Xing *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:18
DOI 10.1186/s13638-017-0805-7

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Self-interference suppression with imperfect channel estimation in a shared-antenna full-duplex massive MU-MIMO system

Pengbo Xing, Ju Liu*, Chao Zhai and Zhiyuan Yu

Abstract

In this paper, we consider a shared-antenna full-duplex massive MU-MIMO system and prove that the self-interference (SI) at the base station (BS) can be suppressed, though the direct-path SI exists and the signal/SI channel is imperfectly estimated. The BS is assumed to employ zero-forcing (ZF) or maximal-ratio transmission/maximal-ratio combining (MRT/MRC) method to linearly process signals. We propose a precoded SI channel training scheme by employing orthogonal sequences and downlink precoding, so the SI channel can be estimated like the signal channel with a much lower dimension. Based on the channel estimation, we further analyze the SI suppression by combining the SI removal and the large-scale antenna linear processing (LALP) method. Numerical results show that an additional 36 dB SI can be suppressed by combining the SI removal and the LALP suppression. We derive the tight closed-form lower bounds of the achievable rates for the combined suppression. Finally, we maximize the system spectral efficiency and energy efficiency based on the simulation and closed-form approximations.

Keywords: In-band full-duplex, Massive MIMO, Self-interference, MU-MIMO, Shared-antenna

1 Introduction

Massive multiple-input multiple-output (MIMO) and in-band full-duplex (IBFD) techniques have attracted intensive research interest because they can be used for broadband green communications [1–8]. IBFD can realize simultaneous bi-directional (uplink and downlink) data transmissions using the same time-frequency resource. Massive MIMO system exploits a large number of antennas at the transmitter to focus the energy into a narrow beam to improve the signal strength and mitigate interference. Both techniques are promising to meet the ever-growing requirements of wireless data transmissions, but bring some design variations over the physical or higher layers [9–17].

The massive MIMO technique can significantly improve the spectral efficiency (SE) and energy efficiency (EE) while keeping a lower transmit power for both uplink and downlink [15, 18–20] by mitigating the detrimental effects of small-scale fading, noise, and interference [4]. With more antennas equipped, the signal processing becomes much easier than the traditional MIMO system [4, 19]. The massive MIMO system usually operates with time-division duplexing (TDD) rather than frequency-division duplexing (FDD). In the FDD system, channel state information (CSI) should be acquired by the feedback, and much overhead will be introduced with the increase of antenna-array scale [15], thus the antenna grouping and CSI compression techniques were studied to reduce the estimation overhead [16, 17]. While in the TDD system, channel reciprocity can be considered to reduce the overhead, as the downlink CSI can be acquired by the base station (BS) through the uplink training [18, 21].

The full-duplex (FD) technique can potentially double the spectral efficiency, but it faces the problem of self-interference (SI), which refers to transmitted signals that can be directly or indirectly received by its own receivers. Recently, much effort has been made to suppress SI

*Correspondence: juliu@sdu.edu.cn
School of Information Science and Engineering, Shandong University, 27 Shanda Nanlu, 250100 Jinan, China

[22–25]. Riihonen et al. designed natural isolation, time-domain cancellation, and spatial SI suppression schemes [22]. Everett et al. proposed to exploit the antenna directional isolation and cross-polarization to suppress the SI in the wireless propagation domain [23]. Bharadia et al. proposed analog circuit domain cancellation methods that can keep a copy of the transmitted signal and subtract it from the received signal [24]. An analog-and-digital hybrid SI cancellation method has been developed by Duarte et al. and can provide approximately 85 dB interference suppression [25]. Recently, Ngo et al. proposed the loop interference cancellation technique in a separate-antenna relay system, where the SI can be suppressed with a large-scale antenna linear processing (LALP) method in a random reflected-path SI environment [26].

Sabharwal et al. discussed two methods of interfacing antennas to an IBFD wireless system: (i) separate-antenna FD and (ii) shared-antenna FD [1]. In the shared-antenna model, a bidirection circulator is used to transmit and receive signals without modifying many architecture standards, thus the RF hardwares such as antenna, RF cable and filters can be saved [2]. At the same time, the nonlinearities of the RF hardware, for example the circulators and amplifiers, introduce weak nonlinear distortions in such systems.

In this paper, we consider a shared-antenna FD massive multiuser MIMO (MU-MIMO) system. In our system, a BS is equipped with a single antenna array, and it serves multiple users, each with a single antenna. Like [2], we ignore the nonlinearities introduced by the RF hardware, and focus on the linear processing in this model. A linear processing method, such as zero-forcing (ZF) or maximal-ratio transmission/maximal-ratio combining (MRT/MRC), is adopted for the signal precoding and decoding over the FD links. Unlike [26], we discuss a Rician SI channel rather than Rayleigh channel model, the uplink and downlink channels in our model are completely dependent on each other, because the same antenna array is employed for the signal receive and transmission. The perfect channel reciprocity can be used to reduce the estimation overhead. Unlike [27], we consider the channel estimation for both the signal and SI channels in this paper. A precode SI channel training scheme is adopted to estimate the SI channel using the minimum mean-square-error (MMSE) technique. The main contributions of our work are summarized as:

- We consider both the non-random direct-path and the random reflected-path SI channels in this model. We prove that the Rician SI channel changes into Rayleigh channel when the downlink channel/SI channel is precoded. We also prove that the SI at BS with imperfect channel estimation can be asymptotically suppressed, though the uplink-downlink channels are totally dependent and the line of sight (LOS) SI exists. It is shown that the capability of LALP suppression is proportional to the square root of the antenna array scale M, which is different from the separate-antenna massive MIMO system. Because of the channel reciprocity, the signal channel estimation of the shared-antenna array is reduced by half compared with the separate-antenna arrays.
- We estimate the SI through the precoded SI channel with MMSE method, which is different from the traditional SI acquisition method. The estimation overhead and the dimension of pilots can be greatly suppressed compared with the direct estimation of the unprecoded SI channel. With the estimated SI, we can further analyze the SI suppression by combining the SI removal and the LALP method. It is shown that an additional 36 dB suppression capability can be achieved through combining the two methods. We also derive the closed-form lower bounds of the uplink achievable rates with imperfect CSI. The optimal SE and EE are further studied based on the rate lower bounds.

The rest of this paper is organized as follows. The system model is described in Section 2. Section 3 studies the SI suppression and the uplink achievable rates. Section 4 analyzes the SE and EE. Numerical results and conclusions are presented in Section 5 and Section 6, respectively.

Notations: The boldface upper and lower-case letters represent matrices and vectors, respectively. The superscripts $[\cdot]^*$, $[\cdot]^T$, and $[\cdot]^H$ stand for conjugate, transpose, and Hermitian transpose of matrices or vectors, respectively. The mathematical expectation is denoted as $\mathbb{E}\{\cdot\}$. $\|\cdot\|$ and represents the *Euclidean* norm. The (i,j)th entry of a matrix is denoted as $[\cdot]_{ij}$. The trace of a square matrix is denoted as $\mathrm{tr}(\cdot)$. The $\xrightarrow{a.s.}$ represents *almost sure convergence*. $u_n = o(v_n)$ and $u_n = O(v_n)$ denote that there exists a constant C, such that $u_n = Cv_n$ and $u_n \leq Cv_n$, $\forall n$, respectively.

2 System model

As shown in Fig. 1, we consider the uplink of a shared-antenna FD massive MU-MIMO system, where a BS is equipped with M antennas and serves K users, each with a single antenna. Like the separate-antenna full-duplex system [23, 25], physical isolation can be adopted between the BS antennas to avoid front-end saturation at the receiver, and active analog SI cancellation is used to suppress the SI before the baseband signal processing. We assume that M is much larger than K, i.e., $M \gg K$. Both the BS and the users are working in the FD mode, i.e., they can transmit and receive signals using the same time-

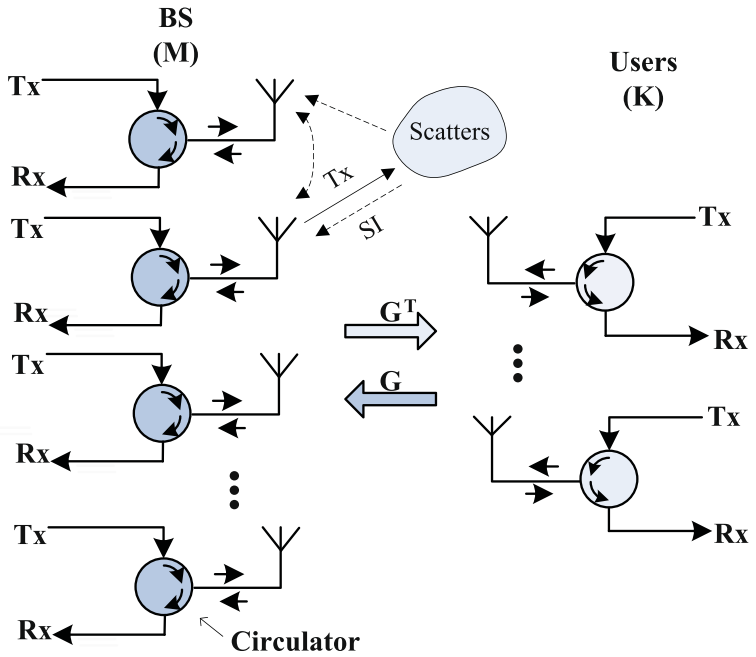

Fig. 1 Data transmission and reception with SI at the BS in the shared-antenna FD massive MU-MIMO system

frequency resource. Therefore, the $M \times 1$ uplink received signal at the BS is given as

$$\boldsymbol{y}_{\mathrm{BS}} = \sqrt{p_{\mathrm{u}}}\boldsymbol{G}\boldsymbol{x}_{\mathrm{u}} + \sqrt{p_{\mathrm{d}}}\boldsymbol{G}_{\mathrm{s}}\boldsymbol{s} + \boldsymbol{n}, \tag{1}$$

where the first term represents the desired signal, and the second term represents the SI from the downlink transmission. $\boldsymbol{G}$ is the $M \times K$ channel matrix between the K users and the BS; $\boldsymbol{x}_{\mathrm{u}}$ denotes the symbols transmitted by the K users with unit power $\mathbb{E}\{\boldsymbol{x}_{\mathrm{u}}\boldsymbol{x}_{\mathrm{u}}^H\} = \boldsymbol{I}_K$, p_{u} is the transmit power of each user; G_{s} is the $M \times M$ SI channel matrix between the BS transceivers; $\boldsymbol{s}$ is the downlink precoded signal transmitted to the K users with unit power $\mathbb{E}\{\boldsymbol{s}^H\boldsymbol{s}\} = 1$; p_{d} is the downlink transmit power; and $\boldsymbol{n}$ represents the additive noise with zero mean and unit variance.

2.1 Channel model

The channel matrix G models the small-scale Rayleigh fading and the large-scale path-loss, so we have

$$\boldsymbol{G} = \boldsymbol{H}\boldsymbol{D}^{1/2}, \tag{2}$$

where $\boldsymbol{H}$ is the $M \times K$ matrix with i.i.d entries [4], representing the small-scale fading between the K users and the BS, and the entry h_{mk} is a circular symmetric complex Gaussian (CSCG) random variable with zero mean and unit variance, i.e., $h_{mk} \sim \mathcal{CN}(0,1)$; $\boldsymbol{D}$ is a $K \times K$ diagonal matrix of the large-scale path-loss with $[\boldsymbol{D}]_{kk} = \beta_k$.

Since the physical isolation and active analog SI cancellation are performed preliminarily, it is reasonable to assume that the SI can be suppressed greatly [26, 28]. The magnitude of the SI channel can be modeled as Rician distribution with LOS SI [29]. Therefore, we can write the matrix $\boldsymbol{G}_{\mathrm{s}}$ as the direct-path plus the reflected-path coefficients, i.e.,

$$\boldsymbol{G}_{\mathrm{s}} = \bar{\boldsymbol{G}}_{\mathrm{s}} + \tilde{\boldsymbol{G}}_{\mathrm{s}}, \tag{3}$$

where $\bar{\boldsymbol{G}}_{\mathrm{s}}$ is a complex deterministic matrix with $[\bar{\boldsymbol{G}}_{\mathrm{s}}]_{mj} = c_{mj}$, which represents the antenna intra-leakage and inter-coupling coefficients related to the direct path, and $\tilde{\boldsymbol{G}}_{\mathrm{s}}$ is a complex random matrix that represents the reflection coefficients related to the reflected path. Similar to channel matrix $\boldsymbol{G}$, random matrix $\tilde{\boldsymbol{G}}_{\mathrm{s}}$ can be expressed as

$$\tilde{\boldsymbol{G}}_{\mathrm{s}} = \tilde{\boldsymbol{H}}_{\mathrm{s}}\tilde{\boldsymbol{D}}_{\mathrm{s}}^{1/2}, \tag{4}$$

where $\tilde{\boldsymbol{H}}_{\mathrm{s}}$ is an $M \times M$ matrix with i.i.d. entries representing the small-scale fading. We suppose $[\tilde{\boldsymbol{H}}_{\mathrm{s}}]_{mj} = h_{\mathrm{s},mj}$, where $h_{\mathrm{s},mj}$ is a CSCG random variable with zero mean and unit variance, and $\tilde{\boldsymbol{D}}_{\mathrm{s}}$ is an $M \times M$ diagonal matrix of the large-scale path-loss, where $[\tilde{\boldsymbol{D}}_{\mathrm{s}}]_{mm} = \beta$, namely, all of the SI large-scale path-loss coefficients are assumed to be identical. This assumption is reasonable, as the antenna array size is much smaller than the distance to the scatters. If these coefficients are not identical, we can set all the elements of SI channel matrix $\tilde{\boldsymbol{G}}_{\mathrm{s}}$ to be the maximum value $\beta_{\max}$, which represents the worst case, then $\tilde{\boldsymbol{G}}_s$ becomes a matrix with identical large-scale fading.

For the shared-antenna FD massive MU-MIMO system, if the array scale is large enough and greatly exceeds the number of users, that is $M \gg K$, the column vectors of the

signal channel matrix are mutually orthogonal [4], and so do the vectors of the SI channel matrix. We then have

$$\frac{\boldsymbol{G}^H\boldsymbol{G}}{M} = \boldsymbol{D}^{1/2}\frac{\boldsymbol{H}^H\boldsymbol{H}}{M}\boldsymbol{D}^{1/2} \approx \boldsymbol{D}, \tag{5}$$

$$\frac{\boldsymbol{G}_s^H\boldsymbol{G}_s}{M} \approx \frac{\bar{\boldsymbol{G}}_s^H\bar{\boldsymbol{G}}_s}{M} + \frac{\tilde{\boldsymbol{G}}_s^H\tilde{\boldsymbol{G}}_s}{M} \approx \bar{\boldsymbol{D}}_s + \tilde{\boldsymbol{D}}_s, \tag{6}$$

$$\mathrm{tr}(\boldsymbol{G}_s^H\boldsymbol{G}_s) \approx \mathrm{tr}(\bar{\boldsymbol{G}}_s^H\bar{\boldsymbol{G}}_s) + \mathrm{tr}(\tilde{\boldsymbol{G}}_s^H\tilde{\boldsymbol{G}}_s) = O(M^2), \tag{7}$$

where $\bar{\boldsymbol{D}}_s$ is an $M \times M$ deterministic matrix and $\tilde{\boldsymbol{D}}_s$ is an $M \times M$ diagonal matrix with $[\tilde{\boldsymbol{D}}_s]_{mm} = \beta$. The proofs of (6) and (7) are given in Appendix.

2.2 Channel estimation

Channel estimation is a prerequisite for SI suppression. Commonly, a series of orthogonal training sequences is sent within a period 2τ, which is a small part of the channel coherence time T. The channel training/estimation is performed in half-duplex. Therefore, the $M \times \tau$ received pilot matrix for the uplink channel and SI channel at the BS can be written as

$$\boldsymbol{Y}_\mathrm{u} = \sqrt{P_\mathrm{u}}\boldsymbol{G}\boldsymbol{\Phi}_\mathrm{u} + \boldsymbol{N}, \tag{8}$$

and

$$\boldsymbol{Y}_\mathrm{s} = \sqrt{P_\mathrm{d}}\boldsymbol{G}_\mathrm{s}\boldsymbol{A}\boldsymbol{\Phi}_\mathrm{d} + \boldsymbol{N}_\mathrm{s}, \tag{9}$$

where P_u and P_d are the transmit powers of pilot sequences for the uplink and downlink, respectively. Suppose that the power of each pilot symbol equals the average power of data transmission, then we have $P_\mathrm{u} = \tau p_\mathrm{u}$ and $P_\mathrm{d} = \tau p_\mathrm{d}$; $\boldsymbol{\Phi}_\mathrm{u} = \boldsymbol{\Phi}_\mathrm{d}$ are $K \times \tau$ matrices whose rows are the pilot sequences for the estimation of the signal channel and SI channel. The pilot matrices satisfy $\boldsymbol{\Phi}_\mathrm{u}\boldsymbol{\Phi}_\mathrm{u}^H = \boldsymbol{\Phi}_\mathrm{d}\boldsymbol{\Phi}_\mathrm{d}^H = \boldsymbol{I}_K$ and $\tau \geq K$; $\boldsymbol{A}$ is an $M \times K$ precoding matrix, which is used for downlink beamforming and updated from the uplink channel estimation; $\boldsymbol{N}$ and $\boldsymbol{N}_\mathrm{s}$ are $M \times \tau$ AWGN matrices with i.i.d $\mathcal{CN}(0,1)$ entries.

2.2.1 Signal channel estimation

The signal channel is estimated in the first training period τ. Given $\boldsymbol{Y}_\mathrm{u}$, the MMSE estimate of $\boldsymbol{G}$ can be written as [26]

$$\hat{\boldsymbol{G}} = \frac{1}{\sqrt{P_\mathrm{u}}}\boldsymbol{Y}_\mathrm{u}\boldsymbol{\Phi}_\mathrm{u}^H\check{\boldsymbol{D}} = \left(\boldsymbol{G} + \frac{1}{\sqrt{\tau p_\mathrm{u}}}\boldsymbol{N}_\mathrm{u}\right)\check{\boldsymbol{D}}, \tag{10}$$

where $\boldsymbol{N}_\mathrm{u} \triangleq \boldsymbol{N}\Phi_\mathrm{u}^H$ and $\check{\boldsymbol{D}} \triangleq (\frac{1}{\tau p_\mathrm{u}}\boldsymbol{D}^{-1} + \boldsymbol{I}_K)^{-1}$. Because $\boldsymbol{\Phi}_\mathrm{u}\boldsymbol{\Phi}_\mathrm{u}^H = \boldsymbol{I}_K$, the entries of $\boldsymbol{N}_\mathrm{u}$ are i.i.d random variables distributed as $\mathcal{CN}(0,1)$. Let $\boldsymbol{\Delta}$ denote the estimation error of $\boldsymbol{G}$, then we have

$$\boldsymbol{G} = \hat{\boldsymbol{G}} + \boldsymbol{\Delta}. \tag{11}$$

According to the property of *statistician's Pythagorean theorem* of the MMSE estimation [30], the estimation error $\boldsymbol{\Delta}$ is independent of $\hat{\boldsymbol{G}}$. Therefore, the kth columns of $\hat{\boldsymbol{G}}$ and $\boldsymbol{\Delta}$ are mutually independent, and the elements of the kth vector are distributed as $\mathcal{CN}(0, \hat{\beta}_k)$ and $\mathcal{CN}(0, \tilde{\beta}_k)$, respectively, with $\hat{\beta}_k \triangleq \frac{\tau p_\mathrm{u}\beta_k^2}{\tau p_\mathrm{u}\beta_k+1}$ and $\tilde{\beta}_k \triangleq \frac{\beta_k}{\tau p_\mathrm{u}\beta_k+1}$.

Note that the shared-antenna array massive MIMO system has a single array, so the transmit and receive CSI are the same during the coherent time. We only need to estimate the uplink channel, and thus the estimation overhead is reduced by half. This is different from the separate-antenna arrays system, where the CSI of the receive and transmit arrays are different, the uplink and downlink CSI have to be estimated separately.

2.2.2 SI channel estimation

The SI channel is estimated in the second training period τ. In fact, it is not necessary to estimate $\boldsymbol{G}_\mathrm{s}$, we can estimate the equivalent or precoded SI channel $\boldsymbol{G}_\mathrm{sa}$ instead, which represents the product of $\boldsymbol{G}_\mathrm{s}$ and $\boldsymbol{A}$. The new precoding matrix $\boldsymbol{A}$ can be updated for the SI channel estimation after $\hat{\boldsymbol{G}}$ is acquired. Because $\boldsymbol{G}_\mathrm{sa}$ has the same dimension as the uplink channel $\boldsymbol{G}$, the training pilot length and pilot amount depend on the user number K rather than the number of BS antennas M. Therefore, the dimension of the SI channel estimation can be decreased from $M \times M$ to $M \times K$, and thus fewer orthogonal pilot sequences are needed. Note that the proposed training and estimation method is different from the traditional SI acquisition method where the SI is directly estimated with/without pilots [22].

The precoding matrix $\boldsymbol{A}$ is derived from $\hat{\boldsymbol{G}}$. For a linear precoding scheme such as ZF or MRT, $\boldsymbol{A}$ is given as

$$\boldsymbol{A} = \begin{cases} \alpha_\mathrm{ZF}\hat{\boldsymbol{G}}^*(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}^*)^{-1}, & \text{for ZF,} \\ \alpha_\mathrm{MRT}\hat{\boldsymbol{G}}^*, & \text{for MRT,} \end{cases} \tag{12}$$

where $\alpha_\mathrm{ZF} \triangleq \sqrt{\frac{M-K}{\sum_{k=1}^K \hat{\beta}_k^{-1}}}$ and $\alpha_\mathrm{MRT} \triangleq \sqrt{\frac{1}{M\sum_{k=1}^K \hat{\beta}_k}}$ are the normalization factors [26] satisfying $\mathbb{E}\{\boldsymbol{s}^H\boldsymbol{s}\} = 1$.

Proposition 1 *In the shared-antenna FD massive MU-MIMO system, the SI equivalent channel $\boldsymbol{G}_{sa} = \boldsymbol{G}_\mathrm{s}\boldsymbol{A}$ is an $M \times K$ complex Gaussian matrix if K is fixed and $M \to \infty$. The (m,i)th entry follows the distribution $\mathcal{CN}\left(0, \rho_{m,i}\right)$, where $\rho_{m,i} = \frac{1}{M^2}\alpha_{ZF}^2\hat{\beta}_i^{-1}\sum_{j=1}^M\left(\left|c_{mj}\right|^2+\beta\right)$ for ZF precoding and $\rho_{m,i} = \alpha_{MRT}^2\hat{\beta}_i\sum_{j=1}^M\left(\left|c_{mj}\right|^2+\beta\right)$ for MRT precoding, $\hat{\beta}_i \triangleq \frac{\tau p_\mathrm{u}\beta_i^2}{\tau p_\mathrm{u}\beta_i+1}$, $i = 1,\ldots,K$.*

Proof. Please see Appendix. □

From Proposition 1, we can see that the equivalent SI channel is a zero-mean complex Gaussian channel, i.e., the Rayleigh channel, even though the SI channel is a Rician channel. Therefore, the precoded SI channel can still be

treated as Rayleigh channel. Like the signal channel $\boldsymbol{G}$, $\boldsymbol{G}_{\mathrm{sa}}$ can be estimated by MMSE if $\boldsymbol{Y}_{\mathrm{s}}$ is given. Let $\boldsymbol{\mathcal{E}}$ denote the estimation error matrix of $\boldsymbol{G}_{\mathrm{sa}}$, we then have

$$\boldsymbol{G}_{\mathrm{sa}} = \hat{\boldsymbol{G}}_{\mathrm{sa}} + \boldsymbol{\mathcal{E}}, \tag{13}$$

the (m,i)th entry of $\hat{\boldsymbol{G}}_{\mathrm{sa}}$ and $\boldsymbol{\mathcal{E}}$ are distributed as $\mathcal{CN}(0,\hat{\rho}_{m,i})$ and $\mathcal{CN}(0,\tilde{\rho}_{m,i})$, respectively, with $\hat{\rho}_{m,i} \triangleq \frac{\tau p_{\mathrm{d}} \rho_{m,i}^2}{\tau p_{\mathrm{d}} \rho_{m,i}+1}$, $\tilde{\rho}_{m,i} \triangleq \frac{\rho_{m,i}}{\tau p_{\mathrm{d}} \rho_{m,i}+1}$, $m = 1,...,M$, $i = 1,...,K$.

2.3 Uplink decoding under precoded SI

Precoding can be used in the training phase for the SI channel estimation, and it can also be used in the data transmission phase for beamforming, which can help enhance the strength of the desired signal and reduce the interference to others. During the data transmission period, the ZF or MRT scheme is adopted by the BS for the downlink beamforming. Correspondingly, the BS receives the precoded SI from the transmitters, and it uses the ZF or MRC linear processing method for the uplink decoding.

Suppose $\boldsymbol{x}_{\mathrm{d}}$ is a $K \times 1$ data vector transmitted from the BS to the users with unit power $\mathbb{E}\{\boldsymbol{x}_{\mathrm{d}}\boldsymbol{x}_{\mathrm{d}}^H\} = \boldsymbol{I}_K$, and $\boldsymbol{s} = \boldsymbol{A}\boldsymbol{x}_{\mathrm{d}}$. (1) can be rewritten as

$$\boldsymbol{y}_{\mathrm{BS}} = \sqrt{p_{\mathrm{u}}}\boldsymbol{G}\boldsymbol{x}_{\mathrm{u}} + \sqrt{p_{\mathrm{d}}}\boldsymbol{G}_{\mathrm{s}}\boldsymbol{A}\boldsymbol{x}_{\mathrm{d}} + \boldsymbol{n}. \tag{14}$$

To decode the uplink data of K users, the BS receiver will multiply a linear matrix $\boldsymbol{W}^T$ before $\boldsymbol{y}_{\mathrm{BS}}$, which is a function of the signal channel estimation, i.e.,

$$\boldsymbol{W}^T = \begin{cases} \left(\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}}\right)^{-1}\hat{\boldsymbol{G}}^H, & \text{for ZF;} \\ \frac{1}{M}\hat{\boldsymbol{G}}^H, & \text{for MRC.} \end{cases} \tag{15}$$

With the uplink linear receiver at the BS, the user data can be decoded as

$$\boldsymbol{r} = \boldsymbol{W}^T\boldsymbol{y}_{\mathrm{BS}} = \sqrt{p_{\mathrm{u}}}\boldsymbol{W}^T\boldsymbol{G}\boldsymbol{x}_{\mathrm{u}} + \sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\boldsymbol{G}_{\mathrm{s}}\boldsymbol{A}\boldsymbol{x}_{\mathrm{d}} + \boldsymbol{W}^T\boldsymbol{n}, \tag{16}$$

where $\sqrt{p_{\mathrm{u}}}\boldsymbol{W}^T\boldsymbol{G}\boldsymbol{x}_{\mathrm{u}}$ is the desired signal, $\sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\boldsymbol{G}_{\mathrm{s}}\boldsymbol{A}\boldsymbol{x}_{\mathrm{d}}$ is the SI between the BS transceivers, and $\boldsymbol{W}^T\boldsymbol{n}$ is the additive noise.

Further, from (11) and (13), we have

$$\begin{aligned} \boldsymbol{r} =& \sqrt{p_{\mathrm{u}}}\boldsymbol{W}^T\hat{\boldsymbol{G}}\boldsymbol{x}_{\mathrm{u}} + \sqrt{p_{\mathrm{u}}}\boldsymbol{W}^T\boldsymbol{\Delta}\boldsymbol{x}_{\mathrm{u}} \\ &+ \sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\hat{\boldsymbol{G}}_{\mathrm{sa}}\boldsymbol{x}_{\mathrm{d}} + \sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\boldsymbol{\mathcal{E}}\boldsymbol{x}_{\mathrm{d}} + \boldsymbol{W}^T\boldsymbol{n}, \end{aligned} \tag{17}$$

where $\sqrt{p_{\mathrm{u}}}\boldsymbol{W}^T\hat{\boldsymbol{G}}\boldsymbol{x}_{\mathrm{u}}$ is the estimated signal, $\sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\hat{\boldsymbol{G}}_{\mathrm{sa}}\boldsymbol{x}_{\mathrm{d}}$ is the SI from the estimation, $\sqrt{p_{\mathrm{u}}}\boldsymbol{W}^T\boldsymbol{\Delta}\boldsymbol{x}_{\mathrm{u}}$ and $\sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\boldsymbol{\mathcal{E}}\boldsymbol{x}_{\mathrm{d}}$ are the estimation errors of the desired signal and SI, respectively.

3 SI suppression and achievable rate

In the massive MIMO system, the white noise and the inter-user interference can be eliminated by an LALP method [15, 19, 31]. In this section, we consider the shared-antenna FD massive MU-MIMO system and discuss the combined SI suppression based on the estimated channels of signal and SI. We also derive the closed-form lower bounds of the uplink achievable rates, which can be used as a system performance metric.

3.1 SI suppression with LALP

In the TDD half-duplex (HD) system, the interferences are asymptotically orthogonal to the subspace spanned by the desired signals when the antenna-array scale M approaches infinity. The desired signal can be recovered by projecting the received signal in a desired subspace with linear processing [19, 26]. In the shared-antenna FD massive MU-MIMO system, the asymptotic orthogonality property still holds true. Unlike the separate-antenna array system [26], the SI in our model includes the non-random direct-path interference, and the uplink channel is totally dependent on the downlink channel, as the same antenna array is employed to transmit and receive signals simultaneously.

Theorem 1 *In the shared-antenna FD massive MU-MIMO system, when the signal and SI channels are not perfectly estimated and ZF or MRT/MRC linear processing is adopted, the SI at the BS will vanish if K is fixed and* $M \to \infty$.

Proof. We consider the SI term $\sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\boldsymbol{G}_{\mathrm{s}}\boldsymbol{A}\boldsymbol{x}_{\mathrm{d}}$ of (16). From (5) and (7), we have $\frac{\boldsymbol{G}^H\boldsymbol{G}}{M} = \boldsymbol{D}$ and $\mathrm{tr}\left(\boldsymbol{G}_s\boldsymbol{G}_s^H\right) = O(M^2)$ when $M \to \infty$. According to the properties of MMSE estimation, we know that the estimated channels $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{G}}_{\mathrm{sa}}$ are independent on the estimation errors $\boldsymbol{\Delta}$ and $\boldsymbol{\mathcal{E}}$, respectively, and they are zero-mean complex Gaussian matrices as well. Substituting (12) and (15) into $\boldsymbol{W}^T\boldsymbol{G}_{\mathrm{s}}\boldsymbol{A}$, we have

$$\boldsymbol{W}^T\boldsymbol{G}_{\mathrm{s}}\boldsymbol{A} = \begin{cases} \alpha_{\mathrm{ZF}}\left(\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}}\right)^{-1}\hat{\boldsymbol{G}}^H\boldsymbol{G}_{\mathrm{s}}\hat{\boldsymbol{G}}^*\left(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}^*\right)^{-1}, & \text{for ZF;} \\ \alpha_{\mathrm{MRT}}\frac{1}{M}\hat{\boldsymbol{G}}^H\boldsymbol{G}_{\mathrm{s}}\hat{\boldsymbol{G}}, & \text{for MRT/MRC.} \end{cases} \tag{18}$$

Note that $\mathrm{tr}(\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}}) = \mathrm{tr}(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}^*) = o(M)$, $\alpha_{\mathrm{ZF}} = o(\sqrt{M})$ and $\alpha_{\mathrm{MRT}} = o(\frac{1}{\sqrt{M}})$, by using Lemma 13 and 14 in [32], we obtain

$$\frac{1}{M}\hat{\boldsymbol{g}}_i^H\boldsymbol{G}_{\mathrm{s}}\hat{\boldsymbol{g}}_i - \frac{1}{M}\mathrm{tr}\boldsymbol{G}_{\mathrm{s}} \xrightarrow{a.s.} 0, \quad i = 1,\ldots,K, \tag{19}$$

$$\frac{1}{M}\hat{\boldsymbol{g}}_i^H\boldsymbol{G}_{\mathrm{s}}\hat{\boldsymbol{g}}_j \xrightarrow{a.s.} 0, \quad i,j = 1,\ldots,K, \quad i \neq j, \tag{20}$$

and

$$\frac{1}{M^{2/3}}\hat{\boldsymbol{G}}^H\boldsymbol{G}_{\mathrm{s}}\hat{\boldsymbol{G}} \xrightarrow{a.s.} \boldsymbol{0}_{K\times K}, \tag{21}$$

where $\hat{\boldsymbol{g}}_i$ and $\hat{\boldsymbol{g}}_j$ are the ith and jth column of $\boldsymbol{G}$, respectively. Thus, we have the SI term of (17)

$$\sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\hat{\boldsymbol{G}}_{\mathrm{sa}}\boldsymbol{x}_{\mathrm{d}} \xrightarrow{a.s.} \boldsymbol{0}_{K\times 1}, \tag{22}$$

and

$$\sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\boldsymbol{\mathcal{E}}\boldsymbol{x}_{\mathrm{d}} \xrightarrow{a.s.} \boldsymbol{0}_{K\times 1}, \tag{23}$$

for ZF and MRT/MRC linear processing schemes. □

Remark: In the shared-antenna FD massive MU-MIMO system, the decoding matrix $\boldsymbol{W}^T$ and the precoding matrix $\boldsymbol{A}$ in the SI term are totally dependent because of the channel reciprocity between uplink and downlink. This is different from the FD relay massive MIMO system where the source-relay and relay-destination channels are independent. According to Lemma 13 in [32] and Theorem 3.7 in [33], the LALP suppression capability is $O(1/M^{1/2})$, and it is different from [26] where the LALP suppression capability is $O(1/M)$ for the separate-antenna arrays.

Proposition 2 *In the shared-antenna FD massive MU-MIMO system with imperfect channel estimation, if K is fixed and $M \to \infty$, the decoded uplink signal vector at the BS is*

$$\boldsymbol{r} \xrightarrow{a.s.} \sqrt{p_{\mathrm{u}}}\boldsymbol{x}_{\mathrm{u}}, \tag{24}$$

for the ZF precoding and decoding and

$$\hat{\boldsymbol{D}}^{-1}\boldsymbol{r} \xrightarrow{a.s.} \sqrt{p_{\mathrm{u}}}\boldsymbol{x}_{\mathrm{u}}, \tag{25}$$

for the MRT/MRC precoding and decoding, where $[\hat{\boldsymbol{D}}]_{kk} = \hat{\beta}_k$ *and* $\hat{\beta}_k \triangleq \frac{\tau p_{\mathrm{u}}\beta_k^2}{\tau p_{\mathrm{u}}\beta_k+1}$, $k = 1,\ldots,K$.

Proof. According to (5), (11), (16), and Theorem 1, the results can be obtained. □

3.2 Combined SI suppression

Practically, the number of antennas cannot be infinite, the LALP method cannot ideally suppress the SI, especially for the dependent shared-antenna array. As mentioned above, the SI in the FD massive MU-MIMO system can be estimated by the MMSE through a precoded SI channel training process, thus the known SI can be removed before the LALP suppression. After the SI removal, the remaining interference becomes much weaker, and the LALP can be further adopted to suppress the remaining SI. Therefore, the suppression capability is greatly improved by combining the two suppression methods.

As shown in (17), the term $\sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\hat{\boldsymbol{G}}_{\mathrm{sa}}\boldsymbol{x}_{\mathrm{d}}$ is known by the BS. We can subtract it from the received signal $\boldsymbol{r}$, so that (17) can be rewritten as

$$\hat{\boldsymbol{r}} = \sqrt{p_{\mathrm{u}}}\boldsymbol{W}^T\hat{\boldsymbol{G}}\boldsymbol{x}_{\mathrm{u}} + \sqrt{p_{\mathrm{u}}}\boldsymbol{W}^T\boldsymbol{\Delta}\boldsymbol{x}_{\mathrm{u}} + \sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\boldsymbol{\mathcal{E}}\boldsymbol{x}_{\mathrm{d}} + \boldsymbol{W}^T\boldsymbol{n}, \tag{26}$$

where $\hat{\boldsymbol{r}} = \boldsymbol{r} - \sqrt{p_{\mathrm{d}}}\boldsymbol{W}^T\hat{\boldsymbol{G}}_{\mathrm{sa}}\boldsymbol{x}_{\mathrm{d}}$. The kth decoded signal can then be expressed as

$$\begin{aligned}\hat{r}_k =& \sqrt{p_{\mathrm{u}}}\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_k x_k + \sqrt{p_{\mathrm{u}}}\sum_{i=1,i\neq k}^{K}\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_i x_i \\ &+ \sqrt{p_{\mathrm{u}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{\delta}_i x_i + \sqrt{p_{\mathrm{d}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{\epsilon}_i x_i^{\mathrm{d}} + \boldsymbol{w}_k^T\boldsymbol{n},\end{aligned} \tag{27}$$

where $\boldsymbol{w}_k$, $\hat{\boldsymbol{g}}_k$, $\boldsymbol{\delta}_i$, and $\boldsymbol{\epsilon}_i$ are the kth or ith column of $\boldsymbol{W}$, $\hat{\boldsymbol{G}}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\mathcal{E}}$, respectively; x_k and x_i^{d} are the kth and ith element of $\boldsymbol{x}_{\mathrm{u}}$ and $\boldsymbol{x}_{\mathrm{d}}$, respectively; $\sqrt{p_{\mathrm{u}}}\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_k x_k$ is the desired signal from the kth user; $\sqrt{p_{\mathrm{u}}}\sum_{i=1,i\neq k}^{K}\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_i x_i$ is the inter-user interference; $\sqrt{p_{\mathrm{u}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{\delta}_i x_i$ is the estimation error from the signal channel $\boldsymbol{G}$; $\sqrt{p_{\mathrm{d}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{\epsilon}_i x_i^{\mathrm{d}}$ is the estimation error from the SI equivalent channel $\boldsymbol{G}_{\mathrm{sa}}$ and $\boldsymbol{w}_k^T\boldsymbol{n}$ is the additive noise.

Theoretically, the SI and other interferences and noise can be suppressed as much as possible, and the achievable rate of users can approach infinity with the increase of the antenna array scale. Practically, tens or hundreds of antennas can be a good approximation of infinity [4, 34].

3.3 Achievable rate

In this subsection, we will discuss the uplink achievable rate of the shared-antenna FD massive MIMO system for the LALP and the combined suppression.

3.3.1 Combined method

From (27), the kth user uplink achievable rate is given as

$$R_k = \mathbb{E}\{\log_2(1+\gamma_k)\}, \tag{28}$$

where γ_k is given as

$$\gamma_k = \frac{p_{\mathrm{u}}\left|\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_k\right|^2}{p_{\mathrm{u}}\sum\limits_{i=1,i\neq k}^{K}\left|\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_i\right|^2 + p_{\mathrm{u}}\sum\limits_{i=1}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{\delta}_i\right|^2 + p_{\mathrm{d}}\sum\limits_{i=1}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{\epsilon}_i\right|^2 + \left\|\boldsymbol{w}_k^T\right\|^2}. \tag{29}$$

3.3.2 LALP only

If we estimate the signal channel and suppress the SI by the LALP method only, we have

$$\begin{aligned}r_k' =& \sqrt{p_{\mathrm{u}}}\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_k x_k + \sqrt{p_{\mathrm{u}}}\sum_{i=1,i\neq k}^{K}\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_i x_i \\ &+ \sqrt{p_{\mathrm{u}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{\delta}_i x_i + \sqrt{p_{\mathrm{d}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{g}_{\mathrm{sa},i}x_i^{\mathrm{d}} + \boldsymbol{w}_k^T\boldsymbol{n},\end{aligned} \tag{30}$$

where $\boldsymbol{g}_{\mathrm{sa},i}$ is the ith column of $\boldsymbol{G}_{\mathrm{sa}}$, $i = 1,...,K$; $\sqrt{p_{\mathrm{d}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{g}_{\mathrm{sa},i}x_i^{\mathrm{d}}$ is the SI. The kth user achievable rate with imperfect estimation and the LALP method only can be expressed as

$$R_k' = \mathbb{E}\{\log_2(1+\gamma_k')\}, \tag{31}$$

where γ_k' is given as

$$\gamma_k' = \frac{p_{\mathrm{u}}\left|\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_k\right|^2}{p_{\mathrm{u}}\sum\limits_{i=1,i\neq k}^{K}\left|\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_i\right|^2+p_{\mathrm{u}}\sum\limits_{i=1}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{\delta}_i\right|^2+p_{\mathrm{d}}\sum\limits_{i=1}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{g}_{\mathrm{sa},i}\right|^2+\left\|\boldsymbol{w}_k^T\right\|^2}. \tag{32}$$

3.3.3 Perfect channel estimation

We compare the achievable rate of the imperfect case with the perfect case and verify the accuracy of the channel estimation. If the perfect CSI can be obtained, the signal channel estimation error and the SI terms can be removed from (32). By setting the estimation error as zero in (27), the uplink decoded signal of the kth user with perfect estimation can be obtained as

$$r_k'' = \sqrt{p_{\mathrm{u}}}\boldsymbol{w}_k^T\boldsymbol{g}_k x_k + \sqrt{p_{\mathrm{u}}}\sum_{i=1,i\neq k}^{K}\boldsymbol{w}_k^T\boldsymbol{g}_i x_i + \boldsymbol{w}_k^T\boldsymbol{n}, \tag{33}$$

where $r_k'' = r_k - \sqrt{p_{\mathrm{d}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{g}_{\mathrm{sa},i}x_i^{\mathrm{d}}$, and $\boldsymbol{w}_k$, $\boldsymbol{g}_k$, and $\boldsymbol{g}_{\mathrm{sa},i}$ are the kth or ith column of $\boldsymbol{W}$, $\boldsymbol{G}$, and $\boldsymbol{G}_{\mathrm{sa}}$, respectively; $\sqrt{p_{\mathrm{u}}}\boldsymbol{w}_k^T\boldsymbol{g}_k x_k$ is the desired signal from the kth user; $\sqrt{p_{\mathrm{u}}}\sum_{i=1,i\neq k}^{K}\boldsymbol{w}_k^T\boldsymbol{g}_i x_i$ is the inter-user interference; and $\sqrt{p_{\mathrm{d}}}\sum_{i=1}^{K}\boldsymbol{w}_k^T\boldsymbol{g}_{\mathrm{sa},i}x_i^{\mathrm{d}}$ is the SI. Therefore, the kth user achievable rate with perfect CSI can be given as

$$R_k'' = \mathbb{E}\left\{\log_2\left(1+\gamma_k''\right)\right\}, \tag{34}$$

where γ_k'' is given as

$$\gamma_k'' = \frac{p_{\mathrm{u}}\left|\boldsymbol{w}_k^T\boldsymbol{g}_k\right|^2}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{g}_i\right|^2+\left\|\boldsymbol{w}_k^T\right\|^2}. \tag{35}$$

3.4 Achievable rate lower bounds of the combined method

From (28), we can derive the lower bounds of the achievable rates for the combined method with ZF and MRT/MRC linear processing.

Lemma 1 *In the shared-antenna FD massive MU-MIMO system, the random variable* $\tilde{g}_i = \frac{\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_i}{\left\|\boldsymbol{w}_k^T\right\|}$ *is a zero-mean complex Gaussian variable with variance* $\hat{\beta}_i$*, which is independent on* $\boldsymbol{w}_k^T$ *if* K *is fixed and* $M\to\infty$*, where* $\hat{\beta}_i = \frac{\tau p_{\mathrm{u}}\beta_i^2}{\tau p_{\mathrm{u}}\beta_i+1}$, $i = 1,...,K$, $i\neq k$.

Proof. Please see Appendix A of [19]. □

Lemma 2 *In the shared-antenna FD massive MU-MIMO system, the random variable* $\tilde{\delta}_i = \frac{\boldsymbol{w}_k^T\boldsymbol{\delta}_i}{\left\|\boldsymbol{w}_k^T\right\|}$ *is a zero-mean complex Gaussian variable with variance* $\tilde{\beta}_i$*, which is independent on* $\boldsymbol{w}_k^T$ *if* K *is fixed and* $M\to\infty$*, where* $\tilde{\beta}_i = \frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i+1}$, $i = 1,...,K$.

Proof. According to the optimum MMSE estimation property, $\boldsymbol{\delta}_i$ is independent of $\hat{\boldsymbol{g}}_i$ for $i = 1,...,K$. Similarly, as the proof of Lemma 1, we can obtain the result. □

Lemma 3 *In the shared-antenna FD massive MU-MIMO system, the random variable* $\tilde{\epsilon}_i = \frac{\boldsymbol{w}_k^T\boldsymbol{\epsilon}_i}{\left\|\boldsymbol{w}_k^T\right\|}$ *is a zero-mean complex Gaussian variable with variance* $\tilde{\rho}_i$*, which is independent on* $\boldsymbol{w}_k^T$ *if* K *is fixed and* $M\to\infty$*, where* $\tilde{\rho}_i = \frac{1}{M}\sum_{m=1}^{M}\frac{\rho_{m,i}}{\tau p_{\mathrm{d}}\rho_{m,i}+1}$, $i = 1,...,K$.

Proof. Please see Appendix. □

Proposition 3 *With the combined scheme of SI removal and ZF large-scale antenna linear processing, the kth uplink achievable rate in the shared-antenna FD massive MU-MIMO system is lower-bounded by*

$$\underline{R}_k^{\mathrm{Z}} = \log_2\left(1+\frac{p_{\mathrm{u}}(M-K)\hat{\beta}_k}{p_{\mathrm{u}}\sum_{i=1}^{K}\tilde{\beta}_i+p_{\mathrm{d}}\sum_{i=1}^{K}\tilde{\rho}_i+1}\right), \tag{36}$$

where $\hat{\beta}_i = \frac{\tau p_{\mathrm{u}}\beta_i^2}{\tau p_{\mathrm{u}}\beta_i+1}$, $\tilde{\beta}_i = \frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i+1}$, *and* $\tilde{\rho}_i = \frac{1}{M}\sum_{m=1}^{M}\frac{\rho_{m,i}}{\tau p_{\mathrm{d}}\rho_{m,i}+1}$.

Proof. From (15), we have $\boldsymbol{W}^T\hat{\boldsymbol{G}} = \boldsymbol{I}_K$ and $\boldsymbol{W}^T\boldsymbol{W}^* = [(\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}})^{-1}]^H$ for the ZF linear processing, so we can get $\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_k = 1$, $\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_i = 0$, and $\left\|\boldsymbol{w}_k^T\right\|^2 = [(\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}})^{-1}]_{kk}$. Similarly, as [19], using the convexity of $\varphi(x) = \log_2(1+x^{-1})$ and Jensen's inequality, from (28), we can obtain

$$\begin{aligned} R_k &\geq \log_2\left(1+\frac{p_{\mathrm{u}}}{\mathbb{E}\left\{p_{\mathrm{u}}\sum\limits_{i=1}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{\delta}_i\right|^2+p_{\mathrm{d}}\sum\limits_{i=1}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{\epsilon}_i\right|^2+\left\|\boldsymbol{w}_k^T\right\|^2\right\}}\right) \\ &= \log_2\left(1+\frac{p_{\mathrm{u}}}{\mathbb{E}\left\{\left(p_{\mathrm{u}}\sum\limits_{i=1}^{K}|\tilde{\delta}_i|^2+p_{\mathrm{d}}\sum\limits_{i=1}^{K}|\tilde{\epsilon}_i|^2+1\right)\left\|\boldsymbol{w}_k^T\right\|^2\right\}}\right). \end{aligned} \tag{37}$$

From Lemma 2 and Lemma 3, we know that the new variables $\tilde{\delta}_i = \frac{\boldsymbol{w}_k^T\boldsymbol{\delta}_i}{\left\|\boldsymbol{w}_k^T\right\|}$ and $\tilde{\epsilon}_i = \frac{\boldsymbol{w}_k^T\boldsymbol{\epsilon}_i}{\left\|\boldsymbol{w}_k^T\right\|}$ are both zero-mean

Gaussian variables with variances of $\tilde{\beta}_i$ and $\tilde{\rho}_i$, respectively, so we have

$$R_k \geq \log_2\left(1+\frac{p_{\mathrm{u}}}{\left(p_{\mathrm{u}}\sum_{i=1}^{K}\tilde{\beta}_i+p_{\mathrm{d}}\sum_{i=1}^{K}\tilde{\rho}_i+1\right)\mathbb{E}\left\{\left\|\boldsymbol{w}_k^T\right\|^2\right\}}\right). \tag{38}$$

Then, by using the property of Wishart matrices $\mathbb{E}\{\mathrm{tr}(\boldsymbol{M}^{-1})\}=K/(M-K)$ (Lemma 2.10) in [34], we have

$$\begin{aligned}\mathbb{E}\left\{\left\|\boldsymbol{w}_k^T\right\|^2\right\} &= \mathbb{E}\left\{[\,(\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}})^{-1}]_{kk}\right\}\\ &= \hat{\beta}_k^{-1}\mathbb{E}\left\{[\,(\hat{\boldsymbol{H}}^H\hat{\boldsymbol{H}})^{-1}]_{kk}\right\}\\ &= \frac{1}{K\hat{\beta}_k}\mathbb{E}\left\{\mathrm{tr}((\hat{\boldsymbol{H}}^H\hat{\boldsymbol{H}})^{-1})\right\}\\ &= \frac{1}{(M-K)\hat{\beta}_k},\quad \text{for}\quad M\geq K+1. \end{aligned}\tag{39}$$

Substitute (39) into (38), we can obtain the result. □

Proposition 4 *With the combined scheme of SI removal and MRT/MRC large-scale antenna linear processing, the kth uplink achievable rate in the shared-antenna FD massive MU-MIMO system is lower-bounded by*

$$\underline{R}_k^{\mathrm{M}} = \log_2\left(1+\frac{p_{\mathrm{u}}(M-1)\hat{\beta}_k}{p_{\mathrm{u}}\left(\sum_{i=1}^{K}\beta_i-\hat{\beta}_k\right)+p_{\mathrm{d}}\sum_{i=1}^{K}\tilde{\rho}_i+1}\right), \tag{40}$$

where $\hat{\beta}_i=\frac{\tau p_{\mathrm{u}}\beta_i^2}{\tau p_{\mathrm{u}}\beta_i+1}$ *and* $\tilde{\rho}_i=\frac{1}{M}\sum_{m=1}^{M}\frac{\rho_{m,i}}{\tau p_{\mathrm{d}}\rho_{m,i}+1}$.

Proof. From (15), we have $\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_k=\frac{1}{M}\left\|\hat{\boldsymbol{g}}_k\right\|^2$ and $\left\|\boldsymbol{w}_k^T\right\|^2=\frac{1}{M^2}\left\|\hat{\boldsymbol{g}}_k\right\|^2$for the MRT/MRC linear processing. Similarly, from (28) and by using Jensen's inequality, we have

$$R_k\geq\log_2\left(1+\frac{p_{\mathrm{u}}}{\mathbb{E}\left\{\frac{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\tilde{g}_i|^2+p_{\mathrm{u}}\sum_{i=1}^{K}|\tilde{\delta}_i|^2+p_{\mathrm{d}}\sum_{i=1}^{K}|\tilde{\epsilon}_i|^2+1}{\left\|\hat{\boldsymbol{g}}_k\right\|^2}\right\}}\right). \tag{41}$$

Then, by using the property of the Wishart matrix (Lemma 2.10) in [34], we get

$$\mathbb{E}\left\{\left\|\hat{\boldsymbol{g}}_k\right\|^{-2}\right\}=\mathbb{E}\left\{(\hat{\boldsymbol{g}}_k^H\hat{\boldsymbol{g}}_k)^{-1}\right\}=\frac{1}{(M-1)\hat{\beta}_k},\quad\text{for}\quad M\geq 2. \tag{42}$$

From Lemma 1, Lemma 2, and Lemma 3, by substituting (42) into (41), we obtain the result. □

4 Spectral efficiency and energy efficiency

The SE and EE are important performance metrics for next-generation mobile communication systems. In this section, we discuss the SE and EE of a shared-antenna FD massive MU-MIMO system with the combined SI suppression method.

4.1 SE and EE

4.1.1 SE and EE for the FD

Generally, the system SE can be defined as the sum-rate per site or sum-rate per cell. According to (28), the SE of the SI suppression and channel estimation in the shared-antenna FD massive MU-MIMO can be given as

$$S_{\mathrm{f}}=\frac{T-2\tau}{T}\sum_{k=1}^{K}R_k. \tag{43}$$

The EE is defined as the SE divided by the total transmit power, so we have

$$E_{\mathrm{f}}=\frac{S_{\mathrm{f}}}{Kp_{\mathrm{u}}}=\frac{T-2\tau}{TKp_{\mathrm{u}}}\sum_{k=1}^{K}R_k. \tag{44}$$

4.1.2 SE and EE for the HD

In comparison, the SE and EE of the HD system can be given as

$$S_{\mathrm{h}}=\frac{T-\tau}{2T}\sum_{k=1}^{K}\mathbb{E}\left\{\log_2\left(1+\gamma_k^{\mathrm{h}}\right)\right\}, \tag{45}$$

$$E_{\mathrm{h}}=\frac{T-\tau}{Kp_{\mathrm{u}}T}\sum_{k=1}^{K}\mathbb{E}\left\{\log_2\left(1+\gamma_k^{\mathrm{h}}\right)\right\}, \tag{46}$$

where γ_k^{h} is given as

$$\gamma_k^{\mathrm{h}}=\frac{p_{\mathrm{u}}\left|\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_k\right|^2}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}\left|\boldsymbol{w}_k^T\hat{\boldsymbol{g}}_i\right|^2+p_{\mathrm{u}}\sum_{i=1}^{K}\left|\boldsymbol{w}_k^T\boldsymbol{\delta}_i\right|^2+\left\|\boldsymbol{w}_k^T\right\|^2}. \tag{47}$$

We can see that the SE of the HD system is smaller than that of the FD system when the antenna scale is large enough, because it transmits data only half of the time. In contrast, the EE is slightly higher because there is no SI and fewer training overhead in the HD system.

4.1.3 Lower bounds of SE and EE

Replacing R_k by $\underline{R}_k^{\mathrm{Z}}$ and $\underline{R}_k^{\mathrm{M}}$, we can obtain the lower bounds of the SE and EE for the shared-antenna FD massive MU-MIMO system with imperfect estimation. By substituting (36) and (40) into (43) and (44), we have

$$\underline{S}_{\mathrm{f}}^{A}=\frac{T-2\tau}{T}\sum_{k=1}^{K}\underline{R}_k^{A}, \tag{48}$$

and

$$\underline{E}_{\mathrm{f}}^{A} = \frac{T-2\tau}{TKp_{\mathrm{u}}} \sum_{k=1}^{K} \underline{R}_{k}^{A}, \tag{49}$$

where the superscript A is "Z" or "M", which represents ZF or MRT/MRC. The lower bounds are in the closed form, so they can be easily used to show the system performance.

4.2 Maximizing SE and EE

4.2.1 Maximizing SE

The SE monotonically grows with the antenna array scale M and the uplink transmit power p_{u}. We suppose that M and p_{u} are fixed and M is large. In this case, only the training length τ affects the SE, and the optimization problem can be formulated as

$$\begin{aligned} &\max_{\tau} \quad S_{\mathrm{f}}(\tau) \\ &\text{s. t.} \quad K \leq \tau \leq T/2, \end{aligned} \tag{50}$$

This problem is neither linear nor convex, and it is even difficult to solve using the linear search because of the expectation operation and the large dimensions of the channels. Fortunately, the SE can be approximated by the lower bounds of (48), because they are very close after the further SI suppression with the combined method, which can be seen in the subsequent simulations in next section. Then, the maximization problem can be approximated as

$$\begin{aligned} &\max_{\tau} \quad \underline{S}_{\mathrm{f}}^{A}(\tau) \\ &\text{s. t.} \quad K \leq \tau \leq T/2. \end{aligned} \tag{51}$$

This problem can be numerically solved by the linear search, the simulation in the next section shows that this solution is global optimal when the antenna scale M is large enough.

4.2.2 Maximizing the EE

The EE of the FD massive MU-MIMO system monotonically increases with M, but not with p_{u}. Because it has the same solution as the SE for τ, here we only focus on p_{u}. The antenna-scale M is large and fixed, so the problem of maximizing the EE can be expressed as

$$\begin{aligned} &\max_{p_{\mathrm{u}}} \quad E_{\mathrm{f}}(p_{\mathrm{u}}) \\ &\text{s. t.} \quad 0 < p_{\mathrm{u}} \leq p_0. \end{aligned} \tag{52}$$

Similarly, the EE can be approximated by the lower bound in (49) when the antenna-array scale M is large enough. The maximization problem can be modified as

$$\begin{aligned} &\max_{p_{\mathrm{u}}} \quad \underline{E}_{\mathrm{f}}^{A}(p_{\mathrm{u}}) \\ &\text{s. t.} \quad 0 < p_{\mathrm{u}} \leq p_0. \end{aligned} \tag{53}$$

It can be numerically solved by the linear search, the simulation in the next section shows that this solution is also global optimal.

4.2.3 Tradeoff of EE versus SE

The maximal EE can be obtained at the point $(\tau^*, p_{\mathrm{u}}^*)$. However, the SE decreases monotonically with the decreasing p_{u}. To keep a reasonable uplink throughput, the SE is lower constrained. Therefore, the optimal p_{u}^* should be constrained by the SE requirement. The optimal value p_{u}^* is adopted if the minimum SE is satisfied; otherwise, the tradeoff value p'^*_{u} ,which meets the SE requirement, is used.

5 Numerical and simulation results

In this section, we present the performance results of the shared-antenna FD massive MU-MIMO system. All the users work in FD mode with a single antenna, and the insertion loss of the circulators is neglected. The number of users is given as $K = 4$. Unless stated otherwise, the system parameters are set as follows: The coherent time T of the uplink data transmission is 400 (symbol length), and the training length τ of each channel is 8. The normalized uplink power p_{u} and the downlink power p_{d} are set as 10 dB and 13 dB, respectively. Similarly, as [26], the large-scale path-loss β_k $(k = 1, 2, 3, 4)$ of the signal channel is set as 0.749, 0.246, 0.125, and 0.635, respectively; The large-scale path-loss β of the SI channel is set as 0.6, and the direct-path SI coefficient c_{mj} is set as arbitrary but deterministic values in a circle on the complex plane with radius 0.33.

5.1 Achievable rate and lower bounds

We obtain the numerical results according to (28), (31), (34), (36), and (40). The SI suppression and the achievable rate for the 2nd user are shown in Fig. 2 and Fig. 3, respectively. We can see that both the large-scale antenna linear processing method and the combined method can suppress the SI as well as the inter-user interferences and noise like [4, 15, 19, 26]. Our combined suppression

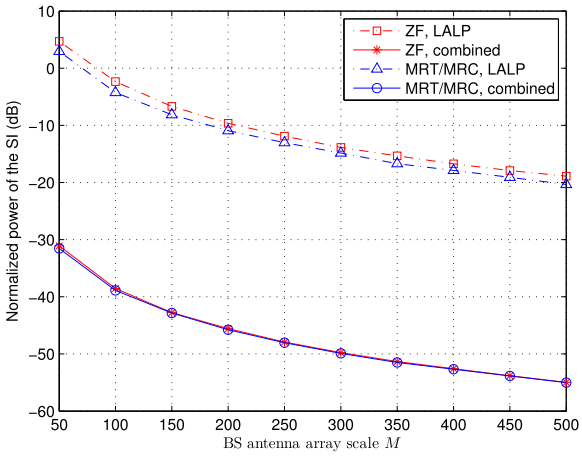

Fig. 2 SI at BS vs. antenna array scale M for the 2nd user

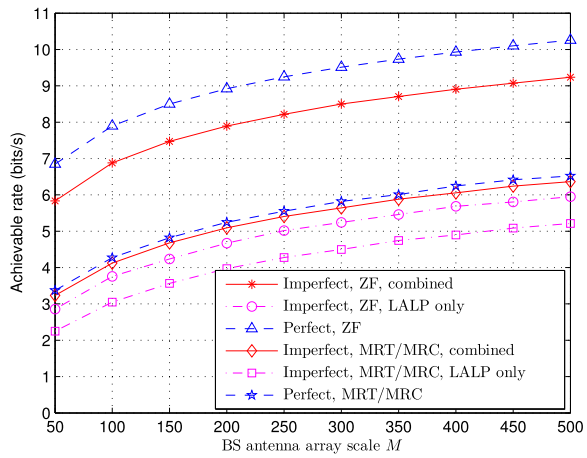

Fig. 3 Achievable rate of the 2nd user with perfect/imperfect estimation vs. the antenna array scale M

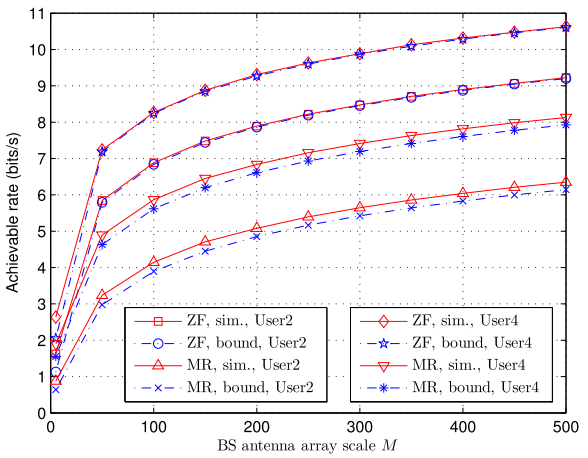

Fig. 4 Achievable rate lower bounds of the 2nd and 4th users with imperfect estimation vs. antenna array scale M

method significantly outperforms the LALP method. As shown in Fig. 2, at the scale of $M = 300$, the SI level is -14 dB by using the ZF linear processing only. In contrast, the SI is -50 dB by using the combined method, providing an additional 36 dB suppression. Hence, the achievable rates for the imperfect estimation are close to the perfect cases. For example, there are small gaps of approximately 1 bits/s/Hz for the ZF and 0.2 bits/s/Hz for the MRT/MRC with $M = 50$ in Fig. 3. Therefore, a higher data rate can be achieved at the same scale of the antenna array, or fewer antennas are required at the same data rate. For instance, the achievable rate increases from 5.3 bits/s/Hz to 8.5 bits/s/Hz at the antenna scale $M = 300$ for the ZF, while the scale M decreases from about 450 to 200 at the achievable rate of 5.1 bits/s/Hz for the MRT/MRC linear processing.

Figure 4 also shows the achievable rates and their lower bounds for the 2nd and 4th users. We can see that the lower bounds are very close to the simulation results when M is large enough. For example, there is a small gap of approximately 0.2 bits/s/Hz when M is 50 for the MRT/MRC, and the gap is even smaller for the ZF (the 2nd user). Therefore, it is reasonable to approximate the SE with the lower bounds.

5.2 SE and EE with SI suppression

Figure 5 and Fig. 6 show the SE and EE of the FD and HD systems versus the antenna array scale according to (43), (44), (45), and (46). We can see that the SE gets larger with the increase of the antenna array scale M for both the HD system and the FD system. The SE of the FD system is almost twice as high as that of the HD, as shown in Fig. 5, when employing our combined suppression method.

The EE of the FD system is slightly lower than that of the HD system because of the residual SI in the FD system, but the gaps are very small, so it is worth acquiring twice the SE by losing some EE. Furthermore, the EE grows with the increase of the antenna array scale M, so the performance is much better than that of the traditional FD system when M is large enough.

5.3 Maximizing SE and EE

Figures 7 and 8 show the SE and EE of the FD system based on (51) and (53). The scale M is supposed to be fixed and large enough. As shown in Fig. 7 and Fig. 8, both the SE and EE get higher at first and then turn lower. The approximate SE and EE are very close to the real values for large M, for example, $M = 500$. Since it is inefficient and time-consuming to search the optimal values with the Monte Carlo simulations, a good choice is to search the optimal values using the approximation. Simulations show that the SE increases when τ is small ($\tau < 10$), and decreases when

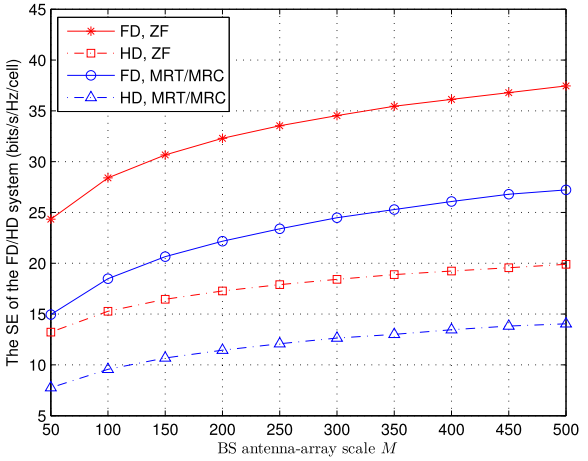

Fig. 5 SE vs. antenna array scale M for the FD and HD systems

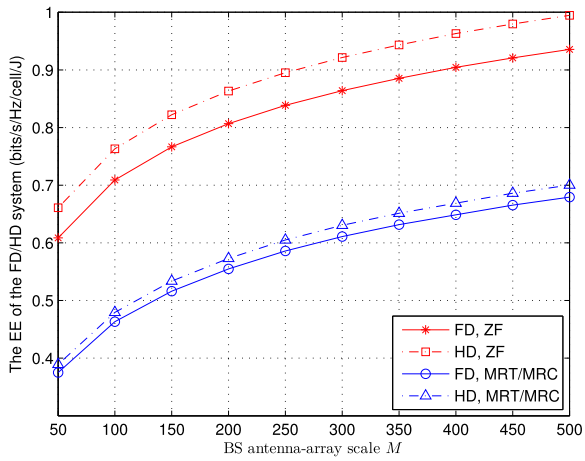

Fig. 6 EE vs. antenna array scale M for the FD and HD systems

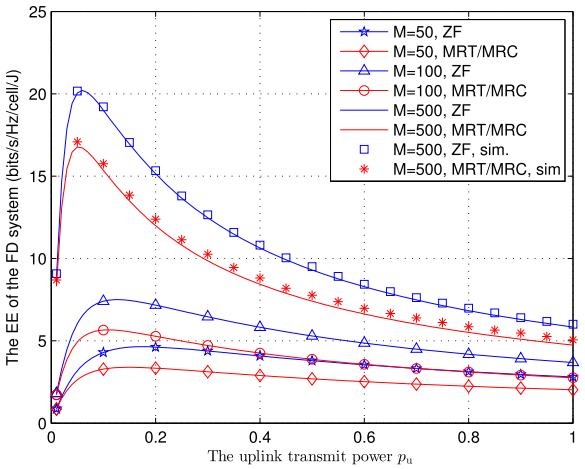

Fig. 8 EE versus uplink transmit power p_u for the FD system

τ is large, this is because the channels can be trained and estimated more sufficiently with the increase of τ. However, the data transmission time will be shortened if τ is too big. The similar results can be seen for the EE in Fig. 8, because the SE increases quadratically with p_u when p_u is small, while it increases logarithmically when p_u is big. The SE or EE increases with the antenna array scale M, because more SI can be suppressed and less training time/uplink transmit power is required when M is larger.

6 Conclusion

We introduced and analyzed a full-duplex massive MU-MIMO system with a single shared-antenna array at the BS. Either the ZF or MRT/MRC method can be adopted to linearly process signals, and a novel training scheme via the precoded SI channel is proposed to estimate the SI with a much lower dimension. We proved that the SI can vanish by using the large-scale antenna linear processing method even if the SI/signal channels are not perfectly estimated and the uplink/downlink channels completely depend on each other. By combining the SI removal and the LALP method, we show that the SI can be sufficiently suppressed. We also derived the closed-form lower bounds of the achievable rates and showed that they are very close to the simulation results. With the combination of the SI suppression, the "large enough" number of antennas can be scaled down, and the FD massive MU-MIMO system significantly outperforms the half-duplex system in terms of SE.

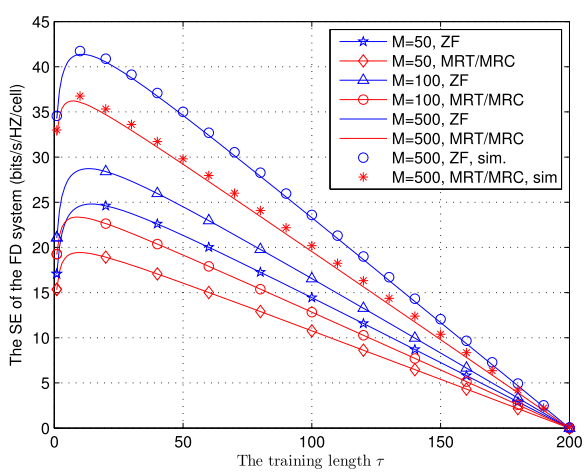

Fig. 7 SE vs. training length τ for the FD system

Appendix

Derivations of (6) and (7)

From (3), we have

$$\frac{\boldsymbol{G}_s^H \boldsymbol{G}_s}{M} = \frac{(\bar{\boldsymbol{G}}_s^H + \tilde{\boldsymbol{G}}_s^H)(\bar{\boldsymbol{G}}_s + \tilde{\boldsymbol{G}}_s)}{M} = \frac{(\bar{\boldsymbol{G}}_s^H \bar{\boldsymbol{G}}_s + \tilde{\boldsymbol{G}}_s^H \bar{\boldsymbol{G}}_s + \bar{\boldsymbol{G}}_s^H \tilde{\boldsymbol{G}}_s + \tilde{\boldsymbol{G}}_s^H \tilde{\boldsymbol{G}}_s)}{M}. \quad (54)$$

Since the direct-path interference is suppressed by the physical isolation and analog domain cancelation, the direct-path channel coefficients c_{mj} are not infinity. Because all of the entries c_{mj} of $\bar{\boldsymbol{G}}_s$ are bounded and all of the entries of $\tilde{\boldsymbol{G}}_s$ are i.i.d. Gaussian random variables with zero mean, we obtain $\frac{[\bar{\boldsymbol{G}}_s^H \tilde{\boldsymbol{G}}_s]_{mj}}{M} \approx 0$ when M is large enough, so we have

$$\frac{\boldsymbol{G}_s^H \boldsymbol{G}_s}{M} \approx \frac{(\bar{\boldsymbol{G}}_s^H \bar{\boldsymbol{G}}_s + \tilde{\boldsymbol{G}}_s^H \tilde{\boldsymbol{G}}_s)}{M} \approx \bar{\boldsymbol{D}}_s + \tilde{\boldsymbol{D}}_s. \quad (55)$$

From (6), we have

$$\frac{1}{M}\mathrm{tr}(\boldsymbol{G}_s^H\boldsymbol{G}_s) \approx \mathrm{tr}(\tilde{\boldsymbol{D}}_s + \tilde{\boldsymbol{D}}_s) = \frac{1}{M}\sum_{m=1}^{M}\sum_{j=1}^{M}|c_{mj}|^2 + \beta M \leq |c_{max}|^2 M + \beta M = O(M), \quad (56)$$

so we have

$$\mathrm{tr}(\boldsymbol{G}_s^H\boldsymbol{G}_s) = O(M^2). \quad (57)$$

Hence, we can obtain (6) and (7) when M is large enough.

Proof of Proposition 1

The ith column $\boldsymbol{g}_{\mathrm{sa},i}$ of matrix $\boldsymbol{G}_{\mathrm{sa}}$ can be written as

$$\boldsymbol{g}_{\mathrm{sa},i} = \boldsymbol{G}_{\mathrm{s}}\boldsymbol{a}_i, \quad (58)$$

where $\boldsymbol{a}_i$ is the ith column of $\boldsymbol{A}$, so the mth entry of $\boldsymbol{g}_{\mathrm{sa},i}$ can be given as

$$g_{\mathrm{sa},mi} = \boldsymbol{g}_{\mathrm{s},m}\boldsymbol{a}_i, \quad (59)$$

where $\boldsymbol{g}_{\mathrm{s},m}$ is the mth row of $\boldsymbol{G}_{\mathrm{s}}$.

For the ZF precoding, $\boldsymbol{a}_i = \frac{1}{M}\alpha_{\mathrm{ZF}}\beta_i^{-\frac{1}{2}}\boldsymbol{h}_i^*$, so we have

$$g_{\mathrm{sa},mi} = \frac{1}{M}\alpha_{\mathrm{ZF}}\beta_i^{-\frac{1}{2}}\sum_{j=1}^{M}\left(c_{mj} + \sqrt{\beta}h_{\mathrm{s},mj}\right)h_{ij}^*. \quad (60)$$

Hence, the variance of mth entry of $\boldsymbol{g}_{\mathrm{sa},i}$ can be given as

$$\begin{aligned}\mathbb{E}\left\{|g_{\mathrm{sa},mi}|^2\right\} &= \frac{1}{M^2}\alpha_{\mathrm{ZF}}^2\beta_i^{-1} \\ &\times \mathbb{E}\left\{\sum_{j=1}^{M}\left(c_{mj}+\sqrt{\beta}h_{\mathrm{s},mj}\right)h_{ij}^*\sum_{j=1}^{M}\left(c_{mj}^*+\sqrt{\beta}h_{\mathrm{s},mj}^*\right)h_{ij}\right\} \\ &= \frac{1}{M^2}\alpha_{\mathrm{ZF}}^2\beta_i^{-1}\sum_{j=1}^{M}\left(|c_{mj}|^2+\beta\right). \quad (61)\end{aligned}$$

Similarly, we can obtain

$$\mathbb{E}\left\{|g_{\mathrm{sa},mi}|^2\right\} = \alpha_{\mathrm{MRT}}^2\beta_i\sum_{j=1}^{M}\left(|c_{mj}|^2+\beta\right), \quad (62)$$

for the MRT precoding.

Proof of Lemma 3

The ith column $\boldsymbol{\epsilon}_i$ of matrix $\boldsymbol{\mathcal{E}}$ can be written as

$$\boldsymbol{\epsilon}_i = \boldsymbol{g}_{\mathrm{sa},i} - \hat{\boldsymbol{g}}_{\mathrm{sa},i}. \quad (63)$$

Note that the correlation between $\boldsymbol{w}_k$ and $\boldsymbol{\epsilon}_i$ is zero because $\boldsymbol{w}_k$ and $\boldsymbol{g}_{\mathrm{sa},i}$ are independent for $i \neq k$, and the correlation disappears for $i = k$ when $M \to \infty$. According to the optimum estimation of MMSE, $\boldsymbol{\epsilon}_i$ is a Gaussian random vector, and the variance of the mth entry for $\boldsymbol{\epsilon}_i$ is $\frac{\rho_{m,i}}{\tau p_{\mathrm{d}}\rho_{m,i}+1}$. Similarly, from the Lyapunov central limit theorem [35] and Proposition 1, we can derive that the variable $\tilde{\epsilon}_i = \frac{\boldsymbol{w}_k^T\boldsymbol{\epsilon}_i}{\|\boldsymbol{w}_k^T\|}$ is a zero-mean complex Gaussian variable, and the variance is

$$\mathbb{E}\left\{|\tilde{\epsilon}_i|^2\right\} = \frac{1}{M}\sum_{j=1}^{M}\frac{\rho_{m,i}}{\tau p_{\mathrm{d}}\rho_{m,i}+1}. \quad (64)$$

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61371188), Research Fund for the Doctoral Program of Higher Education (20130131110029), China Postdoctoral Science Foundation (2014M560553), Special Funds for Postdoctoral Innovative Projects of Shandong Province (201401013).

Competing interests

The authors declare that they have no competing interests.

Received: 3 June 2016 Accepted: 2 January 2017
Published online: 19 January 2017

References

1. A Sabharwal, P Schniter, D Guo, DW Bliss, S Rangarajan, R Wichman, In-band full-dulex wireless: Challenges and opportunities. IEEE J. Sel. Areas Commun. **32**(9), 1637–1652 (2014)
2. S Hong, J Brand, JI Choi, M Jain, J Mehlman, S Katti, P Levis, Applications of self-interference cancellation in 5G and beyond. IEEE Commun. Mag. **52**(2), 114–121 (2014)
3. EG Larsson, O Edfors, F Tufvesson, TL Marzetta, Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
4. F Rusek, D Persson, BK Lau, EG Larsson, TL Marzetta, O Edfors, F Tufvesson, Scaling up MIMO: Opportunities and challenges with very large arrays. IEEE Signal Process. Mag. **30**(1), 40–60 (2013)
5. Z Zhang, X Chai, K Long, AV Vasilakos, L Hanzo, Full duplex techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection. IEEE Commun. Mag. **53**(5), 128–137 (2015)
6. X Jia, P Deng, L Yang, H Zhu, Spectrum and energy efficiencies for multiuser pairs massive MIMO systems with full-duplex amplify-and-forward relay. IEEE Access. **3**, 1907–1918 (2015)
7. TL Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wireless Commun. **9**(11), 3590–3600 (2010)
8. JI Choi, M Jain, K Srinivasan, P Levis, S Katti, in *Proc. ACM MobiCom On*. Achieving single channel, full duplex wireless communications, (Chicago, 2010), pp. 1–12
9. S Goyal, P Liu, SS Panwar, RA Difazio, R Yang, E Bala, Full duplex cellular systems: Will doubling interference prevent doubling capacity? IEEE Commun. Mag. **53**(5), 121–127 (2015)
10. KM Thilina, H Tabassum, E Hossain, DI Kim, Medium access control design for full duplex wireless systems: Challenges and approaches. IEEE Commun. Mag. **53**(5), 112–120 (2015)
11. Y Liao, K Bian, L Song, Z Han, Full–duplex MAC protocol design and analysis. IEEE Commun. Lett. **19**(7), 1185–1188 (2015)
12. L Wang, F Tian, T Svensson, D Feng, M Song, S Li, Exploiting full duplex for device-to-device communications in heterogeneous networks. IEEE Commun. Mag. **53**(5), 146–152 (2015)
13. M Heino, D Korpi, T Huusari, E Antonio-Rodriguez, et al, Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays. IEEE Commun. Mag. **53**(5), 91–101 (2015)
14. L Song, Y Li, Z Han, Resource allocation in full-duplex communicaitons for future wireless networks. IEEE Wireless Commun. **22**(4), 88–96 (2015)
15. J Hoydis, ST Brink, M Debbah, Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? IEEE J. Sel. Areas Commun. **31**(2), 160–171 (2013)
16. X Rao, VKN Lau, Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. IEEE Trans. Signal Process. **62**(12), 3261–3271 (2014)
17. B Lee, J Choi, J Seol, DJ Love, B Shim, Antenna grouping based feedback compression for FDD-based massive MIMO systems. IEEE Trans. Commun. **63**(9), 3261–3274 (2015)

18. L Lu, GY Li, AL Swindlehurst, A Ashikhmin, R Zhang, An overview of massive MIMO: Benefits and challenges. IEEE J. Sel. Topics Signal Process. **8**(5), 742–758 (2014)
19. HQ Ngo, EG Larsson, TL Marzetta, Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans. Commun. **61**(4), 1436–1449 (2013)
20. X Pang, W Hong, T Yang, L Li, Design and implementation of an active multibeam antenna system with 64 RF channels and 256 antenna elements for massive MIMO application in 5G wireless communications. China Commun. **11**(11), 16–23 (2014)
21. TL Marzetta, in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC06) On*. How much training is required for multiuser MIMO? (Pacific Grove, 2006), pp. 359–363
22. T Riihonen, S Werner, R Wichman, Mitigation of loopback self-interference in full-duplex MIMO relays. IEEE Trans. Signal Process. **59**(12), 5983–5993 (2011)
23. E Everett, A Sahai, A Sabharwal, Passive self-interference suppression for full-duplex infrastructure nodes. IEEE Trans. Wireless Commun. **13**(2), 680–694 (2014)
24. D Bharadia, E McMilin, S Katti, in *Proc. ACM SIGCOMM On*. Full duplex radio, (Hong Kong, 2013), pp. 375–386
25. M Duarte, A Sabharwal, V Aggarwal, et al, Design and characterization of a full-duplex multiantenna system for WiFi networks. IEEE Trans. Veh. Technol. **63**(3), 1160–1177 (2013)
26. HQ Ngo, HA Suraweera, M Matthaiou, EG Larsson, Multipair full-duplex relaying with massive arrays and linear processing. IEEE J. Sel. Areas Commun. **32**(9), 1721–1737 (2014)
27. P Xing, J Liu, C Zhai, X Wang, L Zheng, Self-interference suppression for the full-duplex wireless communication with large-scale antenna. Trans. Emerg. Telecom. Technol. **27**(6), 764–774 (2016)
28. Z Zhang, Z Chen, M Shen, B Xia, Spectral and energy efficiency of multi-pair two-way full-duplex relay systems with massive MIMO. IEEE J. Sel. Areas Commun. **34**(4), 848–863 (2016)
29. M Duarte, C Dick, A Sabharwal, Experiment-driven characterization of full-duplex wireless systems. IEEE Trans. Wireless Commun. **11**(12), 4296–4307 (2012)
30. S Haykin, *Adaptive Filter Theory*, 4th edn. (Prentice-Hall, Upper Saddle River, 2002)
31. H Yang, TL Marzetta, Performance of conjugate and zero-forcing beamforming in large-scale antenna systems. IEEE J. Sel. Areas Commun. **31**(2), 172–179 (2013)
32. J Hoydis, *Random matrix theory for advanced communication systems*, Ph.D. dissertation. (Supelec Univ., France, 2012)
33. R Couillet, M Debbah, *Random matrix methods for wireless communications*, 1st edn. (Cambridge Univ. Press, New York, 2011)
34. AM Tulino, S Verdu, Random matrix theroy and wireless communications. Foundations Trends Commun. Inf. Theory. **1**(1), 1–182 (2004)
35. RJ Serfling, *Approximation Theorems of Mathematical Statistics*, 2nd edn. (John Wiley & Sons, New York, 1980)