SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

February 2017, Vol. 60 022309:1–022309:15
doi: 10.1007/s11432-016-5601-1

Energy-efficiency-aware relay selection in distributed full duplex relay network with massive MIMO

Yongzhi LI[1], Cheng TAO[1,2*], Liu LIU[1,2] & Lingwen ZHANG[1]

[1]*Institute of Broadband Wireless Mobile Communications, Beijing Jiaotong University, Beijing* 100044*, China;*
[2]*National Mobile Communications Research Laboratory, Southeast University, Nanjing* 211189*, China*

Received December 11, 2015; accepted March 2, 2016; published online September 13, 2016

Abstract This paper considers a distributed full duplex relay network where multiple sources simultaneously transmit their signals to multiple destinations via the cooperation of a set of relay stations (RSs). Each RS is assumed to be equipped with large antenna arrays while all sources and destinations only have a single antenna. We assume the channels are Rician fading and the RSs use linear processing to process the signals with imperfect channel state information (CSI). We derive the closed-form expressions of the end-to-end achievable rates for the maximum-ratio combining/matched-filter (MRC/MF) and the zero-forcing (ZF) processing. These results are then used to pursue a detailed analysis of the power saving of the relay network. Then we study the energy-efficiency-aware relay selection strategy since the energy efficiency affects network lifetime in future wireless network. We propose a sub-optimal strategy with low complexity that only requires the statistical CSI. Simulations show that the energy efficiency of the system can be improved with optimal power allocation and our proposed strategy performs very close to the exhaustive search algorithm which is optimal.

Keywords relay network, full duplex, massive MIMO, relay selection, energy efficiency.

Citation Li Y Z, Tao C, Liu L, et al. Energy-efficiency-aware relay selection in distributed full duplex relay network with massive MIMO. Sci China Inf Sci, 2017, 60(2): 022309, doi: 10.1007/s11432-016-5601-1

1 Introduction

Compared with traditional MIMO technology, massive MIMO technique, which is an emerging technology, has attracted much attention currently [1, 2]. The most important benefit in massive MIMO system is that the interference among the users and the effects of noise can be reduced while the spectral efficiency can be significantly improved using simple signal processing techniques [3] such as the maximum-ratio combining/matched-filter (MRC/MF) and the zero-forcing (ZF) processing.

On a parallel avenue, relay technology, as a kind of cooperative communication technique, has been extensively explored in the literatures due to their potential of offering a number of significant performance enhancement in terms of outage behavior, achievable rate region and error probability [4, 5]. As a result, some communication standards such as LTE-Advanced are expected to support relay based communication. While initial work on the problem focused on the half duplex (HD) relaying where relays

* Corresponding author (email: chtao@bjtu.edu.cn)

retransmit the source data in orthogonal and dedicated channels, attention has shifted recently to the full duplex (FD) relaying where the relay receives and transmits simultaneously on the same channel. Hence, FD relaying is able to overcome the bandwidth loss and improve the spectral efficiency. However, the FD relay suffers from several interferences which mainly limit its performance, i.e., the loop interference due to the signal leakage from the same relay's output to the input, and the cross talk interference caused by the signal interference among different relays' output and input [6, 7]. Therefore, the loop interference mitigation is a highly active area in FD research. Recent progresses have been made on both theory and hardware implementation to make FD wireless communication a viable practical solution.

Refs. [7–10] provide details on recent practical FD implementations which demonstrate the feasibility of the FD operation mode with the existing technology. More precisely, three types of loop interference mitigation schemes were proposed in [7]: (i) physical isolation; (ii) time-domain cancelation and (iii) spatial suppression. The results showed that the proposed schemes in [7] offer significant mitigation such that the residual interference can be regarded as noise. More recently, exploiting a large number of degrees of freedom, Ref. [11] combined massive MIMO technique and FD relaying system and proposed two techniques to significantly reduce the loop interference. Importantly, the two loop interference mitigation techniques can be achieved by utilizing simple signal processing techniques. Therefore, massive MIMO combined with FD relaying is a worthy candidate for the future wireless networks.

In addition, relay selection has been widely investigated in HD relay network as an attractive relaying solution offering high performance, efficient use of system resources and simple hardware implementation [12, 13]. Therefore, not surprisingly, applying the relay selection concept to FD relay network provides an efficient approach to combine space diversity benefits and FD spectral efficiency. The authors in [14] focused on the relay selection in amplify-and-forward full-duplex relay network and investigated serval suboptimal relay selection schemes requiring partial CSI. The relay selection schemes have been analyzed in terms of outage probability. The authors in [15] studied the employment of a relay selection policy for a FD multi-relay network of decode-and-forward (DF) protocol and presented a closed-form expression for the average capacity as well as the bit error rate (BER). The authors in [16] considered a two-way FD relay system using amplify-and-forward protocol and proposed a relay slection scheme in maximizing the effective signal-to-interference-noise ratio (SINR).

However, most of the previous papers only considered the full duplex relays with only a few antennas, in which the loop interference and cross-talk interference may not be suppressed easily and, hence, limit the performance of the relay network. Different with previous papers, we consider a distributed decode-and-forward massive MIMO full-duplex relaying architecture since a large number of degrees of freedom in massive MIMO can be used to significantly suppress the loop interference in the spatial domain. However, the increasing physical size of massive antenna-array is a fundamental problem for practical deployment. Fortunately, the millimeter wave finds a way out since its small wavelengths can reduce the antenna spacing and makes the massive MIMO technique more feasible in practice [17, 18]. Hence, based on this technique, we model the channel as Rician fading due to the fact that, with the highly directional and quasi-optical nature of propagation at millimeter wave, line-of-sight (LOS) propagation is dominating [19–21]. We first derive the closed-form expressions of the end-to-end achievable rates for MRC/MF and ZF processing over Rician fading channels, respectively. Based on the expressions, we evaluate the system performance of the full duplex relay network and show that, due to the LOS component, the transmit power of each source and of each selected relay station can be scaled down proportionally to $1/N_r$ and $1/N_t$ as $N_r, N_t \to \infty$, respectively, which is regardless of the pilot power, to maintain a desirable quality of service. Moreover, we propose an energy-efficiency-aware relay selection strategy with low complexity to maximize the energy efficiency of the relay network. The advantages of our proposed strategy are: (i) the complexity is quite low which increases linearly in the network size; (ii) only statistical CSI is needed in the relay network system. Therefore, the overhead of the information exchanges can be very small. The simulations show that our proposed strategy performs very close to the optimal one and the energy efficiency of the system can be significantly improved with optimal power allocation.

The rest of this paper is organized as follows. Section 2 presents the system architecture. In Section 3,

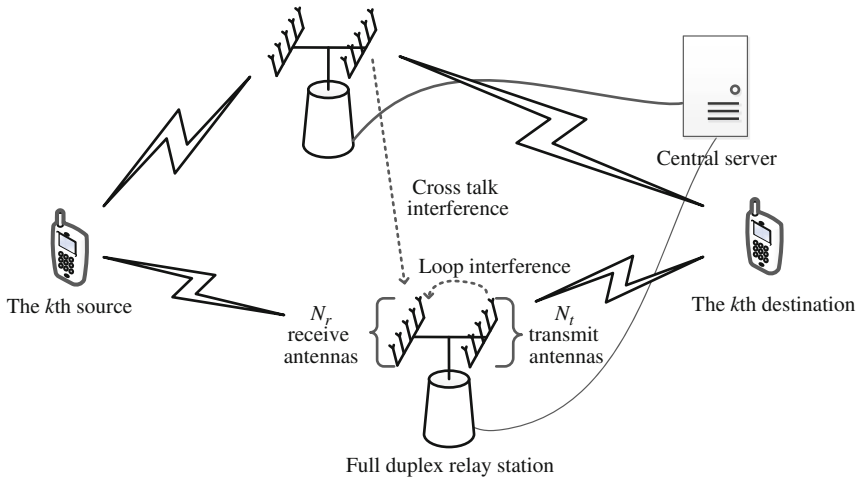

Figure 1 System architecture of a distributed full duplex relay network.

we derive the closed-form expressions of the end-to-end achievable rate for MRC/MF and ZF processing. Utilizing the closed-form expressions, we then propose the energy-efficiency-aware relay selection strategy in Section 4. Simulation results are presented in Section 5 and we conclude this paper in Section 6.

2 System model

2.1 System architecture and channel models

Figure 1 illustrates the considered distributed decode-and-forward relay network, which consists of M distributed full duplex relay stations, K mobile communication pairs, and one central server. Each relay station has N_r receive antennas and N_t transmit antennas, while each source and each destination is equipped with a single antenna. All distributed RSs are connected to the central server through the high-speed optical connections and the central server will dynamically select a subset $\mathcal{L}$ of RS's where $\mathcal{L} = \{\pi_1, \ldots, \pi_L\}$ and $\pi_l \in \{1, 2, \ldots, M\}$ for $l = 1, \ldots, L$. Then the kth source communicates with kth destination with the help of the cooperation of these selected RSs. We denote by $|\mathcal{L}| = L$ the cardinality of the selected RS subset (the number of selected RSs) for the distributed full duplex relay transmission. Following the assumption in [11] and [22], the channels from the L selected RSs to the destinations are reciprocal and the direct links among sources and destinations do not exist due to the strong attenuation. Once L RSs are selected, all K sources transmit their data symbols to the L selected RSs simultaneously, while the L RSs broadcast the precoded symbols to all K destinations and the remaining $M - L$ RSs keep silence (i.e., the transmit power equals to zero). Hereafter, we user the term *transmission mode* L to denote the case where the cardinality of the selected RS subset is equal to L. For a given transmission mode L, we only consider the transmission between the K communication pairs and the L selected RSs. For convenience, we hereafter rewrite the index of the lth selected RS π_l as l for $l = 1, \ldots, L$.

Denote $\boldsymbol{G}_{S,l} = [\boldsymbol{g}_{S,l1}, \ldots, \boldsymbol{g}_{S,lK}] \in \mathbb{C}^{N_r \times K}$ and $\boldsymbol{G}_{D,l} = [\boldsymbol{g}_{D,l1}, \ldots, \boldsymbol{g}_{D,lK}] \in \mathbb{C}^{N_t \times K}$ for $l = 1, \ldots, L$ as the channel matrices from the K sources to the receive antennas of the lth selected RS and from the K destinations to the transmit antennas of the lth selected RS, respectively. We assume the channel embraces independent fast fading, geometric attenuation and log-normal shadowing fading. More precisely, $\boldsymbol{G}_{S,l}$ and $\boldsymbol{G}_{D,l}$ can be expressed as $\boldsymbol{G}_{S,l} = \boldsymbol{H}_{S,l}\boldsymbol{D}_{S,l}^{1/2}$ and $\boldsymbol{G}_{D,l} = \boldsymbol{H}_{D,l}\boldsymbol{D}_{D,l}^{1/2}$, where $\boldsymbol{H}_{S,l}$ and $\boldsymbol{H}_{D,l}$ are the channel matrices modeling fast fading, while $\boldsymbol{D}_{S,l}$ and $\boldsymbol{D}_{D,l}$ are the large-scale fading diagonal matrices whose kth diagonal entries are denoted by $\beta_{S,lk}$ and $\beta_{D,lk}$, respectively. According to [19, 21], we model the fast fading channels as Rician fading with a deterministic LOS component which can be written as

$$\boldsymbol{h}_{A,lk} = \sqrt{\frac{K_{Af,lk}}{1 + K_{Af,lk}}}\overline{\boldsymbol{h}}_{A,lk} + \sqrt{\frac{1}{1 + K_{Af,lk}}}\boldsymbol{h}_{Aw,lk}, \tag{1}$$

where $A \in \{S, D\}$, $\boldsymbol{h}_{A,lk}$ is the kth column of $\boldsymbol{H}_{A,l}$, $\boldsymbol{h}_{Aw,lk} \sim \mathcal{CN}(0, \boldsymbol{I})$ denotes the random components, $\overline{\boldsymbol{h}}_{A,lk}$ is the deterministic components and its mth entries can be expressed as

$$\left[\bar{\boldsymbol{h}}_{A,lk}\right]_m = \exp\left(-\mathrm{j}\frac{2\pi\lambda}{d}(m-1)\sin(\theta_{A,lk})\right), \tag{2}$$

where d is the antenna spacing, λ is the wavelength, $\theta_{A,lk}$ is the arrival angles of the kth source ($A = S$) and the kth destination ($A = D$) to the lth selected RS, respectively. For convenience, we will set $d = \lambda/2$ in the rest of this paper.

Since all L selected RSs operate in full duplex mode, the received signal at the lth RS is interfered by the signals transmitted from its own transmit antennas and other RS's transmit antennas, which is called the loop interference and the cross-talk interference, respectively [23]. Then the received signals at the lth relay stations and at the kth destination are given by

$$\boldsymbol{y}_{R,l}(t) = \sqrt{p_S}\boldsymbol{G}_{S,l}\boldsymbol{x}(t) + \sqrt{p_R}\sum_{j=1}^{L}\boldsymbol{G}_{R,lj}\boldsymbol{s}_j(t) + \boldsymbol{n}_{R,l}(t), \tag{3}$$

$$y_{D,k}(t) = \sqrt{p_R}\sum_{l=1}^{L}\boldsymbol{g}_{D,lk}^{\mathrm{T}}\boldsymbol{s}_l(t) + n_{D,k}(t), \tag{4}$$

respectively, where p_S and p_R represent the average transmit power of each source and of the L RSs, $\boldsymbol{x}(t)$ and $\boldsymbol{s}_l(t)$ are the transmit vectors from the K sources and the lth selected RS, respectively. $\boldsymbol{G}_{R,lj} \in \mathbb{C}^{N_r \times N_t}$ represents the cross-talk interference channel from the transmit antennas of the jth RS to the receive antennas of the lth RS. Particularly, $\boldsymbol{G}_{R,ll}$ denotes the loop interference channel matrix. Note that all these RSs are connected to the central server, each RS is able to know the signals transmitted from the transmit antennas of its own and other RSs' through the coordination between the central server and the RSs. Thus we assume each relay station subtracts an estimation of the loop interference and the cross talk interference from its input and applies hardware loop interference cancellation [24], then $\boldsymbol{G}_{R,lj}$ can be interpreted as the residual interference due to the imperfect interference cancellation. As a common assumption made in the existing literature, the residual interfering links can be modeled via the Rayleigh fading distribution [7]. Therefore, the elements of $\boldsymbol{G}_{R,lj}$ can be modeled as independent and identically distributed (i.i.d.) $\mathcal{CN}(0, \sigma_{R,l}^2)$ random variables, where $\sigma_{R,l}$ only depends on the capability of the hardware interference cancellation technique of the lth selected RS. In addition, $\boldsymbol{n}_{R,l}(t) \sim \mathcal{CN}(0, \boldsymbol{I}_{N_r})$ and $n_{D,k}(t) \sim \mathcal{CN}(0, 1)$ represent the additive white Gaussian noise (AWGN) received at the lth RS and the kth destination, respectively.

2.2 Channel estimation

In real situations, the channel matrices $\boldsymbol{G}_{A,l}$ is estimated at the lth relay station. For our considered Rician fading channel model, we assume that both the deterministic LOS component and the Rician K-factor are perfectly known at the relay stations, such that the kth column of the estimate of $\boldsymbol{G}_{A,l}$ can be expressed as

$$\hat{\boldsymbol{g}}_{A,lk} = \sqrt{\frac{K_{Af,lk}}{1+K_{Af,lk}}}\overline{\boldsymbol{g}}_{A,lk} + \sqrt{\frac{1}{1+K_{Af,lk}}}\hat{\boldsymbol{g}}_{Aw,lk}, \tag{5}$$

where $\overline{\boldsymbol{g}}_{A,lk} = \sqrt{\beta_{A,lk}}\bar{\boldsymbol{h}}_{A,lk}$ denotes the deterministic components of $\boldsymbol{g}_{A,lk}$. $\hat{\boldsymbol{g}}_{Aw,lk}$ represents the estimates of the random components $\boldsymbol{g}_{Aw,lk} = \sqrt{\beta_{A,lk}}\boldsymbol{h}_{Aw,lk}$.

In this paper, we consider that the channels are estimated using pilot sequences. In the training stage, all sources and destinations simultaneously transmit their orthogonal pilot sequences of τ ($\tau \geqslant 2K$) symbols to the RSs. Then the lth selected RS uses the minimum mean square error (MMSE) metric to estimate $\boldsymbol{G}_{Aw,l} = [\boldsymbol{g}_{Aw,l1}, \ldots, \boldsymbol{g}_{Aw,lK}]$. According to [25], the estimation of $\boldsymbol{G}_{Aw,l}$ can be shown as

$$\hat{\boldsymbol{G}}_{Aw,l} = \left(\boldsymbol{G}_{Aw,l} + \frac{1}{\sqrt{\tau p_P}}\boldsymbol{N}_{A,p}\right)\tilde{\boldsymbol{D}}_{A,l}, \tag{6}$$

where p_P is the transmit power of each pilot symbol, $\boldsymbol{N}_{A,p} \sim \mathcal{CN}(0, \boldsymbol{I})$, $\tilde{\boldsymbol{D}}_{A,l} = (\frac{1}{\tau p_P}\boldsymbol{D}_{A,l}^{-1} + \boldsymbol{I}_K)^{-1}$. Let $\boldsymbol{\varepsilon_{Aw,l}} = \boldsymbol{G}_{Aw,l} - \hat{\boldsymbol{G}}_{Aw,l}$ be the estimate error matrices of $\boldsymbol{G}_{Aw,l}$. Then, from the property of MMSE estimation, the rows of $\hat{\boldsymbol{G}}_{Aw,l}$, $\boldsymbol{\varepsilon_{Aw,l}}$ are independent and distributed as $\mathcal{CN}(0, \hat{\boldsymbol{D}}_{A,l})$, $\mathcal{CN}(0, \boldsymbol{D}_{A,l} - \hat{\boldsymbol{D}}_{A,l})$, respectively, where $\hat{\boldsymbol{D}}_{A,l}$ is diagonal matrices whose kth diagonal entries are

$$\sigma_{A,lk}^2 = \frac{\tau p_P \beta_{A,lk}^2}{1 + \tau p_P \beta_{A,lk}}.$$

2.3 Data transmission with MRC/MF and ZF processing

In our considered system architecture, each of the selected RSs uses linear processing to decode the signals transmitted from the K sources while employing linear precoding to forward the signals to the K destinations. Let $\boldsymbol{W}_l^{\mathrm{T}}$ be the linear receiver matrix and $\boldsymbol{A}_l$ be the precoding matrix of the lth selected RS. After linear reception, the estimated signal for the kth source is given by

$$r_{R,lk}(t) = \sqrt{p_S}\sum_{i=1}^{K}\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,li}x_i(t) + \sqrt{p_R}\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,ll}\boldsymbol{s}_l(t) + \sqrt{p_R}\boldsymbol{w}_{lk}^{\mathrm{T}}\sum_{j\neq l}^{L}\boldsymbol{G}_{R,lj}\boldsymbol{s}_j(t) + \boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{n}_{R,l}(t), \tag{7}$$

where $\boldsymbol{w}_{lk}^{\mathrm{T}}$ is the kth column of $\boldsymbol{W}_l^{\mathrm{T}}$. Meanwhile, through the high-speed optical connection, the central server instructs each selected RSs to precode a time-delay version of $\boldsymbol{x}(t)$ and broadcasts the precoded symbols to all K destinations. More explicitly, $\boldsymbol{s}_l(t) = \boldsymbol{A}_l\boldsymbol{x}(t - t_d)$ for $l = 1, \ldots, L$, where $t_d > 1$ guarantees that the receive and transmit signals at the RSs are uncorrelated. Therefore, the received signal at the kth destination can be expressed as

$$y_{D,k}(t) = \sqrt{p_R}\sum_{l=1}^{L}\sum_{i=1}^{K}\boldsymbol{g}_{D,lk}^{\mathrm{T}}\boldsymbol{a}_{lk}x_i(t - t_d) + n_{D,k}(t), \tag{8}$$

where $\boldsymbol{a}_{lk}$ is the kth column of $\boldsymbol{A}_l$.

In this paper, we consider two prominent linear processing techniques, MRC/MF processing and ZF processing: (i) for MRC/MF processing case, $\boldsymbol{W}_l^{\mathrm{T}} = \hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}$ and $\boldsymbol{A}_l = \alpha_{\mathrm{MF},l}\hat{\boldsymbol{G}}_{D,l}^{*}$; (ii) for ZF processing case, $\boldsymbol{W}_l^{\mathrm{T}} = (\hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}\hat{\boldsymbol{G}}_{S,l})^{-1}\hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}$ and $\boldsymbol{A}_l = \alpha_{\mathrm{ZF},l}\hat{\boldsymbol{G}}_{D,l}^{*}(\hat{\boldsymbol{G}}_{D,l}^{\mathrm{T}}\hat{\boldsymbol{G}}_{D,l}^{*})^{-1}$, where both $\alpha_{\mathrm{MF},l}$ and $\alpha_{\mathrm{ZF},l}$ are the normalization constants to satisfy a long term total transmit power constraint, such that $\mathrm{E}\{\mathrm{tr}\{\boldsymbol{A}_l\boldsymbol{A}_l^{\mathrm{H}}\}\} = 1$, at the lth relay.

3 Achievable rate analysis and performance evaluation

3.1 Achievable rate analysis

In this section, we investigate the end-to-end achievable rate of the transmission link from the kth source to the kth destination via the help of L selected RSs for MRC/MF and ZF processing in the transmission mode L. Since it is sufficient that at least one relay succeeds in decoding and the optical connections then share the information to all other relays, therefore the rate of the source-relay link is limited by the strongest source-relay rate. However, in order to avoid data accumulation, the end-to-end achievable rate is limited by the weakest link, i.e., the end-to-end achievable rate equals to the minimum of the transmission between the source-relay and relay-destination link [26]. Therefore, the ergodic end-to-end achievable rate can be given by

$$R_k^{(\mathcal{L})} = \min\left\{\max\left\{R_{S,1k}^{(\mathcal{L})}, \ldots, R_{S,Lk}^{(\mathcal{L})}\right\}, R_{D,k}^{(\mathcal{L})}\right\}, \tag{9}$$

where $R_{S,lk}^{(\mathcal{L})}$ and $R_{D,k}^{(\mathcal{L})}$ are the achievable rates of the transmission link from the kth source to the lth RS and from all selected RSs to the kth destination under the given subset $\mathcal{L}$ of RSs and given by (10) and (11).

$$R_{S,lk}^{(\mathcal{L})} = \mathrm{E}\left\{\log_2\left(1 + \frac{p_S|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}|^2}{p_S\sum_{i\neq k}^{K}|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,li}|^2 + p_R\|\sum_{j=1}^{|\mathcal{L}|}\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,lj}\boldsymbol{A}_j\|^2 + \|\boldsymbol{w}_{lk}^{\mathrm{T}}\|^2}\right)\right\}, \tag{10}$$

$$R_{D,k}^{(\mathcal{L})} = \mathrm{E}\left\{\log_2\left(1+\frac{p_R|\sum_{j=1}^{|\mathcal{L}|}\boldsymbol{g}_{D,jk}^{\mathrm{T}}\boldsymbol{a}_{jk}|^2}{p_R\sum_{i\neq k}^{K}|\sum_{j=1}^{|\mathcal{L}|}\boldsymbol{g}_{D,jk}^{\mathrm{T}}\boldsymbol{a}_{ji}|^2+1}\right)\right\}. \tag{11}$$

Next we employ the widely used technique from [27] to obtain the closed-form expression for the achievable rate since the technique does not require instantaneous CSI at the destinations. Thus, for a given subset $\mathcal{L}$ of indices of the selected RSs', the end-to-end achievable rate of the transmission link is given by

$$\tilde{R}_k^{(\mathcal{L})} = \min\left\{\max\left\{\tilde{R}_{S,1k}^{(\mathcal{L})},\ldots,\tilde{R}_{S,Lk}^{(\mathcal{L})}\right\},\tilde{R}_{D,k}^{(\mathcal{L})}\right\}, \tag{12}$$

where $\tilde{R}_{S,lk}^{(\mathcal{L})}$ and $\tilde{R}_{D,k}^{(\mathcal{L})}$ are given by (13) and (14):

$$\tilde{R}_{S,lk}^{(\mathcal{L})} = \log_2\left(1+\frac{p_S|\mathrm{E}\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}\}|^2}{p_S\mathrm{Var}\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}\}+\mathrm{IP}_{S,lk}+\mathrm{LCI}_{lk}+\mathrm{AN}_{lk}}\right), \tag{13}$$

$$\tilde{R}_{D,k}^{(\mathcal{L})} = \log_2\left(1+\frac{p_R|\mathrm{E}\{\sum_{j=1}^{|\mathcal{L}|}\boldsymbol{g}_{D,jk}\boldsymbol{a}_{jk}\}|^2}{p_R\mathrm{Var}\{\sum_{j=1}^{|\mathcal{L}|}\boldsymbol{g}_{D,jk}\boldsymbol{a}_{jk}\}+\mathrm{IP}_{D,k}+1}\right), \tag{14}$$

where

$$\begin{aligned}
&\mathrm{IP}_{S,lk} = p_S\sum_{i\neq k}^{K}\mathrm{E}\left\{\left|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,li}\right|^2\right\}, && \mathrm{IP}_{D,lk} = p_R\sum_{i\neq k}^{K}\mathrm{E}\left\{\left|\sum_{j=1}^{|\mathcal{L}|}\boldsymbol{g}_{D,ji}\boldsymbol{a}_{jk}\right|^2\right\},\\
&\mathrm{LCI}_{lk} = p_R\sum_{j=1}^{|\mathcal{L}|}\mathrm{E}\left\{\left|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,lj}\boldsymbol{s}_j(t)\right|^2\right\}, && \mathrm{AN}_{lk} = \mathrm{E}\left\{\left\|\boldsymbol{w}_{lk}^{\mathrm{T}}\right\|\right\}.
\end{aligned} \tag{15}$$

To obtain the closed-form expression of (12), we first give a key lemma that will be useful in deriving our main results.

Lemma 1. When N is asymptotically large, according to the law of large numbers, we have

$$\frac{1}{N}\hat{\boldsymbol{G}}_{A,l}^{\mathrm{H}}\hat{\boldsymbol{G}}_{A,l}\xrightarrow{\text{a.s.}}\boldsymbol{P}_{A,l}, \tag{16}$$

where $\boldsymbol{P}_{A,l}$ is a diagonal matrix whose kth diagonal entry is $\rho_{A,lk}=\frac{(K_{Af,lk}\beta_{A,lk}+\sigma_{A,lk}^2)}{(1+K_{Af,lk})}$ for $\{A,N\}\in\{\{S,N_r\},\{D,N_t\}\}$.

Proof. According to (5), the inner product between two columns of $\hat{\boldsymbol{G}}_{A,l}$ can be expressed as

$$\frac{1}{N}\hat{\boldsymbol{g}}_{A,lk}^{\mathrm{H}}\hat{\boldsymbol{g}}_{A,li}=\frac{\sqrt{K_{Af,lk}K_{Af,li}}\bar{\boldsymbol{g}}_{A,lk}^{\mathrm{H}}\bar{\boldsymbol{g}}_{A,li}+\hat{\boldsymbol{g}}_{Aw,lk}^{\mathrm{H}}\hat{\boldsymbol{g}}_{Aw,li}}{N\sqrt{(1+K_{Af,lk})(1+K_{Af,li})}}+\frac{\sqrt{K_{Af,lk}}\bar{\boldsymbol{g}}_{A,lk}^{\mathrm{H}}\hat{\boldsymbol{g}}_{Aw,li}+\sqrt{K_{Af,li}}\hat{\boldsymbol{g}}_{Aw,lk}^{\mathrm{H}}\bar{\boldsymbol{g}}_{A,li}}{N\sqrt{(1+K_{Af,lk})(1+K_{Af,li})}}. \tag{17}$$

If $i=k$ and N is asymptotically large, it can be readily obtained that $\frac{1}{N}\hat{\boldsymbol{g}}_{A,lk}^{\mathrm{H}}\hat{\boldsymbol{g}}_{A,lk}\xrightarrow{\text{a.s.}}\rho_{A,lk}$. If $i\neq k$, the only remaining term of (17) is

$$\frac{\sqrt{K_{Af,lk}K_{Af,li}}}{N\sqrt{(1+K_{Af,lk})(1+K_{Af,li})}}\bar{\boldsymbol{g}}_{A,lk}^{\mathrm{H}}\bar{\boldsymbol{g}}_{A,li}=\frac{\sqrt{K_{Af,lk}K_{Af,li}}\mathrm{e}^{\frac{\mathrm{j}(N-1)\pi\left(\sin\left(\theta_{A,lk}\right)-\sin\left(\theta_{A,li}\right)\right)}{2}}}{N\sqrt{(1+K_{Af,lk})(1+K_{Af,li})}}\phi_{A,lki}, \tag{18}$$

where

$$\phi_{A,lki}=\frac{\sin\left(N\pi\left(\sin\left(\theta_{A,lk}\right)-\sin\left(\theta_{A,li}\right)\right)/2\right)}{\sin\left(\pi\left(\sin\left(\theta_{A,lk}\right)-\sin\left(\theta_{A,li}\right)\right)/2\right)}. \tag{19}$$

When $N\to\infty$, we get $\frac{1}{N}\hat{\boldsymbol{g}}_{A,lk}^{\mathrm{H}}\hat{\boldsymbol{g}}_{A,li}\xrightarrow{\text{a.s.}}0$ since $\phi_{A,lki}$ is bounded.

We next provide the closed-form expressions of the end-to-end achievable rate given in (12) for MRC/MF processing and ZF processing.

Theorem 1. With MRC/MF processing, the end-to-end achievable rate, for a finite number of antennas at the L selected RS, is given by

$$\tilde{R}_{\mathrm{MF},k}^{(\mathcal{L})}=\log_2\left(1+\min\left\{\max\left\{\gamma_{S,1k}^{(\mathcal{L})},\ldots,\gamma_{S,Lk}^{(\mathcal{L})}\right\},\gamma_{D,k}^{(\mathcal{L})}\right\}\right), \tag{20}$$

where

$$\gamma_{S,lk}=\frac{p_S N_r \rho_{S,lk}}{\frac{p_S}{N_r\rho_{S,lk}}\sum_{i=1}^{K}\Delta_{S,li}+p_R\sum_{l=1}^{L}\sigma_{R,l}^2+1},\quad \gamma_{D,k}=\frac{p_R N_t(\sum_{j=1}^{L}\mathrm{tr}(\boldsymbol{P}_{D,j})^{-\frac{1}{2}}\rho_{D,jk})^2}{p_R\sum_{i=1}^{K}(\sum_{j=1}^{L}\Delta_{D,ji}+(\sum_{j=1}^{L}\dot{\Delta}_{D,ji})^2)+1},$$

and $\Delta_{S,i}$, $\Delta_{D,i}$ and $\dot{\Delta}_{D,i}$ are defined in (21).

$$\begin{aligned}\Delta_{S,li}&=\frac{\beta_{S,li}(K_{Sf,lk}K_{Sf,li}\beta_{S,lk}\psi_{S,lki}^2+N_r(K_{Sf,lk}\beta_{S,lk}+(1+K_{Sf,li})\sigma_{S,lk}^2))}{(1+K_{Sf,lk})(1+K_{Sf,li})},\\ \Delta_{D,li}&=\frac{\beta_{D,li}(K_{Df,lk}\beta_{D,lk}+(1+K_{Df,li})\sigma_{D,lk}^2)}{\mathrm{tr}(\boldsymbol{P}_{D,j})(1+K_{Df,lk})(1+K_{Df,li})},\\ \dot{\Delta}'_{D,li}&=\psi_{D,lki}\sqrt{\frac{K_{Df,lk}K_{Df,li}\beta_{D,lk}\beta_{D,li}}{N_t\mathrm{tr}(\boldsymbol{P}_{D,j})(1+K_{Df,lk})(1+K_{Df,li})}}.\end{aligned}\tag{21}$$

In addition, $\psi_{S,lki}$ ($\psi_{D,lki}$) equals to $\phi_{S,lki}$ ($\phi_{D,lki}$) for $i\neq k$, and equals to zero otherwise.

Proof. See detail proof in Appendix A.

Theorem 2. With ZF processing, the end-to-end achievable rate, for a finite number of receive antennas at the L selected RSs and $N_t\gg 1$, can be approximated as

$$\tilde{R}_{\mathrm{ZF},k}^{(\mathcal{L})}\approx\log_2\left(1+\min\left\{\max\left\{\gamma_{S,1k}^{(\mathcal{L})},\ldots,\gamma_{S,Lk}^{(\mathcal{L})}\right\},\gamma_{D,k}^{(\mathcal{L})}\right\}\right),\tag{22}$$

where

$$\gamma_{S,lk}^{(\mathcal{L})}=\frac{p_S(N_t-K)}{(p_S\sum_{i=1}^{K}\chi_{S,li}+\frac{p_R\sigma_{R,l}^2}{N_t}\sum_{j=1}^{|\mathcal{L}|}\frac{\mathrm{tr}(\boldsymbol{P}_{D,j}^{-1})}{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{D,j}^{-1})}+1)[\hat{\boldsymbol{\Sigma}}_{S,l}^{-1}]_{kk}},\quad \gamma_{D,k}^{(\mathcal{L})}=\frac{p_R(N_t-K)(\sum_{j=1}^{|\mathcal{L}|}\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{D,j}^{-1})^{1/2})^2}{p_R\sum_{j=1}^{|\mathcal{L}|}\chi_{D,jk}+1},\tag{23}$$

and

$$\chi_{A,lk}=\frac{\beta_{A,lk}-\sigma_{A,lk}^2}{1+K_{Af,lk}},\quad \hat{\boldsymbol{\Sigma}}_{A,j}=\boldsymbol{\Lambda}_{A,l}+\frac{1}{N}\boldsymbol{\Omega}_{A,l}^{1/2}\bar{\boldsymbol{G}}_{A,l}^{\mathrm{H}}\bar{\boldsymbol{G}}_{A,l}\boldsymbol{\Omega}_{A,l}^{1/2},\tag{24}$$

for $\{A,N\}=\{\{S,N_r\},\{D,N_t\}\}$, where $\boldsymbol{\Lambda}_{A,l}$ and $\boldsymbol{\Omega}_{A,l}$ are diagonal matrices whose the kth diagonal entries are $\sigma_{A,lk}^2/(1+K_{Af,lk})$ and $K_{Af,lk}/(1+K_{Af,lk})$, respectively.

Proof. See detail proof in Appendix B.

3.2 Performance evaluation

In this section we evaluate the system performance considering the sum spectral efficiency. Denote T as the length of the coherent interval (in symbols). Thus, the sum spectral efficiency is given by

$$\tilde{S}_B^{(\mathcal{L})}=\frac{T-\tau}{T}\sum_{k=1}^{K}\tilde{R}_{B,k}^{(\mathcal{L})},\tag{25}$$

where $B\in\{\mathrm{MF},\mathrm{ZF}\}$. Note that in the case of ZF processing, $\tilde{R}_{\mathrm{ZF},k}^{(\mathcal{L})}$ is an approximate result. However, we show that this approximation is very tight and fairly accurate in Section 5. For this reason, we hereafter consider the result of ZF processing as exact.

Next we emphasis on the general case where the Rician K-factors are nonzero and $L>1$ and study the power saving by using very large antenna arrays at the RSs.

Theorem 3. Whether p_P is fixed or not, when N_r and N_t go to infinity with the same speed, we can scale down the transmit power for each source and for each selected RS proportionally to $1/N_r$ and $1/N_t$, i.e., $p_S=E_S/N_r$ and $p_R=E_R/N_t$ where E_S and E_R are fixed, while maintaining a desirable rate.

Proof. See detail proof in Appendix C.

Remark 1. The result in Theorem 3 is different with that in [11] which demonstrated that in Rayleigh fading case, if the pilot power and the data power are the same, the transmit power of each source and of the relay can be scaled down proportionally to $1/\sqrt{N_r}$ and $1/\sqrt{N_t}$, respectively. However, for Rician channels with LOS components, the transmit power of sources and selected RSs can be cut down by $1/N_r$ and $1/N_t$ regardless of p_P. This is due to the fact that the LOS propagation reduces fading fluctuations and increases the received SINR. As a result, the transmit power can be scaled down more aggressively for Rician fading channels.

Remark 2. By intuitions, since the loop interference and the cross talk interference are caused by the fact that receive antennas of the lth selected RS are interfered by the signals transmitted from its own and other RSs' transmit antennas. If we scale down the transmit power of each selected RS and make it very low (i.e., approach to zero), the loop interference and the cross talk interference interference can be significantly reduced. According to Theorem 3, we can maintain a desirable spectral efficiency with a low transmit power of selected RSs. Therefore, we conclude that the loop interference and cross talk interference can be cancelled out by using a large transmit antenna array and a scaled-down transmit power.

Corollary 1. According to Theorem 3, when $p_S = E_S/N_r$ and $p_R = E_R/N_t$ for fixed E_S and E_R, where $N_r, N_t \to \infty$, the sum spectral efficiency of MRC/MF and ZF processing are given by

$$\tilde{S}_B^{(\mathcal{L})} \xrightarrow{\text{a.s.}} \frac{T-\tau}{T} \sum_{k=1}^{K} \log_2 \left(1 + \min\left\{\max\left\{E_S\rho_{S,1k}, \ldots, E_S\rho_{S,Lk}\right\}, \gamma_{B,D,k}^{PE}\right\}\right), \tag{26}$$

for $B \in \{\text{MF}, \text{ZF}\}$, where

$$\gamma_{\text{MF},D,k}^{PE} = E_R \left(\sum_{j=1}^{|\mathcal{L}|} \text{tr}\left(\boldsymbol{P}_{D,j}\right)^{-\frac{1}{2}} \rho_{D,jk}\right)^2, \quad \gamma_{\text{ZF},D,k}^{PE} = E_R \left(\sum_{j=1}^{|\mathcal{L}|} \text{tr}(\hat{\boldsymbol{P}}_{D,j}^{-1})^{\frac{1}{2}}\right)^2. \tag{27}$$

If we neglect the large-scale fading effects and assume the Rician K-factors are all the same (i.e., $\beta_{S,lk} = \beta_{D,lk} = 1$ and $K_{Sf,lk} = K_{Df,lk} = K_f$, for $\forall\{l,k\}$), then the sum spectral efficiency of MRC/MF processing and ZF processing converge to the same value and shown as

$$\tilde{S}^{(\mathcal{L})} \xrightarrow{\text{a.s.}} \frac{T-\tau}{T} K \log_2 \left(1 + \rho_1 \min\left(E_S, \frac{L^2 E_R}{K}\right)\right), \tag{28}$$

where $\rho_1 = (K_f + \frac{\tau p_P}{1+\tau p_P})/(1+K_f)$. In particular, when $p_P = p_S = E_S/N_r$ and $N_r \to \infty$, we have $\rho_1 \xrightarrow{\text{a.s.}} K_f/(1+K_f)$.

4 Energy-efficiency-aware relay selection

In previous sections, we mainly focused on the sum spectral efficiency under a given transmission mode L and supposed that the transmit powers of all sources are the same. However, energy efficiency is critical as it affects network lifetime in future wireless network. An interesting issue comes up that how to choose L RSs and allocate different power to different sources to maximize the energy efficiency of the system.

Thus in this section, we assume that the transmit powers of different sources are different. We are interested in designing an energy-efficiency-aware relay selection strategy to maximize the energy efficiency of the relay network subject to the constraints of maximum power transmitted from sources and RSs.

Mathematically, the optimization problem can be formulated as

$$\begin{aligned} \underset{|\mathcal{L}|=L}{\text{maximize}} \quad & \text{EE}_{\text{B}}^{(\mathcal{L})} \\ \text{subject to} \quad & 0 \leqslant p_{S,k} \leqslant p_0, \forall k;\ 0 \leqslant p_R \leqslant p_1, \end{aligned} \tag{29}$$

where p_0 and p_1 are the peak power constraints of $p_{S,k}$ and p_R, respectively. We define the energy efficiency as the sum spectral efficiency divided by the total transmit power

$$\mathrm{EE}_B^{(\mathcal{L})} = \frac{S_B^{(\mathcal{L})}}{\sum_{k=1}^{K} p_{S,k} + Lp_R}. \tag{30}$$

However, finding the optimal energy efficiency of (29) is guaranteed by making an exhaustive search over all reasonable relay combinations of $\binom{M}{L}$ and computing the optimal power allocation for each combination. Although the exhaustive algorithm is feasible for off-line processing, a low complexity approach is of particular interest for adaption to changes in propagation environment. Therefore, we utilize standard alternating optimization: (1) given an initial set of transmit power $\{p_{S,k}, p_R\}$, for $k = 1, \ldots, K$; (2) select $|\mathcal{L}^*| = L$ RSs that maximizes the spectral efficiency with the initial set of transmit power; (3) optimize the transmit power $p_{S,k}$ and p_R to maximize the energy efficiency under the maximum spectral efficiency obtained in step (2). Thus, we next focus on the relay selection strategy with an initial set of transmit power of sources and RSs to maximize the sum spectral efficiency.

4.1 Approximate incremental relay selection strategy

To maximize the sum spectral efficiency, it is conceivable that we need to examine all $\binom{M}{L}$ possible combinations, which lead to the prohibitively high computational complexity as M gets large.

Alternatively, we consider the incremental selection strategy as a suboptimal approach with low complexity [28]. The algorithm only examines the spectral efficiency for $L(M-(L-1)/2)$ different combinations, resulting in much less complexity than the exhaustive approach. Conventionally, to implement the incremental selection strategy, the central server needs to have the knowledge of the instantaneous CSI between all RSs and K communication pairs. However since each RS is equipped with a large antenna array, the system requires significant overhead to deliver the instantaneous CSI to the central server which may cause congestion in the optical connection. Therefore, this is not feasible in practice.

We note that in massive MIMO setup, things that are random become deterministic and the lower bound of the spectral efficiency is fairly accurate. Thus, we propose a sub-optimal relay selection strategy and call it as the approximate incremental relay selection in that the central server only needs to calculate the lower bound of the spectral efficiency to complete the selection among the M relay stations. In particular, the idea of this algorithm is to determine the subset of $\mathcal{L}^*$ through L steps, where one RS is selected in each step, as shown in Algorithm 1. More precisely, in each step, one RS is added to the RS subset denoted by $\bar{\mathcal{L}}$, where the selection criterion is to maximize the sum spectral efficiency of the updated RS subset $\bar{\mathcal{L}}$. Then after L steps, we have totally added L RSs into $\bar{\mathcal{L}}$ and then assign $\mathcal{L}^* = \bar{\mathcal{L}}$.

Note that, employing our proposed selection strategy has the advantage that the central server does not need the feedback of instantaneous CSI. Hence, the overhead is much smaller than the conventional selection strategy.

Algorithm 1 Approximate incremental relay selection strategy

Initialization: Let $\bar{\mathcal{M}} = \{1, \ldots, M\}$, $\bar{\mathcal{L}} = \emptyset$ and $Z = |\bar{\mathcal{M}}|$, % Use $\bar{\mathcal{L}}$ and $\bar{\mathcal{M}}$ to store all selected RSs and all other RSs, respectively.

1: **for** $l = 1 : L$ % Add one RS to $\bar{\mathcal{L}}$ in each step. **do**
2: **for** $z = 1 : Z$ **do**
3: $\hat{\mathcal{L}}_z = \bar{\mathcal{L}} \cup \{\mathcal{M}_z\}$, % $\bar{\mathcal{M}}_z$ is the zth element of $\bar{\mathcal{M}}$.
4: $\tilde{S}_{B,z} = \tilde{S}_B^{(\hat{\mathcal{L}}_z)}$ based on (25), % Calculate the spectral efficiency of $\hat{\mathcal{L}}_z$.
5: **end for**
6: $z^* = \arg\max_z \ \tilde{S}_{B,z}$, % Select the RS to maximize the spectral efficiency.
7: $\bar{\mathcal{L}} = \hat{\mathcal{L}}_{z^*}$, $\bar{\mathcal{M}} = \bar{\mathcal{M}} \setminus \hat{\mathcal{L}}_{z^*}$, $Z = |\bar{\mathcal{M}}|$. % Update the set $\bar{\mathcal{L}}$ and $\bar{\mathcal{M}}$.
8: **end for**
9: $\mathcal{L}^* = \bar{\mathcal{L}}$. % Complete the RS selection and obtain $\mathcal{L}^*$.

4.2 Optimal power allocation

After selecting L RSs that maximize the sum spectral efficiency, we then focus on the power allocation to maximize the energy efficiency, subject to the maximum spectral efficiency and the constraints of maximum powers transmitted from the sources and the L selected relay stations. Therefore, the optimization problem can be formulated as

$$\begin{aligned} \underset{|\mathcal{L}^*|=L}{\text{maximize}} \quad & \mathrm{EE}_{\mathrm{B}}^{(\mathcal{L}^*)} \\ \text{subject to} \quad & \tilde{S}_B^{(\mathcal{L}^*)} = \tilde{S}_{B0}^{(\mathcal{L}^*)}; \\ & 0 \leqslant p_{S,k} \leqslant p_0, \forall k;\ 0 \leqslant p_R \leqslant p_1, \end{aligned} \tag{31}$$

where $\tilde{S}_{B0}^{(\mathcal{L}^*)}$ is the maximum spectral efficiency achieved by the $|\mathcal{L}^*| = L$ selected RSs. From (20), (22) and (25), the optimal power allocation problem in (31) is equivalent to

$$\begin{aligned} \underset{|\mathcal{L}^*|=L}{\text{minimize}} \quad & \sum_{k=1}^{K} p_{S,k} + L p_R \\ \text{subject to} \quad & \prod_{k=1}^{K} (1+\gamma_{B,k}) = 2^{\frac{T}{T-\tau} S_{B0}^{(\mathcal{L}^*)}} \\ & \gamma_{B,k} \leqslant \max\left\{ \frac{p_{S,k} a_{B,lk}}{p_{S,k} \sum_{i=1}^{K} b_{B,lk} + p_R \sum_{j=1}^{L} c_{B,lk} + 1} \mid l = 1, \ldots, L \right\}, \\ & \gamma_{B,k} \leqslant \frac{p_R d_{B,k}}{p_R \sum_{j=1}^{L} e_{B,jk} + 1};\ 0 \leqslant p_{S,k} \leqslant p_0;\ 0 \leqslant p_R \leqslant p_1, \end{aligned} \tag{32}$$

for $B = \{\mathrm{MF}, \mathrm{ZF}\}$, where $a_{B,lk}, b_{B,lk}, c_{B,lk}, d_{B,k}, e_{B,jk}$ are constant value which are depended on the MRC/MF and ZF processing. More explicitly, we have

- For MRC/MF processing,

$$\begin{aligned} & a_{\mathrm{MF},lk} = N_r \rho_{S,lk}, && b_{\mathrm{MF},lk} = \Delta_{S,li}/N_r \rho_{S,lk}, && c_{\mathrm{MF},lk} = \sigma_{R,l}^2, \\ & d_{\mathrm{MF},k} = \left(\sum_{j=1}^{L} \mathrm{tr}(\boldsymbol{P}_{D,j})^{-\frac{1}{2}} \rho_{D,jk} \right)^2, && e_{\mathrm{MF},jk} = \sum_{i=1}^{K} (N_t \mathrm{tr}(\boldsymbol{P}_{D,j}))^{-1} \Delta_{D,jk}. \end{aligned} \tag{33}$$

- For ZF processing,

$$\begin{aligned} & a_{\mathrm{ZF},lk} = (N_t - K) / \left[\hat{\boldsymbol{\Sigma}}_{S,l}^{-1} \right]_{kk}, && b_{\mathrm{ZF},lk} = \chi_{S,li}, \quad c_{\mathrm{ZF},lk} = \sigma_{R,l}^2 \mathrm{tr}\left(\boldsymbol{P}_{D,j}^{-1} \right) / N_t \mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_{D,j}^{-1} \right), \\ & d_{\mathrm{ZF},k} = (N_t - K) \left(\sum_{j=1}^{L} \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{D,j}^{-1})^{\frac{1}{2}} \right)^2, && e_{\mathrm{ZF},jk} = \chi_{D,jk}. \end{aligned} \tag{34}$$

We can see that the optimal problem in (32) is very similar to a geometric programming (GP) which can be reformulated as a convex problem. However, the problem cannot be solved directly since the equality constraint in (32) is polynomial function. Thus, we are trying to approximate the equality constraint to a monomial function. According to [29, Lemma 1], we can use $\kappa_{B,k} \gamma_{B,k}^{\eta_{B,k}}$ to approximate $1 + \gamma_{B,k}$ near a point $\hat{\gamma}_{B,k}$, where $\eta_{B,k} = \hat{\gamma}_{B,k}(1 + \hat{\gamma}_{B,K})^{-1}$ and $\kappa_{B,k} = \hat{\gamma}_{B,k}^{-\eta_{B,k}}(1 + \hat{\gamma}_{B,k})$. Therefore, the equality constraint of (32) can be rewritten as

$$\prod_{k=1}^{K} \kappa_{B,k} \gamma_{B,k}^{\eta_{B,k}} = 2^{\frac{T}{T-\tau} \tilde{S}_{B0}^{(\mathcal{L}^*)}}, \tag{35}$$

which is a monomial function.

In summary, we formulate the algorithm to handle the energy-efficiency-aware relay selection algorithm shown in Algorithm 2.

Algorithm 2 Energy-efficiency-aware relay selection

Initialization: Given a initial set $\{p_{S,k}, p_R\}$ for $k = 1, \ldots, K$. Define the maximum number of iterations I, a tolerance ϵ and a parameter μ.

1: Implement Algorithm 1 to select RSs with the subset $\mathcal{L}^*$ and obtain the maximum sum spectral efficiency $S_B^{(\mathcal{L}^*)}$.
2: Let $i = 1$, $S_{B,0}^{(\mathcal{L}^*)} = S_B^{(\mathcal{L}^*)}$ and choose the initial values of $\gamma_{B,k}$ for $k = 1, \ldots, K$.
3: **while** $i \leqslant I$ **do**
4: Solve the optimal problem of (32) by replacing the first constraint to (35).
5: Let $\gamma_{B,k}^*$ for $k = 1, \ldots, K$ be the solutions.
6: **if** $\max\limits_k |\gamma_{B,k} - \gamma_{B,k}^*| < \epsilon$ **then**
7: break;
8: **end if**
9: **end while**
10: Obtain the optimum $p_{S,k}^*$ and p_R^*

5 Numerical results

For our simulations, we consider a distributed full duplex relay network with $M = 7$ relay stations, the number of communication pairs $K = 5$, the length of pilot sequence $\tau = 2K$. We suppose the Rician K-factors between K communication pairs and L selected RS are all the same such that $K_{Sf,lk} = K_{Df,lk} = K_f$ for $\forall\{l, k\}$ and the level of the loop interference and the cross talk interference $\sigma_{R,l} = \sigma_R$ for $l = 1, \ldots, L$. The parameter of the coherent interval is chosen according to the LTE standard, i.e., $T \approx 200$ symbols. In all these plots, "Simulation" represents the results obtained by utilizing Monte-Carlo simulation, and "Analysis" denotes the analytical results for large scale system.

Firstly we evaluate the validity of the end-to-end achievable rate given by (9) as well as the closed-form expressions given in Theorems 1 and 2. Figure 2 shows the sum rate versus SNR for MRC/MF and ZF processing. We assume the level of the loop interference and the cross talk interference $\sigma_{R,l}^2 = 0$ dB for $l = 1, \ldots, L$, the large scale fading coefficient $\beta_{S,lk} = \beta_{D,lk} = 1$ for $\forall\{l, k\}$ and the K-factor $K_f = 10$. Obviously, we can see that the relative performance gap between the "Simulation" curves and the "Analysis" curves is very small. In particular, for ZF processing, we can see that our derived approximate expression is very tight for large antenna array. Therefore, we conclude that using the mean channel gain for signal detection is fairly reasonable, and the closed-form expressions of the achievable rate are good predictors of the system performance.

Figure 3 investigates the power efficiency using large antenna array when the transmit power of the sources is $p_S = E_S/N_r$ and of the selected RSs is $p_R = E_R/N_t$. In the simulations, we choose $E_S = E_R = 0$ dB and the K-factors $K_f = 10$. Once again, a precise agreement can be seen between the simulation results and our analytical results for both MRC/MF and ZF processing cases. We also observe that the sum spectral efficiencies of MRC/MF and ZF processing tend to the same value as N_r and N_t growing large for $p_P = E_S/N_r$ and fixed $p_P = 10$ dB, respectively. This is owing to the benefit of the LOS propagation, which can reduce the fading fluctuations and enhance the received SINR. In particular, for the cases of $\sigma_R^2 = 0$ dB and $\sigma_R^2 = 10$ dB, both of the sum spectral efficiencies converge to a fixed value. This implies that the loop interference and the cross talk interference can be canceled out by scaling down the transmit power of the RSs proportionally to $1/N_t$ when N_t grows to infinity.

Next we give numerical examples for the sum spectral efficiency and the energy efficiency performance of our proposed approximate incremental relay selection strategy. In addition, we use the conventional exhaustive search relay selection as a reference selection scheme, which employs the instantaneous CSI to exhaustively select L optimal RSs. We consider a more practical scenario that incorporates small-scale fading and large-scale fading. We also assume that all the sources and destinations are located at random inside a disk with a radius of 500 m and all the M RSs are randomly located in the center of the disk with a radius of 200 m. The large-scale fading coefficient is modeled as $\beta_{A,lk} = z_{A,lk}/(r_{A,lk}/r_0)^\nu$ for $A \in \{S, D\}$ [21], where $z_{A,lk}$ is a log-normal random variable with standard deviation of 8 dB, $r_{A,lk}$ is the distance between the kth source (destination) and the lth RS, $\nu = 3.8$ is the path loss exponent and $r_0 = 200$ m.

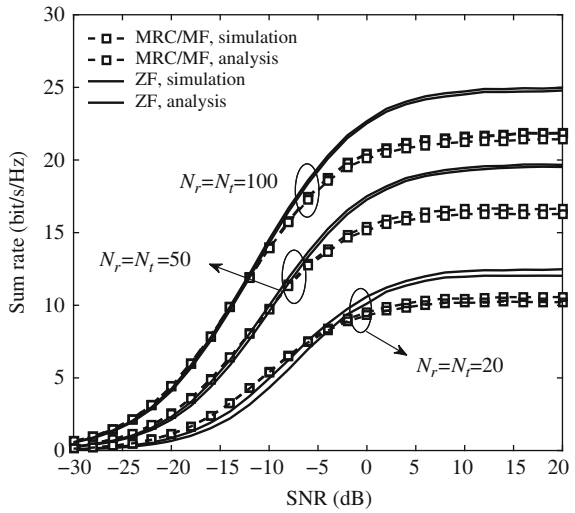

Figure 2 Sum rate versus SNR for MRC/MF and ZF processing with $K = 5$, $L = 3$, $\sigma_R^2 = 0$ dB, $K_f = 10$, and $p_S = p_R = p_P = \text{SNR}$.

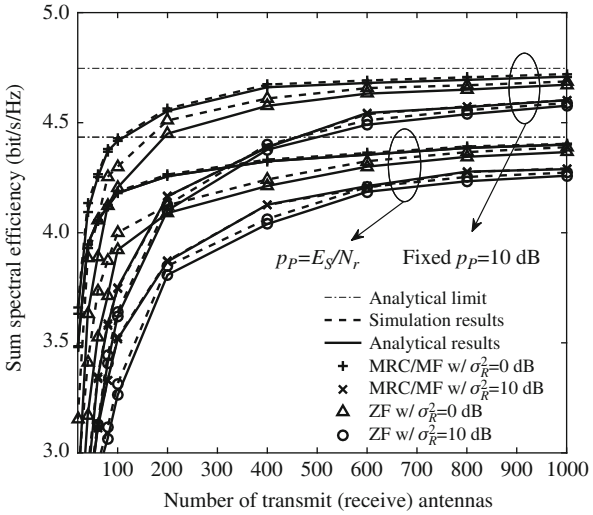

Figure 3 Sum rate versus the number of transmit (receive) antennas with scaled-down power $p_S = E_S/N_r$ and $p_R = E_R/N_t$, where $E_S = E_R = 0$ dB.

Figure 4 illustrates the comparison of the cumulative distribution of the sum spectral efficiency performance for the two relay selection strategies with $L = 3$, $K_f = 10$, $\sigma_R^2 = 0$ dB, $p_P = p_S = p_R = 10$ dB. "exhaustive strategy" denotes the exhaustive relay selection strategy and "proposed strategy" stands for our proposed approximate incremental relay selection strategy obtained via Algorithm 1. Not surprisingly, the "exhaustive strategy" outperforms our "proposed strategy" for both MRC/MF and ZF processing cases, since the former employs the instantaneous CSI to exhaustively search $\binom{M}{L}$ relay combinations to maximize the sum spectral efficiency, while the latter only utilizes the statistical CSI and much lower complexity method to achieve the same aim. However, it can be found that the gap between the "exhaustive strategy" curves and the "proposed strategy" curves are very small for the cases of $N_r = N_t = 50$ and $N_r = N_t = 200$. Therefore, we conclude that our proposed strategy can be seen as suboptimal with low complexity.

Figure 5 shows the energy efficiency versus the sum spectral efficiency with $N_r = N_t = 100$, $L = 3$ and $p_P = 10$ dB. In each subplot, three curves are plotted: (1) the energy efficiency with power allocation which is obtained by solving the problem of (29), with legend "exhaustive strategy w/ power allocation"; (2) the energy efficiency with power allocation obtained via Algorithm 2, with legend "proposed strategy w/ power allocation"; (3) the energy efficiency without power allocation where all sources and all L selected RSs use their maximum powers, with legend "exhaustive strategy w/o power allocation". Obviously, the system performance can be significantly improved with optimal power allocation, especially in the high spectral efficiency region. For instance, compared to the case of no power allocation, the exhaustive selection strategy with optimal power allocation can improve the energy efficiency by factors of 2.5 and 2 for MRC/MRT and ZF processing, respectively, to achieve the same sum spectral efficiency of 7 bits/s/Hz. Moreover, we can see that our proposed relay selection strategy performs very close to the exhaustive one and the energy efficiency difference between the "exhaustive strategy w/ power allocation" curve and the "proposed strategy w/ power allocation" is very small. For example, the energy efficiency obtained via Algorithm 2 is about 0.8 bit/J worse than that obtained by solving the problem of (29) to achieve the sum spectral efficiency of 7 bits/s/Hz for both MRC/MF and ZF processing.

6 Conclusion

In this paper, we analyze a distributed full duplex relay network, where the RSs are equipped with massive arrays, while all sources and destinations are equipped with a single antenna each. We model the channel as Rician fading and assume that the RSs employ MRC/MF and ZF to process the signals. First

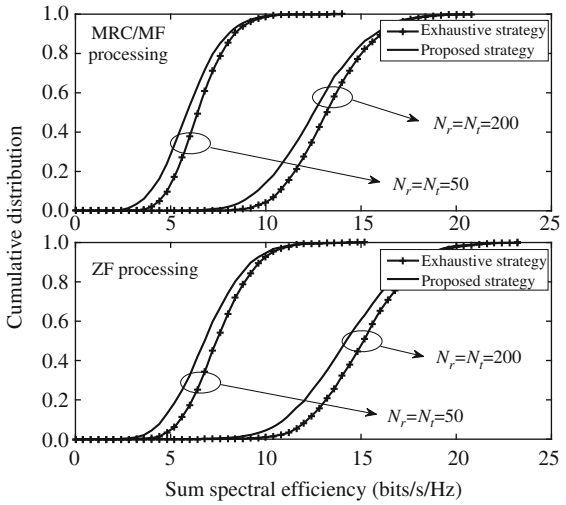

Figure 4 Cumulative distribution of the sum spectral efficiency for the exhaustive and the proposed approximate incremental relay selection strategy with $K = 5$, $L = 3$, $\sigma_R^2 = 0$ dB, $p_S = p_R = 0$ dB and $p_P = 10$ dB.

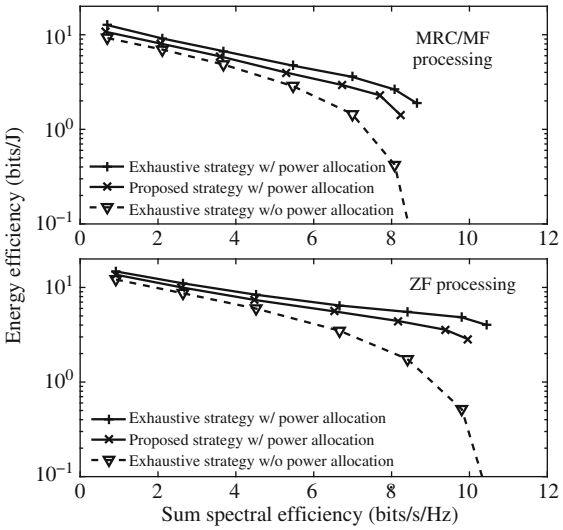

Figure 5 Energy efficiency versus sum spectral efficiency for MRC/MF and ZF processing with $K = 5$, $N_r = N_t = 100$, $L = 3$, $\sigma_{R,l}^2 = 0$ dB and $p_P = 10$ dB.

we derive closed-form expressions of the end-to-end achievable rate. These results are used to evaluate the system performance of the power efficiency and how the sum spectral efficiency vary with the Rician K-factor. Then based on the closed-form expressions, we work on the relay selection strategy and propose an energy-efficiency-aware relay selection scheme to maximize the energy efficiency of the relay network. Compared with the optimal relay selection strategy, the complexity of our proposed strategy is low and only statistical CSI are required. We demonstrate that the energy efficiency of the system can be significantly improved with power allocation and our proposed sub-optimal strategy performs close to the optimal one.

Acknowledgements This work was supported in part by National High-tech R&D Program of China (863) (Grant No. 2014AA01A706), Fundamental Research Funds for the Central Universities (Grant No. 2014JBZ001), National Natural Science Foundation of China (Grant No. 61471027), Beijing Nova Programme (Grant No. xx2016023), Research Fund of National Mobile Communications Research Laboratory, Southeast University (Grant No. 2014D05), and Beijing Natural Science Foundation Project (Grant No. 4152043). We would like to thank the anonymous referees for their help in improving this paper.

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Marzetta T. Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans Wirel Commun, 2010, 9: 3590–3600
2 Larsson E, Edfors O, Tufvesson F, et al. Massive MIMO for next generation wireless systems. IEEE Commun Mag, 2014, 52: 186–195
3 Ngo H Q, Larsson E, Marzetta T. Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans Commun, 2013, 61: 1436–1449
4 Laneman J, Tse D, Wornell G W. Cooperative diversity in wireless networks: efficient protocols and outage behavior. IEEE Trans Inf Theory, 2004, 50: 3062–3080
5 Wang T, Cano A, Giannakis G, et al. High-performance cooperative demodulation with decode-and-forward relays. IEEE Trans Commun, 2007, 55: 1427–1438
6 Riihonen T, Werner S, Wichman R. Hybrid full-duplex/half-duplex relaying with transmit power adaptation. IEEE Trans Wirel Commun, 2011, 10: 3074–3085
7 Riihonen T, Werner S, Wichman R. Mitigation of loopback self-interference in full-duplex MIMO relays. IEEE Trans Signal Process, 2011, 59: 5983–5993
8 Choi J I, Jain M, Srinivasan K, et al. Achieving single channel, full duplex wireless communication. In: Proceedings

of the 16th Annual International Conference on Mobile Computing and Networking, Chicago, 2010. 1–12
9 Duarte M, Sabharwal A. Full-duplex wireless communications using off-the-shelf radios: feasibility and first results. In: Proceedings of the 44th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2010. 1558–1562
10 Everett E, Duarte M, Dick C, et al. Empowering full-duplex wireless communication by exploiting directional diversity. In: Proceedings of the 45th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2011. 2002–2006
11 Ngo H Q, Suraweera H, Matthaiou M, et al. Multipair full-duplex relaying with massive arrays and linear processing. IEEE J Sel Areas Commun, 2014, 32: 1721–1737
12 Bletsas A, Khisti A, Reed D, et al. A simple cooperative diversity method based on network path selection. IEEE J Sel Areas Commun, 2006, 24: 659–672
13 Costa D, Aissa S. Performance analysis of relay selection techniques with clustered fixed-gain relays. IEEE Signal Process Lett, 2010, 17: 201–204
14 Krikidis I, Suraweera H, Smith P, et al. Full-duplex relay selection for amplify-and-forward cooperative networks. IEEE Trans Wirel Commun, 2012, 11: 4381–4393
15 Rui X, Hou J, Zhou L. On the performance of full-duplex relaying with relay selection. Electron Lett, 2010, 46: 1674–1676
16 Cui H Y, Ma M, Song L Y, et al. Relay selection for two-way full duplex relay networks with amplify-and-forward protocol. IEEE Trans Wirel Commun, 2014, 13: 3768–3777
17 Rappaport T, Sun S, Mayzus R, et al. Millimeter wave mobile communications for 5G cellular: It will work. IEEE Access, 2013, 1: 335–349
18 Swindlehurst A L, Ayanoglu E, Heydari P, et al. Millimeter-wave massive MIMO: the next wireless revolution. IEEE Commun Mag, 2014, 52: 56–62
19 Brady J, Behdad N, Sayeed A. Beamspace MIMO for millimeter-wave communications: system architecture, modeling, analysis, and measurements. IEEE Trans Antenn Propag, 2013, 61: 3814–3827
20 Sayeed A, Brady J. Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies. In: Proceedings of IEEE Global Communications Conference, Atlanta, 2013. 3679–3684
21 Zhang Q, Jin S, Wong K K, et al. Power scaling of uplink massive MIMO systems with arbitrary-rank channel means. IEEE J Sel Top Signal Process, 2014, 8: 966–981
22 Suraweera H A, Ngo H Q, Duong T Q, et al. Multi-pair amplify-and-forward relaying with very large antenna arrays. In: Proceedings of IEEE International Conference on Communications, Budapest, 2013. 4635–4640
23 Liu Y, Xia X G, Zhang H. Distributed linear convolutional space-time coding for two-relay full-duplex asynchronous cooperative networks. IEEE Trans Wirel Commun, 2013, 12: 6406–6417
24 Riihonen T, Werner S, Wichman R. Optimized gain control for single-frequency relaying with loop interference. IEEE Trans Wirel Commun, 2009, 8: 2801–2806
25 Kay S M. Fundamentals of Statistical Signal Processing: Estimation Theory. Upper Saddle River: Prentice Hall, 1993. 380–392
26 Riihonen T, Werner S, Wichman R. Transmit power optimization for multiantenna decode-and-forward relays with loopback self-interference from full-duplex operation. In: Proceedings of the 45th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2011. 1408–1412
27 Yang H, Marzetta T. Performance of conjugate and zero-forcing beamforming in large-scale antenna systems. IEEE J Sel Areas Commun, 2013, 31: 172–179
28 Gharavi A M, Gershman A. Fast antenna subset selection in MIMO systems. IEEE Trans Signal Process, 2004, 52: 339–347
29 Weeraddana P, Codreanu M, Latva-aho M, et al. Resource allocation for cross-layer utility maximization in wireless networks. IEEE Trans Veh Technol, 2011, 60: 2790–2809

Appendix A

We firstly provide the proof for the MRC receiver case. Since $\boldsymbol{W}_l^{\mathrm{T}} = \hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}$, we have $\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk} = \left\|\hat{\boldsymbol{g}}_{S,lk}\right\|^2 + \hat{\boldsymbol{g}}_{S,lk}^{\mathrm{H}}\boldsymbol{\varepsilon}_{S,lk}$. Therefore, according to Lemma 1, $\mathrm{E}\left\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}\right\} = N_r\rho_{S,lk}$.

- Compute $\mathrm{Var}\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}\}$: According to the Lindeberg-Levy central limit theorem, the variance of $\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}$ is given by

$$\mathrm{Var}\left\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}\right\} = \frac{N_r\beta_{S,lk}}{\left(1+K_{Sf,lk}\right)^2}\left(K_{Sf,lk}\beta_{S,lk} + \left(1+K_{Sf,lk}\right)\sigma_{S,lk}^2\right). \tag{A1}$$

- Compute $\mathrm{IP}_{S,lk}$: According to Lemma 1, we have

$$\mathrm{E}\left\{\left|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,li}\right|^2\right\} = \frac{N_r\beta_{S,li}\left(K_{Sf,lk}\beta_{S,lk} + \left(1+K_{Sf,li}\right)\sigma_{S,lk}^2\right)}{\left(1+K_{Sf,lk}\right)\left(1+K_{Sf,li}\right)} + \frac{K_{Sf,lk}K_{Sf,li}\beta_{S,lk}\beta_{S,li}}{\left(1+K_{Sf,lk}\right)\left(1+K_{Sf,li}\right)}\phi_{S,lki}^2, \tag{A2}$$

where $\phi_{S,lki}$ is defined in Lemma 1.

- Compute LCI_{lk}: Substituting $\boldsymbol{A}_j = \alpha_{\mathrm{MF},j}\hat{\boldsymbol{G}}_{D,j}^*$, we can obtain

$$\mathrm{E}\left\{\left|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,lj}\boldsymbol{A}_j\right|^2\right\} = \alpha_{\mathrm{MF},j}^2\mathrm{E}\left\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,lj}\hat{\boldsymbol{G}}_{D,j}^{\mathrm{H}}\hat{\boldsymbol{G}}_{D,j}\boldsymbol{G}_{R,lj}^{\mathrm{H}}\boldsymbol{w}_{lk}^*\right\} = N_rN_t\alpha_{\mathrm{MF},j}^2\sigma_{R,l}^2\rho_{S,lk}\sum_{i=1}^{K}\rho_{D,ji}, \tag{A3}$$

where

$$\alpha_{\mathrm{MF},j}^2 = \mathrm{E}\left\{\left(\mathrm{tr}\left(\hat{\boldsymbol{G}}_{D,j}^{\mathrm{H}}\hat{\boldsymbol{G}}_{D,j}\right)\right)^{-1}\right\} = \left(N_t \mathrm{tr}(\boldsymbol{P}_{D,j})\right)^{-1}. \tag{A4}$$

- Compute AN_{lk}: According to Lemma 1, we have $\mathrm{AN}_{lk} = N_r\rho_{S,lk}$.

The proof for the MF precoding case follows the same methodology above and we achieved the results given in Theorem 1.

Appendix B

We first provide the proof for the ZF receiver case. Since $\boldsymbol{W}_l^{\mathrm{T}} = (\hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}\hat{\boldsymbol{G}}_{S,l})^{-1}\hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}$, we have $\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk} = 1 + \boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{\varepsilon}_{S,lk}$, where $\boldsymbol{\varepsilon}_{S,lk} = \boldsymbol{\varepsilon}_{Sw,lk}/(1+K_{Sf,lk})$ and $\boldsymbol{\varepsilon}_{Sw,lk}$ is the kth column of $\boldsymbol{\varepsilon}_{Sw,l}$. Since $\boldsymbol{w}_{lk}^{\mathrm{T}}$ and $\boldsymbol{\varepsilon}_{S,lk}$ are independent, we can obtain $\mathrm{E}\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}\} = 1$.

- Compute $\mathrm{Var}\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}\}$: The variance of $\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}$ is given by

$$\mathrm{Var}\left\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,lk}\right\} = \mathrm{E}\left\{\left|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{\varepsilon}_{S,lk}\right|^2\right\} = \frac{(\beta_{S,lk}-\sigma_{S,lk}^2)}{(1+K_{Sf,lk})}\mathrm{E}\left\{\left[\left(\hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}\hat{\boldsymbol{G}}_{S,l}\right)^{-1}\right]_{kk}\right\} = \frac{(\beta_{S,lk}-\sigma_{S,lk}^2)[\hat{\boldsymbol{\Sigma}}_{S,l}^{-1}]_{kk}}{(1+K_{Sf,lk})(N_r-K)}. \tag{B1}$$

Note that the last equality is obtained by approximating $\hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}\hat{\boldsymbol{G}}_{S,l}$ as a central Wishart distribution with covariance matrix $\hat{\boldsymbol{\Sigma}}_{S,l}$ [21].

- Compute $\mathrm{IP}_{S,lk}$: We have that $\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{g}_{S,li} = \boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{\varepsilon}_{S,li}$, for $i \neq k$. Following the same step of (B1) we obtain

$$\mathrm{E}\left\{\left|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{\varepsilon}_{S,li}\right|^2\right\} = \frac{(\beta_{S,li}-\sigma_{S,li}^2)[\hat{\boldsymbol{\Sigma}}_{S,l}^{-1}]_{kk}}{(1+K_{Sf,li})(N_r-K)}. \tag{B2}$$

- Compute LCI_{lk}: Substituting $\boldsymbol{A}_j = \alpha_{\mathrm{ZF},j}\hat{\boldsymbol{G}}_{D,j}^*(\hat{\boldsymbol{G}}_{D,j}^{\mathrm{T}}\hat{\boldsymbol{G}}_{D,j}^*)^{-1}$ and obtain

$$\mathrm{E}\left\{\left|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,lj}\boldsymbol{A}_j\right|^2\right\} = \alpha_{\mathrm{ZF},j}^2\mathrm{E}\left\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,lj}\hat{\boldsymbol{G}}_{D,j}^*\left(\hat{\boldsymbol{G}}_{D,j}^{\mathrm{T}}\hat{\boldsymbol{G}}_{D,j}^*\right)^{-2}\hat{\boldsymbol{G}}_{D,j}^{\mathrm{T}}\boldsymbol{G}_{R,lj}^{\mathrm{H}}\boldsymbol{w}_{lk}^*\right\}. \tag{B3}$$

When $N_t \gg K$, we use Lemma 1 to approximate $\hat{\boldsymbol{G}}_{D,j}^{\mathrm{T}}\hat{\boldsymbol{G}}_{D,j}^*$ as $N_t\boldsymbol{P}_{D,j}$. Therefore, we substitute the approximation into (B3) and obtain

$$\mathrm{E}\left\{\left|\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,lj}\boldsymbol{A}_j\right|^2\right\} \approx \frac{\alpha_{\mathrm{ZF},j}^2}{N_t^2}\mathrm{tr}\left(\boldsymbol{P}_{D,j}^{-1}\right)\mathrm{E}\left\{\boldsymbol{w}_{lk}^{\mathrm{T}}\boldsymbol{G}_{R,lj}\boldsymbol{G}_{R,lj}^{\mathrm{H}}\boldsymbol{w}_{lk}^*\right\} = \frac{\alpha_{\mathrm{ZF},j}^2\sigma_{R,l}^2}{N_t}\mathrm{tr}\left(\boldsymbol{P}_{D,j}^{-1}\right)\mathrm{E}\left\{\left[\left(\hat{\boldsymbol{G}}_{S,l}^{\mathrm{H}}\hat{\boldsymbol{G}}_{S,l}\right)^{-1}\right]_{kk}\right\},$$

and

$$\alpha_{\mathrm{ZF},j}^2 = \mathrm{E}\left\{\left(\mathrm{tr}\left(\hat{\boldsymbol{G}}_{D,j}^{\mathrm{T}}\hat{\boldsymbol{G}}_{D,j}^*\right)^{-1}\right)^{-1}\right\} = \frac{(N_t-K)}{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{D,j}^{-1})}. \tag{B4}$$

- Compute AN_{lk}: Following the same step of (B1), we have $\mathrm{AN}_{lk} = [\hat{\boldsymbol{\Sigma}}_{S,l}^{-1}]_{kk}/(N_r-K)$.

Similarly, we can obtain the expression for $R_{D,k}$ and we arrive at (22).

Appendix C

For ZF processing, let $p_S = E_S/N_r^{c_1}$ and $p_R = E_R/N_t^{c_2}$, where $c_1, c_2 > 0$, E_S and E_R are fixed regardless of N_r and N_t. According to the law of large numbers, as $N_r, N_t \to \infty$, we have $\hat{\boldsymbol{\Sigma}}_{D,l} \xrightarrow{\text{a.s.}} \boldsymbol{P}_{D,l}$. The inverse matrix of $\boldsymbol{P}_{D,l}$ can be easily obtained by calculating the reciprocals of its main diagonal entries. Hence, in the large-antenna limit, we have $[\hat{\boldsymbol{\Sigma}}_{D,l}^{-1}]_{kk} \xrightarrow{\text{a.s.}} \rho_{D,lk}^{-1}$, and rewrite (22) as

$$R_k \approx \log_2\left(1+\min\left(\frac{\frac{E_R(1-K/N_t)}{N_t^{c_2-1}}(\sum_{j=1}^L(\sum_{i=1}^K\rho_{D,ji}^{-1})^{-\frac{1}{2}})^2}{\frac{E_R}{N_t^{c_2}}\sum_{j=1}^L\chi_{D,jk}+1}, \frac{\frac{E_S}{N_r^{c_1-1}}(1-K/N_r)}{(\frac{E_S}{N_r^{c_1}}\sum_{i=1}^K\chi_{S,li}+\frac{E_R}{N_t^{c_2}}\sum_{j=1}^L\sum_{i=1}^K\frac{(1-K/N_t)}{\sum_{i=1}^K\rho_{D,ji}^{-1}}\sigma_{R,l}^2\rho_{D,li}^{-1}+1)\rho_{S,lk}^{-1}}\right)\right). \tag{C1}$$

According to the expression given above, (i) for the fixed channel estimate accuracy case (i.e., p_P is fixed), we find that only $c_1 = c_2 = 1$ leads the end-to-end achievable rate to fixed value; (ii) for the case that the transmit power of pilot sequences and data transmission are equal, we also find that only $c_1 = c_2 = 1$ leads the end-to-end achievable rate to a fixed value.

Following with the same lines of reasoning, the results for the MRC/MF processing case can be easily obtained.