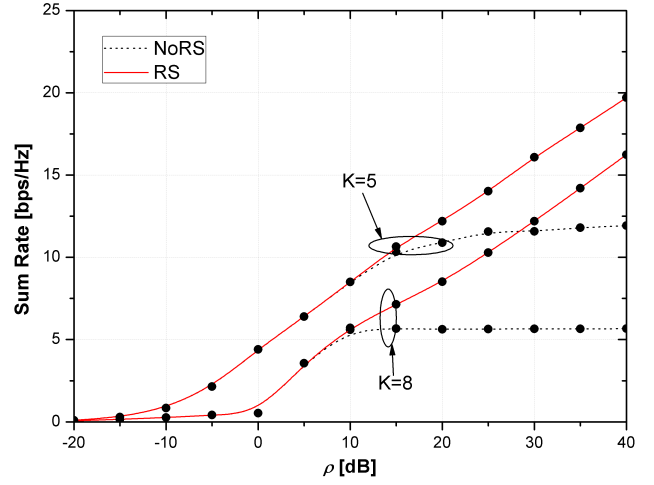
APPENDIX A
PROOF OF THEOREM 1

After appropriate substitution of the MMSE decoder and scaling by $\frac{1}{N}$, the desired signal becomes

$$\frac{1}{N}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}_{\mathrm{SR}}\mathbf{g}_{\mathrm{SR},k} = \frac{1}{N}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}_{\mathrm{SR}}\hat{\mathbf{g}}_{\mathrm{SR},k} \tag{41}$$

$$= \frac{\frac{1}{N}\operatorname{tr}\hat{\boldsymbol{\Sigma}}_{\mathrm{SR},k}\hat{\mathbf{D}}_{\mathrm{SR},k}}{1 + \frac{1}{N}\operatorname{tr}\hat{\boldsymbol{\Sigma}}_{\mathrm{SR},k}\hat{\mathbf{D}}_{\mathrm{SR},k}} \tag{42}$$

$$= \frac{\frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},k}}{1 + \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},k}}, \tag{43}$$

where we have used [45, Eq. 2.2], and $\hat{\Sigma}_{\mathrm{SR},k}$ is defined as $\hat{\Sigma}_{\mathrm{SR},k} = \left(\hat{\mathbf{G}}_{\mathrm{SR}}\hat{\mathbf{G}}_{\mathrm{SR}}^{\mathsf{H}} - \hat{\mathbf{g}}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}} + \mathbf{Z}_{\mathrm{SR}} + N\alpha_{\mathrm{SR}}\mathbf{I}_N\right)^{-1}$. Applying [42, Lem. B.26] and [41, Thm. 1], we have $\frac{1}{N}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR}}\hat{\mathbf{g}}_{\mathrm{SR},k} \asymp \frac{\delta_{\mathrm{SR},k}}{1+\delta_{\mathrm{SR},k}} \xrightarrow{\text{a.s.}}_{N_t\to\infty} 0$, where $\delta_{\mathrm{SR},k} = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},k}$. The DE of the power of the term concerning the estimation error becomes

$$\frac{1}{N^2}\left|\mathbf{g}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR}}\tilde{\mathbf{e}}_{\mathrm{SR},k}\right|^2 = \frac{1}{N^2}\left|\frac{\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\tilde{\mathbf{e}}_{\mathrm{SR},k}}{1+\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}}\right|^2 \quad (44)$$

$$\asymp \frac{1}{N}\frac{\delta'_{\mathrm{SR},k}}{\left(1+\delta_{\mathrm{SR},k}\right)^2}, \quad (45)$$

where $\delta'_{\mathrm{SR},k} = \frac{1}{N}\operatorname{tr}(\mathbf{D}_{\mathrm{SR},k} - \hat{\mathbf{D}}_{\mathrm{SR},k})\mathbf{T}'_{\mathbf{D}_{\mathrm{SR},k}}$ and $\mathbf{K} = \mathbf{D}_{\mathrm{SR},k} - \hat{\mathbf{D}}_{\mathrm{SR},k}$. The DE of the term, including the expression of the received AWGN noise, is obtained as

$$\frac{1}{N^2}\left|\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR}}\right|^2 = \frac{1}{N^2}\left|\frac{\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}}{1+\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\hat{b}g_{\mathrm{SR},k}}\right|^2 \quad (46)$$

$$= \frac{1}{N^2}\frac{\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}^2\hat{\mathbf{g}}_{\mathrm{SR},k}}{\left(1+\frac{1}{N}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}\right)^2}$$

$$\asymp \frac{1}{N}\frac{\delta''_{\mathrm{SR},k}}{\left(1+\delta_{\mathrm{SR},k}\right)^2}, \quad (47)$$

where $\delta''_{\mathrm{SR},k} = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}''_{\mathrm{SR},k}$ and $\mathbf{K} = \mathbf{I}_N$. Note that we have applied [42, Lem. B.26] and [41, Thm. 1]. The derivation of the DE of the power of the multi-user interference follows. We have

$$\frac{1}{N^2}\left|\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR}}\mathbf{g}_{\mathrm{SR},j}\right|^2 = \frac{1}{N^2}\left|\frac{\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\mathbf{g}_{\mathrm{SR},j}}{1+\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}}\right|^2$$

$$\asymp \frac{1}{N^2}\frac{\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\mathbf{D}_{\mathrm{SR},j}\hat{\Sigma}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}}{\left(1+\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}\right)^2}. \quad (48)$$

Since $\hat{\Sigma}_{\mathrm{SR},k}$ is not independent of $\mathbf{g}_{\mathrm{SR},j}$, the use of [16, Lemma 2] gives

$$\hat{\Sigma}_{\mathrm{SR},k} = \hat{\Sigma}_{\mathrm{SR},jk} - \frac{\hat{\Sigma}_{\mathrm{SR},jk}\hat{\mathbf{g}}_{\mathrm{SR},j}\hat{\mathbf{g}}_{\mathrm{SR},j}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},jk}}{1+\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},jk}\hat{\mathbf{g}}_{\mathrm{SR},k}}, \quad (49)$$

which introduces a new matrix $\hat{\Sigma}_{\mathrm{SR},jk}$ to (48) defined as $\hat{\Sigma}_{\mathrm{SR},jk} = \left(\hat{\mathbf{G}}_{\mathrm{SR}}\hat{\mathbf{G}}_{\mathrm{SR}}^{\mathsf{H}} - \hat{\mathbf{g}}_{\mathrm{SR},j}\hat{\mathbf{g}}_{\mathrm{SR},j}^{\mathsf{H}} - \hat{\mathbf{g}}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}} + \mathbf{Z}_{\mathrm{SR},k} + N\alpha_{\mathrm{SR}}\mathbf{I}_N\right)^{-1}$. By substituting (49) into (48) and applying [42, Lem. B.26] and [41, Thm. 1], we obtain $\frac{1}{N^2}\left|\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\mathbf{g}_{\mathrm{SR},j}\right|^2 \asymp \frac{\mu_{\mathrm{SR},jk}}{N\left(1+\hat{\delta}_{\mathrm{SR},k}\right)^2}$, where $\mu_{\mathrm{SR},jk}$ is given by (50). The derivation of the DE of each term of (50) follows. In particular, we have

$$\frac{1}{N^2}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},jk}\hat{\mathbf{D}}_{\mathrm{SR},j}\hat{\Sigma}_{\mathrm{SR},jk}\hat{\mathbf{g}}_{\mathrm{SR},k}$$

$$\asymp \frac{1}{N^2}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\hat{\Sigma}_{\mathrm{SR},jk}\hat{\mathbf{D}}_{\mathrm{SR},j}\hat{\Sigma}_{\mathrm{SR},jk}$$

$$\asymp \frac{1}{N^2}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},j}\mathbf{T}'_{\mathrm{SR},jk} = \frac{\delta'_{\mathrm{SR},jk}}{N}, \quad (51)$$

where we have used [41, Thm. 1] and [16, Thm. 2] as well as we have defined $\delta'_{\mathrm{SR},jk} = \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},j}\mathbf{T}'_{\mathrm{SR},jk}$. Moreover, we have used $\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},jk}\hat{\mathbf{g}}_{\mathrm{SR},k} \asymp \frac{1}{N}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},jk} = \delta_{\mathrm{SR},jk}$ by means of [41, Thm. 1] and [16, Thm. 2]. As shown above, application of Lemma [42, Lem. B.26] as well as [41, Thm. 1] and [16, Thm. 2] to the first and second term of (50) gives

$$\frac{1}{N^2}\left|\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},jk}\hat{\mathbf{g}}_{\mathrm{SR},k}\right|^2 \asymp \frac{1}{N^2}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\hat{\Sigma}_{\mathrm{SR},jk}\hat{\mathbf{D}}_{\mathrm{SR},k}\hat{\Sigma}_{\mathrm{SR},jk}$$

$$\asymp \frac{1}{N^2}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{SR},k}\mathbf{T}''_{\mathrm{SR},jk} = \frac{\delta''_{\mathrm{SR},jk}}{N}. \quad (52)$$

Regarding the term including the SI, we have

$$\frac{1}{N^2}\mathbf{w}_k^{\mathsf{H}}\mathbf{G}_{\mathrm{RR}}\mathbb{E}\left[\mathbf{s}[n]\mathbf{s}^{\mathsf{H}}[n]\right]\mathbf{G}_{\mathrm{RR}}^{\mathsf{H}}\mathbf{w}_k = \frac{1}{N^2}\mathbf{w}_k^{\mathsf{H}}\mathbf{G}_{\mathrm{RR}}\mathbf{G}_{\mathrm{RR}}^{\mathsf{H}}\mathbf{w}_k$$

$$= \frac{1}{N^2}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR}}\mathbf{G}_{\mathrm{RR}}\mathbf{G}_{\mathrm{RR}}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR}}\hat{\mathbf{g}}_{\mathrm{SR},k} \quad (53)$$

$$= \frac{1}{N}\frac{\frac{1}{N}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\mathbf{G}_{\mathrm{RR}}\mathbf{G}_{\mathrm{RR}}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}}{\left(1+\frac{1}{N}\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}\hat{\mathbf{g}}_{\mathrm{SR},k}\right)^2} \quad (54)$$

$$\asymp \frac{1}{N}\frac{\frac{1}{N}\operatorname{tr}\mathbf{D}_{\mathrm{SR},k}\hat{\Sigma}_{\mathrm{SR},k}\mathbf{G}_{\mathrm{RR}}\mathbf{G}_{\mathrm{RR}}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{SR},k}}{\left(1+\frac{1}{N}\operatorname{tr}\mathbf{D}_{\mathrm{SR},k}\hat{\Sigma}_{\mathrm{SR},k}\right)^2} \quad (55)$$

$$\asymp \frac{1}{N}\frac{\frac{1}{N}\operatorname{tr}\mathbf{D}_{\mathrm{SR},k}\mathbf{T}''_{\mathrm{SR},k}\frac{1}{N}\operatorname{tr}\mathbf{G}_{\mathrm{RR}}\mathbf{G}_{\mathrm{RR}}^{\mathsf{H}}}{\left(1+\frac{1}{N}\operatorname{tr}\mathbf{D}_{\mathrm{SR},k}\mathbf{T}_{\mathrm{SR},k}\right)^2} \quad (56)$$

$$= \frac{\delta''_{\mathrm{SR},jk}\frac{1}{N}\operatorname{tr}\mathbf{T}_{\mathrm{RR}}}{\left(1+\delta_{\mathrm{SR},k}\right)^2}, \quad (57)$$

where in (54), we have used [45, Eq. 2.2] twice. Next, in (55), we have applied [42, Lem. B.26], while in (56), [45, Eq. 2.2] is used. Note that $\frac{1}{N}\operatorname{tr}\mathbf{T}_{\mathrm{RR}} = \frac{1}{N}\operatorname{tr}\mathbf{G}_{\mathrm{RR}}\mathbf{G}_{\mathrm{RR}}^{\mathsf{H}}$ due to Theorem with the covariance of the ith column of $\mathbf{G}_{\mathrm{RR}}$ equal to $\sigma_{\mathrm{SI}}^2\mathbf{I}_N$. If we make the necessary substitutions, we obtain the corresponding deterministic equivalent $\bar{\gamma}_{\mathrm{SR},k}$ and this concludes the proof.

APPENDIX B
PROOF OF THEOREM 2

First, we derive the DE of the normalization parameter λ. We start with a simple algebraic manipulation to λ. We obtain $\lambda = \frac{K}{\operatorname{tr}\hat{\mathbf{G}}_{\mathrm{RD}}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{RD}}^2\hat{\mathbf{G}}_{\mathrm{RD}}} = \frac{K}{\Psi}$. Next, we have

$$\Psi = \sum_{k=1}^{K}\hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}}\hat{\Sigma}_{\mathrm{RD}}^2\hat{\mathbf{g}}_{\mathrm{RD},k} \quad (58)$$

$$\asymp \frac{1}{M}\sum_{k=1}^{K}\frac{\frac{1}{M}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{RD},k}\hat{\Sigma}_{\mathrm{RD},k}^2}{\left(1+\frac{1}{M}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{RD},k}\hat{\Sigma}_{\mathrm{RD},k}\right)^2} \quad (59)$$

$$\asymp \frac{1}{M}\sum_{k=1}^{K}\frac{\delta'_{\mathrm{RD},k}}{\left(1+\delta_{\mathrm{RD},k}\right)^2}, = \frac{K}{\bar{\lambda}}, \quad (60)$$

where we have defined $\hat{\Sigma}_{\mathrm{RD},k} \triangleq \left(\hat{\mathbf{W}}_{\mathrm{RD}} - \hat{\mathbf{g}}_{\mathrm{RD},k}\hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} + \mathbf{Z}_{\mathrm{RD}} + \alpha_{\mathrm{RD}}M\mathbf{I}_M\right)^{-1}$ with $\hat{\mathbf{W}}_{\mathrm{RD}} = \hat{\mathbf{G}}_{\mathrm{RD}}\hat{\mathbf{G}}_{\mathrm{RD}}^{\mathsf{H}}$. Also, we have applied [41, Thm. 1] and [16, Thm. 2] for $\mathbf{L} = \hat{\mathbf{D}}_{\mathrm{RD},k}$ and $\mathbf{K} = \mathbf{I}_M$, and we have denoted $\delta_{\mathrm{RD},k} = \frac{1}{M}\operatorname{tr}\hat{\mathbf{D}}_{\mathrm{RD}k}\mathbf{T}_{\mathrm{RD},k}$ and

$$\mu_{\mathrm{SR},k} = \hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{D}}_{\mathrm{SR},j} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{g}}_{\mathrm{SR},k} + \frac{\left|\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{g}}_{\mathrm{SR},k}\right|^2 \hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{D}}_{\mathrm{SR},j} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{g}}_{\mathrm{SR},k}}{\left(1 + \hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{g}}_{\mathrm{SR},k}\right)^2}$$
$$- 2\mathrm{Re}\left\{\frac{\hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \mathbf{g}_{\mathrm{SR},k} \mathbf{g}_{\mathrm{SR},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{D}}_{\mathrm{SR},j} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{g}}_{\mathrm{SR},k}}{1 + \hat{\mathbf{g}}_{\mathrm{SR},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{SR},jk} \hat{\mathbf{g}}_{\mathrm{SR},k}}\right\}. \tag{50}$$

$\delta'_{\mathrm{RD},k} = \frac{1}{M} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \mathbf{T}'_{\mathrm{RD},k}$. Hence, $\lambda \asymp \bar{\lambda}$. Regarding the rest part of the desired signal power, we substitute the RZF precoder, and after dividing by $\frac{1}{M}$ we have

$$\frac{1}{M} \mathbf{g}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD}} \hat{\mathbf{g}}_{\mathrm{RD},k} = \frac{1}{M} \hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD}} \hat{\mathbf{g}}_{\mathrm{RD},k} \tag{61}$$

$$= \frac{\frac{1}{M} \mathrm{tr}\, \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},k} \hat{\mathbf{D}}_{\mathrm{RD},k}}{1 + \frac{1}{M} \mathrm{tr}\, \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},k} \hat{\mathbf{D}}_{\mathrm{RD},k}} \tag{62}$$

$$= \frac{\frac{1}{M} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \mathbf{T}_{\mathrm{RD},k}}{1 + \frac{1}{M} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \mathbf{T}_{\mathrm{RD},k}}, \tag{63}$$

where we have applied Lemmas [45, Eq. 2.2], [42, Lem. B.26], and [46, p. 207] in (61) and (62), respectively. Moreover, we have exploited [41, Thm. 1] in (63) for $\mathbf{L} = \hat{\mathbf{D}}_{\mathrm{RD},k}$. Writing in a concise form the last equation, we obtain $\frac{1}{M} \mathbf{g}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD}} \hat{\mathbf{g}}_{\mathrm{RD},k} = \frac{\delta_{\mathrm{RD},k}}{1+\delta_{\mathrm{RD},k}}$. We continue the proof with the derivation of the DE of the term of the interference part of $\frac{\rho_j}{K} \sum_{j \neq k}^{K} |\mathbf{g}_{\mathrm{RD},k}^{\mathsf{H}} \mathbf{f}_{\mathrm{RD},j}|^2$. Making use of [45, Eq. 2.2], we obtain by means of [42, Lem. B.26] and [45, Eq. 2.2]

$$\frac{1}{M^2} |\mathbf{g}_{\mathrm{RD},k}^{\mathsf{H}} \mathbf{f}_{\mathrm{RD},j}|^2 = \frac{1}{M^2} |\hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD}} \hat{\mathbf{g}}_{\mathrm{RD},j}|^2 \tag{64}$$

$$= \frac{1}{M^2} \frac{\hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},j} \hat{\mathbf{g}}_{\mathrm{RD},j} \hat{\mathbf{g}}_{\mathrm{RD},j}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},j} \hat{\mathbf{g}}_{\mathrm{RD},k}}{\left(1 + \hat{\mathbf{g}}_{\mathrm{RD},j}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},j} \hat{\mathbf{g}}_{\mathrm{RD},j}\right)^2} \tag{65}$$

$$= \frac{1}{M^2} \frac{\hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},j} \hat{\mathbf{D}}_{\mathrm{RD},j} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},j} \hat{\mathbf{g}}_{\mathrm{RD},k}}{\left(1 + \hat{\mathbf{g}}_{\mathrm{RD},j}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},j} \hat{\mathbf{g}}_{\mathrm{RD},j}\right)^2}, \tag{66}$$

where in (66), we have taken into consideration that $\hat{\mathbf{g}}_{\mathrm{RD},k}$ and $\hat{\mathbf{g}}_{\mathrm{RD},j}$ are mutually independent. Given that $\hat{\boldsymbol{\Sigma}}_{\mathrm{RD},j}$ is not independent of $\hat{\mathbf{g}}_{\mathrm{RD},k}$, we employ [16, Lemma 2], which yields

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{RD},j} = \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} - \frac{\hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \hat{\mathbf{g}}_{\mathrm{RD},k} \hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk}}{1 + \hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \hat{\mathbf{g}}_{\mathrm{RD},k}}, \tag{67}$$

where the new matrix $\hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk}$ is defined as $\hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} = (\hat{\mathbf{W}}_{\mathrm{RD}} - \hat{\mathbf{g}}_{\mathrm{RD},k} \hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} - \hat{\mathbf{g}}_{\mathrm{RD},j} \hat{\mathbf{g}}_{\mathrm{RD},j}^{\mathsf{H}} + \mathbf{Z}_{\mathrm{RD}} + \alpha_{\mathrm{RD}} \mathbf{I}_M)^{-1}$. After substituting (67) into (66), we have $\frac{1}{M^2} \left|\hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD}} \hat{\mathbf{g}}_{\mathrm{RD},j}\right|^2 = \frac{Q_{\mathrm{RD},jk}}{M(1+\delta_{\mathrm{RD},k})^2}$, where $Q_{\mathrm{RD},jk}$ is given by (70). We proceed with the derivation of the DE of each term in (70). Specifically, we have

$$\frac{1}{M^2} \hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \hat{\mathbf{D}}_{\mathrm{RD},j} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \hat{\mathbf{g}}_{\mathrm{RD},k}$$
$$\asymp \frac{1}{M^2} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \hat{\mathbf{D}}_{\mathrm{RD},j} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \tag{68}$$
$$\asymp \frac{1}{M^2} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},j} \mathbf{T}''_{\mathrm{RD},jk} = \frac{\delta''_{\mathrm{RD},jk}}{M}, \tag{69}$$

where we have applied [42, Lem. B.26] and [16, Thm. 2] for $\mathbf{L} = \hat{\mathbf{D}}_{\mathrm{RD},k}$ and $\mathbf{K} = \hat{\mathbf{D}}_{\mathrm{RD},j}$. Similarly, we have

$$\frac{1}{M^2} \left|\hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \hat{\mathbf{g}}_{\mathrm{RD},k}\right|^2$$
$$\asymp \frac{1}{M^2} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \hat{\mathbf{D}}_{\mathrm{RD},k} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk}$$
$$= \frac{1}{M^2} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \hat{\mathbf{D}}_{\mathrm{RD},k} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \tag{71}$$
$$\asymp \frac{1}{M^2} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \mathbf{T}'''_{\mathrm{RD},jk} = \frac{\delta'''_{\mathrm{RD},jk}}{M}, \tag{72}$$

where $\mathbf{L} = \hat{\mathbf{D}}_{\mathrm{RD},k}$ and $\mathbf{K} = \hat{\mathbf{D}}_{\mathrm{RD},k}$, and $\delta'''_{\mathrm{RD},jk} = \frac{1}{M} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},jk} \mathbf{T}'''_{\mathrm{RD},jk}$. The next term is written as

$$\frac{1}{M^2} \hat{\mathbf{g}}_{\mathrm{RD},k}^{\mathsf{H}} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \mathbf{g}_{\mathrm{RD},k} \asymp \frac{1}{M} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \hat{\boldsymbol{\Sigma}}_{\mathrm{RD},jk} \tag{73}$$
$$\asymp \frac{1}{M} \mathrm{tr}\, \hat{\mathbf{D}}_{\mathrm{RD},k} \mathbf{T}_{\mathrm{RD},jk} = \delta_{\mathrm{RD},jk}, \tag{74}$$

where in (73), we have applied both Lemmas [42, Lem. B.26] and [46, p. 207], while in the next equation, we have applied [41, Thm. 1]. Hence, (70) becomes

$$Q_{\mathrm{RD},jk} \asymp \frac{\delta'_{\mathrm{RD},j}}{M} + \frac{\left|\delta''_{\mathrm{RD},k}\right|^2 \delta'_{\mathrm{RD},k}}{M\left(1+\delta_{\mathrm{RD},k}\right)^2}$$
$$- 2\mathrm{Re}\left\{\frac{\delta''_{\mathrm{RD},jk} \delta'_{\mathrm{RD},k}}{M\left(1+\delta_{\mathrm{RD},k}\right)}\right\}. \tag{75}$$

Making the necessary substitutions, the proof is concluded.

Appendix C
Proof of Proposition 2

The private messages of the RS transmission in the second link achieve almost the same sum rate as the conventional BC with full power, when $\mathrm{R}_{\mathrm{RD},k}^{\mathrm{p}} = \mathrm{R}_{\mathrm{RD},k}^{\mathrm{NoRS}}$ holds. Actually, this happens, when

$$\frac{\bar{\lambda} \frac{\rho t}{K} \left(\delta_{\mathrm{RD},k}\right)^2}{\bar{\lambda} \frac{\rho t}{K} \sum_{j \neq k}^{K} \frac{Q_{\mathrm{RD},jk}}{M} + \left(1+\delta_{\mathrm{RD},k}\right)^2} > 1. \tag{76}$$

Then, thinking that the number of users K is generally much larger than 1, we set

$$\frac{\bar{\lambda} \frac{\rho t}{K} \left(\delta_{\mathrm{RD},k}\right)^2}{\bar{\lambda} \frac{\rho t}{K} \sum_{j \neq k}^{K} \frac{Q_{\mathrm{RD},jk}}{M} + \left(1+\delta_{\mathrm{RD},k}\right)^2} = K. \tag{77}$$

$$Q_{\text{RD},jk} = \hat{\mathbf{g}}_{\text{RD},k}^{\text{H}} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \hat{\mathbf{D}}_{\text{RD},j} \hat{\boldsymbol{\Sigma}}_{jk} \hat{\mathbf{g}}_{\text{RD},k} + \frac{\left| \hat{\mathbf{g}}_{\text{RD},k}^{\text{H}} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \hat{\mathbf{g}}_{\text{RD},k} \right|^2 \hat{\mathbf{g}}_{\text{RD},k}^{\text{H}} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \hat{\mathbf{D}}_{\text{RD},j} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \hat{\mathbf{g}}_{\text{RD},k}}{\left(1 + \hat{\mathbf{g}}_{\text{RD},k}^{\text{H}} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \hat{\mathbf{g}}_{\text{RD},k} \right)^2}$$
$$- 2\text{Re} \left\{ \frac{\hat{\mathbf{g}}_{\text{RD},k}^{\text{H}} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \mathbf{g}_{\text{RD},k} \mathbf{g}_{\text{RD},k}^{\text{H}} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \hat{\mathbf{D}}_{\text{RD},j} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \hat{\mathbf{g}}_{\text{RD},k}}{1 + \hat{\mathbf{g}}_{\text{RD},k}^{\text{H}} \hat{\boldsymbol{\Sigma}}_{\text{RD},jk} \hat{\mathbf{g}}_{\text{RD},k}} \right\}. \tag{70}$$

Another reason for the dependence of this setting on K follows. Specifically, let us denote

$$\bar{Y} = \frac{\bar{\lambda} \frac{1}{K} \left(\delta_{\text{RD},k} \right)^2}{\bar{\lambda} \frac{1}{K} \sum_{j \neq k}^{K} \frac{Q_{\text{RD},jk}}{M} + \left(1 + \delta_{\text{RD},k} \right)^2}. \tag{78}$$

Since the common message should be decoded by all destination users, less power is allocated at the common message as K increases, and equivalently, the rate of the common message reduces. Thus, $\bar{Y}$ should depend on K.

The power splitting ratio t is chosen as $t = K / \left(\rho \bar{Y} \right)$. If the choice of t is the smallest value between $t = K / \left(\rho \bar{Y} \right)$ and 1, the inequality $\text{R}_{\text{RD},k}^{\text{p}} \leq \text{R}_{\text{RD},k}^{\text{NoRS}}$ becomes equality. Note that at low SNR ($\rho \to 0$), t becomes 1, which means that transmission of the common message is not beneficial, while increasing the SNR, the common message boosts the sum-rate. Similar to [19], the rate loss in the second link between the private messages of the NoRS and RS is proved to be upper bounded as

$$\sum_{j=1}^{K} \left(\text{R}_{\text{RD},j}^{\text{NoRS}} - \text{R}_{\text{RD},j}^{\text{p}} \right) \leq \log_2 e. \tag{79}$$

Substituted (79) into (38), we obtain the desired result.

Acknowledgement

References

[1] A. Papazafeiropoulos and T. Ratnarajah, "Impact of rate-splitting robustness in multi-pair massive mimo Relay systems," in *IEEE Global Commun. Conf. (GLOBECOM) 2017*, Singapore, Dec. 2017.

[2] "FP7 integrating project METIS (lCT 317669)." [Online]. Available: https://www.metis2020.com

[3] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] E. Larsson *et al.*, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] D. Bliss, P. Parker, and A. Margetts, "Simultaneous transmission and reception for improved wireless network performance," in *IEEE/SP 14th Workshop on Statistical Signal Processing, 2007, SSP'07.* IEEE, 2007, pp. 478–482.

[6] D. Bliss, T. Hancock, and P. Schniter, "Hardware and environmental phenomenological limits on full-duplex MIMO relay performance," in *Proc. Annual Asilomar Conf. Signals, Syst. and Comput (ASILOMAR)*, 2012.

[7] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, 2011.

[8] G. Zheng, I. Krikidis, and B. Ottersten, "Full-duplex cooperative cognitive radio with transmit imperfections," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2498–2511, 2013.

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

[10] T. Riihonen, S. Werner, and R. Wichman, "Transmit power optimization for multiantenna decode-and-forward relays with loopback self-interference from full-duplex operation," in *Forty Fifth Asilomar Conf. on Signals, Syst. and Comput. (ASILOMAR)*. IEEE, 2011, pp. 1408–1412.

[11] Y. Sung *et al.*, "Loop-interference suppression strategies using antenna selection in full-duplex MIMO relays," in *International Symp. on Intelligent Signal Process. and Commun. Systems (ISPACS), 2011.* IEEE, 2011, pp. 1–4.

[12] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas in Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[13] L. Lu *et al.*, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[14] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[15] A. K. Papazafeiropoulos and T. Ratnarajah, "Deterministic equivalent performance analysis of time-varying massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5795–5809, 2015.

[16] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[17] S. Yang *et al.*, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, 2013.

[18] C. Hao, Y. Wu, and B. Clerckx, "Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3232–3246, 2015.

[19] M. Dai *et al.*, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. PP, no. 99, pp. 1–1, 2016.

[20] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Mitigation of phase noise in massive MIMO systems: A rate-splitting approach," in *Proc. of IEEE Int. Conf. Commun.*, 2017. [Online]. Available: arXiv:1702.01169

[21] ——, "Rate-splitting to mitigate residual transceiver hardware impairments in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8196–8211, Sept. 2017.

[22] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, 2016.

[23] S.-Y. Chung *et al.*, "Capacity enhancement schemes for forward and reverse links of distributed cellular base stations," Apr. 3 2007, US Patent 7,200,391.

[24] C. Hoymann *et al.*, "Relaying operation in 3GPP LTE: Challenges and solutions," *IEEE Commun. Mag.*, vol. 50, no. 2, 2012.

[25] J. Chen *et al.*, "Spectral-energy efficiency tradeoff in relay-aided massive MIMO cellular networks with pilot contamination," *IEEE Access*, vol. 4, pp. 5234–5242, 2016.

[26] H. Cui, L. Song, and B. Jiao, "Multi-pair two-way amplify-and-forward relaying with very large number of relay antennas," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2636–2645, 2014.

[27] X. Xiong *et al.*, "Channel estimation for full-duplex relay systems with large-scale antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6925–6938, Oct. 2016.

[28] X. Sun *et al.*, "Multi-pair two-way massive MIMO AF full-duplex relaying with imperfect CSI over Ricean fading channels," *IEEE Access*, vol. 4, pp. 4933–4945, 2016.

[29] X. Xia *et al.*, "Hardware impairments aware transceiver for full-duplex massive MIMO relaying," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6565–6580, 2015.

[30] L. Chen *et al.*, "Green full-duplex self-backhaul and energy harvesting

small cell networks with massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3709–3724, Dec. 2016.

[31] H. Q. Ngo *et al.*, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, 2014.

[32] X. Xia *et al.*, "Multi-pair full-duplex amplify-and-forward relaying with very large antenna arrays," in *IEEE Wireless Commun. and Networking Conf. (WCNC)*, 2015, pp. 304–309.

[33] H. A. Suraweera *et al.*, "Multi-pair amplify-and-forward relaying with very large antenna arrays," in *IEEE International Conf. on Commun. (ICC)*, 2013, pp. 4635–4640.

[34] Y. Yang *et al.*, "Relay technologies for WiMAX and LTE-advanced mobile systems," *IEEE Commun. Mag.*, vol. 47, no. 10, pp. 100–105, 2009.

[35] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River: Prentice Hall PTR, 1993.

[36] B. Clerckx *et al.*, "Rate splitting for MIMO wireless networks: A promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, 2016.

[37] C. Lim *et al.*, "Recent trend of multiuser MIMO in LTE-advanced," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 127–135, 2013.

[38] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[39] P. Billingsley, *Probability and measure*, 3rd ed. John Wiley & Sons, Inc., 2008.

[40] A. W. van der Vaart, *Asymptotic statistics (Cambridge series in statistical and probabilistic mathematics)*. New York: Cambridge University Press, 2000.

[41] S. Wagner *et al.*, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[42] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*. Springer, 2010, vol. 20.

[43] Z. Xiang, M. Tao, and X. Wang, "Massive MIMO multicasting in noncooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1180–1193, 2014.

[44] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. no.8, pp. 4353–4368, Aug. 2015.

[45] J. W. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[46] T. Tao, *Topics in random matrix theory*. American Mathematical Soc., 2012, vol. 132.